Optimal Income Taxation:

Mirrlees Meets Ramsey[*]

Jonathan Heathcote
Minneapolis Fed and CEPR

Hitoshi Tsujiyama
Goethe University Frankfurt

April 2, 2016

Abstract

What is the optimal shape of the income tax schedule? This paper compares the optimal (Mirrlees) tax and transfer policy to various simple parametric (Ramsey) alternatives. The environment features distinct roles for public and private insurance. We explore a flexible class of social welfare functions, one special case of which captures the taste for redistribution reflected in the current tax system. Optimal marginal tax rates generally increase in income, a pattern qualitatively similar to the U.S. system, and are neither flat nor U-shaped. The optimal tax schedule can be well approximated by a simple two parameter function that is easily communicated to policymakers.

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

1 Introduction

In this paper 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?

A natural starting point for characterizing the optimal structure of taxation is 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 to first formulate a planner's problem that features incentive constraints such that individuals are willing to choose the allocations intended for their unobserved productivity types. Given a solution to this problem, an earnings tax schedule can be inferred, such that the same allocations are decentralized as a competitive equilibrium given those taxes. This approach is attractive because it places no constraints on the shape of the tax schedule, and because the implied allocations are constrained efficient.

The alternative Ramsey approach to tax design is to restrict the planner to choose a tax schedule within a parametric class. Although there are no theoretical foundations for imposing ad hoc restrictions on the design of the tax schedule, the practical advantage of doing so is that one can then consider tax design in richer models. In this paper we systematically compare the fully optimal non-parametric Mirrlees policy with two common parametric functional forms for the income tax schedule, T, that maps income, y, into taxes net of transfers, $T(y)$. The first is an affine tax: $T(y) = \tau_0 + \tau_1 y$, where τ_0 is a lump-sum tax or transfer, and τ_1 is a constant marginal tax rate. Under this specification, a higher marginal tax rate τ_1 translates into larger lump-sum transfers and thus more redistribution. The second tax function is $T(y) = y - \lambda y^{1-\tau}$. This specification rules out lump-sum transfers, but for $\tau > 0$ implies marginal tax rates that increase with income. Heathcote, Storesletten, and Violante (2016) (henceforth HSV) show that this function closely approximates the current U.S. tax and transfer system.

By comparing welfare in the two cases, we will learn whether in designing a tax system it is more important to allow for lump-sum transfers (as in the affine case) or to allow for marginal tax rates to increase with income (as in the HSV case). We will also be interested in whether either

1

affine or HSV tax systems come close to decentralizing constrained efficient allocations, or whether a more flexible functional form is required.

Our paper adds to an extensive literature investigating the optimal shape of the tax and transfer system. Many authors have argued for an affine "flat tax" system, with constant marginal tax rates and redistribution being achieved via universal transfers. For example, Friedman (1962) advocated a "negative income tax," which effectively combines a lump-sum transfer with a constant marginal tax rate. Mirrlees (1971) found the optimal schedule to be close to linear, a finding that extends to many other papers in the subsequent literature (see, e.g., Mankiw et al. 2009). In contrast, others, and most notably Saez (2001), have argued that marginal tax rates should be U-shaped, with higher rates at low and high incomes than in the middle of the income distribution. In our baseline calibration we find neither flat taxes nor U-shaped taxes to be optimal. Rather, we find that the optimal system features marginal tax rates that are generally increasing across the entire income distribution, a pattern qualitatively similar to the system currently in place in the United States.

We study a standard environment in which agents differ with respect to productivity, and where the government chooses an income tax system to redistribute and to finance exogenous government purchases. The analysis builds on the existing literature in two dimensions that are important for offering quantitative guidance on the welfare-maximizing shape of the tax function. First, we assume that agents are able to privately insure a share of idiosyncratic labor productivity risk and emphasize the role of the tax system in addressing risks that are not privately insurable. Second, rather than focussing exclusively on a utilitarian welfare criterion, we evaluate alternative tax systems using a wide range of alternatives, including a social welfare function that is designed to be consistent with the amount of redistribution embedded in the U.S. tax code. We now discuss these two innovations in more detail.

The existing literature mostly abstracts from private insurance, but for the purposes of providing concrete practical advice on tax system design it is very important to appropriately specify the relative roles of public and private insurance. Private insurance in the model operates as follows. 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 is also unobserved by the planner, but can be perfectly

privately insured by agents. What we have in mind is that agents face shocks that the tax authority does not directly observe, but which agents can smooth in a variety of ways. One important source of private insurance, which will be the focus of our exposition, is insurance within the family, where it is reasonable to assume that family members have much better information about each other's productivity than does the tax authority. When agents can insure more risks privately, the government has a smaller role in providing social insurance, and the optimal tax schedule is less progressive.

The shape of the optimal tax schedule in any social insurance problem is necessarily sensitive to the social welfare function that the planner is assumed to be maximizing. We will consider a class of social welfare functions in which the weight on an agent with uninsurable idiosyncratic productivity component α takes the form $\exp(-\theta\alpha)$. Here the parameter θ determines the taste for redistribution. To facilitate comparison with the existing literature, we use the utilitarian case ($\theta = 0$) as our baseline, but we will also assess how robust our policy prescriptions are to alternative values for θ. We will argue that the degree of progressivity built into the actual U.S. tax and transfer system is informative about U.S. policymakers' taste for redistribution. In particular, we characterize in closed form the mapping between the taste for redistribution parameter θ in our class of social welfare functions and the progressivity parameter τ that maximizes welfare within the HSV class of tax / transfer systems. This mapping can be inverted to infer the U.S. taste for redistribution θ^* that would lead a planner to choose precisely the observed degree of tax progressivity τ^*.

The form of the distribution of uninsurable risk is known to be critical for the shape of the optimal tax function. In our calibration we are therefore careful to replicate observed dispersion in U.S. wages. Using cross-sectional data from the Survey of Consumer Finances, we show that the empirical earnings distribution is very well approximated by an Exponentially-Modified Gaussian (EMG) distribution. We estimate the corresponding parameters of the labor productivity distribution by maximum likelihood. We then use external estimates and evidence on consumption inequality to discipline the relative variances of the uninsurable and insurable components of wage risk.

Our key findings are as follows. First, in our baseline model, the welfare gains of moving from the current tax system to the tax system that decentralizes the Mirrlees solution are sizable. The

best policy in the HSV class is preferred to the best policy in the affine class, indicating that it is more important that marginal tax rates increase with income than that the tax system allows for lump-sum transfers. The best policy in the HSV class generates 84 percent of the maximum possible welfare gains from tax reform.

Second, the potential for large welfare gains from tax reform is very sensitive to the assumed taste for redistribution in the social welfare function. When we consider the case $\theta = \theta^*$ (the "empirically motivated" social welfare function), the potential gains from tax reform shrink to less than 0.1 percentage points of consumption, and moving to the best affine tax system is now welfare-reducing by around 0.6 percentage points of consumption. Thus, unless one believes strongly that a utilitarian social welfare objective is appropriate, one should be cautious about advocating drastic changes to the shape of the tax schedule.

Third, counter-factually assuming away private insurance while retaining our baseline social welfare function leads to a larger role for government redistribution and thus more progressive taxation. In this case, an affine tax function is preferred to the best policy in the HSV class. Thus, if we were to abstract from the existence of private insurance we would draw the wrong conclusions about the shape of the optimal tax function.

Fourth, we challenge the view that marginal rates should be high at low income levels in order to target transfers only to the very poor. In contrast to this view, optimal rates are relatively low at low income levels in our baseline calibration. Moreover, we use a version of the Diamond-Saez formula to show that better insurance at the bottom of the income distribution translates into lower optimal marginal tax rates for the poor. We illustrate this result with a sensitivity analysis with respect to the exogenous level of government consumption the government must finance: more expenditure reduces transfers to the poor, and raises marginal tax rates at the bottom of the income distribution.

In an extension to our baseline model, we introduce a third component of idiosyncratic productivity, κ, which is privately uninsurable but observed by the planner. This component is designed to capture differences in wages related to observable characteristics such as age and education. Because wages vary systematically by these characteristics, a constrained efficient tax system should explicitly index taxes to these observables (see, e.g., Weinzierl 2011). In our model we assume that κ is drawn before the agent can trade in financial markets and therefore cannot be insured privately.

We set the variance of this observable fixed effect to reflect the amount of wage dispersion that can be accounted for by standard observables in a Mincer regression. We find that if the planner can condition taxes on the observable component of labor productivity, it can generate large welfare gains, in part because tagging translates into lower marginal rates on average.

Related Literature   Seminal papers in the literature on taxation in the Mirrlees tradition include Mirrlees (1971), Diamond (1998), and Saez (2001). More recent work has focused on extending the approach to dynamic environments: Farhi and Werning (2013) and Golosov et al. (forthcoming) are perhaps 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 tradition in economies with heterogeneity and incomplete private insurance markets. Recent examples include Conesa and Krueger (2006), who explore the Gouveia and Strauss (1994) functional form for the tax schedule, and Heathcote et al. (2016), who explore the function used by Feldstein (1969), Persson (1983), and Benabou (2000). Our paper shares the goal of offering quantitative and practical guidance on tax design, but also provides systematic comparisons to the optimal Mirrlees policy.

Our interest in constructing social welfare functions that are broadly consistent with observed tax progressivity is related to Werning (2007). Werning's goal is to characterize the Pareto efficiency or inefficiency of any given tax schedule, given an underlying skill distribution. In contrast, our focus will be on quantifying the extent of inefficiency in the current system, rather than on a zero-one classification of efficiency.[1]

Recent papers by Bourguignon and Spadaro (2012), Brendon (2013), and Lockwood and Weinzierl (forthcoming) address the inverse of the optimal taxation problem, which is to characterize the profile for social welfare weights that rationalize a particular observed tax system: given these weights, the observed tax system is optimal by construction. Heathcote and Tsujiyama (2016) pursue the inverse-optimum approach in an environment similar to the present paper.

Our approach is similar to the inverse-optimum approach in that it uses the progressivity built into the observed tax system to learn about the shape of the planner's social welfare function. In contrast to the inverse-optimum approach, however, our approach restricts the social welfare

---

[1]In our model environment, the distribution of productivity will be bounded above. It follows immediately that the current tax system is not Pareto efficient, since it violates the zero-marginal-tax-at-the-top prescription.

function to a one parameter functional form which only allows for a simple tilt in social preferences toward (or against) relatively high productivity workers. We find this parametric assumption attractive because it is flexible enough to nest most of the standard social welfare functions used in the literature and also allows for a closed-form mapping between structural model parameters, including the observed progressivity of the tax system, and the planner's taste for redistribution. Restricting the social welfare function to belong to a simple parametric class rather than solving for the non-parametric inverse optimum Pareto weights is analogous to restricting the tax function to a simple parametric class (a la Ramsey) rather than solving for the fully optimal non-parametric Mirrlees schedule. We see merit in both approaches, and hope that the simplicity and flexibility of our approach will prove useful in future quantitative work on tax design.

Saez and Stantcheva (2016), Hendren (2014), and Weinzierl (2014) propose various interesting ways to generalize inter-personal 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. One advantage of our approach, which uses fixed Pareto weights that are specified ex ante, is that we can evaluate alternative functional forms for taxes that correspond to 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. They consider a range of alternative environments, in most of which agents face a single idiosyncratic shock that can be insured privately or publicly. Section III of their paper explores a more similar environment to ours, in which there are two components of productivity and differential roles for public versus private insurance with respect to the two components. Like us, they conclude that the government should focus on insuring the source of risk that cannot be insured privately. Relative to Chetty and Saez (2010), our contributions are twofold: (i) we consider optimal Mirrleesian tax policy in addition to affine tax systems, and (ii) our analysis is more quantitative in nature.

2   Environment

Labor Productivity   There is a unit mass of agents. Agents differ only with respect to labor productivity w, which has two orthogonal components: $\log w = \alpha + \varepsilon$. These two idiosyncratic components differ with respect to whether or not they can be observed and insured privately. 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. In Appendix B.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. For the purposes of optimal tax design, the details of how private insurance is delivered do not matter as long as the set of risks that is privately insurable remains independent of the choice of tax system, which is our maintained assumption.

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

In the simplest description of the model environment, the world is static, and each agent draws α and ε only once. However, it will become clear that there is an isomorphic dynamic interpretation in which agents draw new values for the insurable shock ε in each period. In that case, the differential insurance assumption could be reinterpreted as assuming that α represents fixed effects that are drawn before agents enter the economy, whereas ε captures life-cycle productivity shocks against which agents can purchase insurance.[2] A more challenging extension to the framework would be to allow for persistent shocks to the unobservable noninsurable component of productivity α. However, Heathcote et al. (2014) estimate that life-cycle uninsurable shocks account for only 17 percent of the observed cross-sectional variance of log wages.

Preferences   Agents have identical preferences over consumption, c, and work effort, h. The

---

[2]Although explicit insurance against life-cycle shocks may not exist, households can almost perfectly smooth transitory shocks to income by borrowing and lending.

utility function is separable between consumption and work effort and takes the form

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

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

Technology   Aggregate output in the economy is simply aggregate effective labor supply

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

Aggregate output is divided between private consumption and a publicly provided good G that is nonvalued:

$$Y = \int \int c(\alpha, \varepsilon) dF_\alpha(\alpha) dF_\varepsilon(\varepsilon) + G.$$

The resource constraint of the economy is thus given by

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

Insurance   We imagine insurance against ε shocks as occurring via a family planner who dictates hours worked and private within-family transfers for a continuum of agents who share a common uninsurable component α and whose insurable shocks ε are distributed according to F_ε. As will become clear, by modeling private insurance as occurring within the family, it will be very clear that there is no way for the government to monopolize all provision of insurance in the economy.

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

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

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

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

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

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

subject to (2) and the family budget constraint (3). In Appendix B.2 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.

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.

3. The government budget constraint is satisfied:

$$\int T(y(\alpha)) dF_\alpha(\alpha) = G. \tag{5}$$

3   Planner's Problems

The planner maximizes a social welfare function characterized by weights $W(\alpha)$ that potentially vary with α.[3]

---

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

3.1 Ramsey Problem

The Ramsey planner chooses the optimal tax function in a given parametric class $\mathcal{T}$. For example, for the class of affine functions, $\mathcal{T} = \{T : \mathbb{R}_+ \to \mathbb{R} | T(y) = \tau_0 + \tau_1 y \text{ for } y \in \mathbb{R}_+, \tau_0 \in \mathbb{R}, \tau_1 \in \mathbb{R}\}$.

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

$$\max_{T \in \mathcal{T}} \int W(\alpha) \int u(c(\alpha, \varepsilon), h(\alpha, \varepsilon)) dF_\varepsilon(\varepsilon) dF_\alpha(\alpha) \tag{6}$$

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

The first-order conditions (FOCs) to the family head's problem are

$$c(\alpha, \varepsilon) = c(\alpha) = y(\alpha) - T(y(\alpha)), \tag{7}$$

$$h(\alpha, \varepsilon)^\sigma = [y(\alpha) - T(y(\alpha))]^{-\gamma} \exp(\alpha + \varepsilon) \left[1 - T'(y(\alpha))\right]. \tag{8}$$

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

If the tax function satisfies

$$T''(y) > -\gamma \frac{[1 - T'(y)]^2}{y - T(y)} \tag{9}$$

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 (7) and (8) are therefore sufficient for optimality.

We now offer a sharper characterization of the efficient allocation of labor supply within the family for the tax functions in which we are particularly interested.

Affine Taxes Suppose taxes are an affine function of income, $T(y) = \tau_0 + \tau_1 y$.[4] Then we have

---

[4]Note that in this case, condition (9) is satisfied because

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

the following explicit solution for hours worked as a function of productivity $\exp(\alpha + \varepsilon)$ and family income $y(\alpha)$:

$$h(\alpha, \varepsilon) = \left[(y(a)(1 - \tau_1) - \tau_0)^{-\gamma} \exp(\alpha + \varepsilon) (1 - \tau_1) \right]^{\frac{1}{\sigma}} .$$

HSV Taxes   Suppose income taxes are in the HSV class, $T(y) = y - \lambda y^{1-\tau}$.[5] Then hours worked are given by

$$h(\alpha, \varepsilon) = \left[\exp(\alpha + \varepsilon) (1 - \tau) \lambda^{1-\gamma} y(\alpha)^{-(1-\tau)\gamma - \tau} \right]^{\frac{1}{\sigma}} . \tag{10}$$

3.2   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, where each family contains a continuum of members whose insurable component is distributed according to the common density F_ε. 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. It is clear that, by choosing taxes, the tax authority can choose the difference between income and consumption. It is less obvious that the planner can also dictate income *levels* as a function of type. To achieve this, the Mirrlees formulation of the planner's problem includes incentive constraints that guarantee that for each and every type α, a family of that type weakly prefers to deliver to the planner the value for income $y(\alpha)$ the planner intends for that type, thereby receiving $c(\alpha)$, rather than delivering any alternative level of income.

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

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

---

[5]Then condition (9) becomes

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

This is satisfied for any progressive tax, $\tau \in [0, 1)$, because $\tau + \gamma (1 - \tau) > 0$. It is also satisfied for any regressive tax, $\tau < 0$, if $\gamma \geq 1$, because $\gamma \geq 1 > \frac{-\tau}{1-\tau}$. Therefore, for all relevant parameterizations, condition (9) is also satisfied for this class of tax functions.

head solves

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

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

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

The first-order conditions to this problem give

$$
\begin{aligned}
c(\alpha, \tilde{\alpha}, \varepsilon) &= c(\tilde{\alpha}), \\
h(\alpha, \tilde{\alpha}, \varepsilon) &= \frac{y(\tilde{\alpha})}{\exp(\alpha)} \frac{\exp(\varepsilon)^{\frac{1}{\sigma}}}{\int \exp\left(\frac{1+\sigma}{\sigma}\varepsilon\right) dF_\varepsilon(\varepsilon)}.
\end{aligned}
\tag{11}
$$

Substituting in the allocations above, we get

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

First Stage Planner's Problem The planner maximizes social welfare, evaluated according to $W(\alpha)$, subject to the resource constraint, and subject to incentive constraints that ensure that family utility from reporting α truthfully and receiving the associated allocation is weakly larger than expected welfare from any alternative report and associated allocation:

$$\max_{\{c(\alpha),y(\alpha)\}} \int W(\alpha) U(\alpha,\alpha) dF_\alpha(\alpha), \tag{12}$$

$$\text{subject to} \quad \int c(\alpha) dF_\alpha(\alpha) + G = \int y(\alpha) dF_\alpha(\alpha), \tag{13}$$

$$U(\alpha,\alpha) \geq U(\alpha,\tilde{\alpha}) \quad \text{for all } \alpha \text{ and } \tilde{\alpha}. \tag{14}$$

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.[6] We will solve this problem numerically.

---

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

Decentralization with Income Taxes We have formulated the Mirrlees problem with the planner inviting families to report their unobservable characteristic α and then assigning the family an allocation for income $y(\tilde{\alpha})$ and consumption $c(\tilde{\alpha})$ on the basis of the report $\tilde{\alpha}$. But since the planner is offering agents a choice between a menu of alternative pairs for income and consumption, it is clear that an alternative way to think about what the planner does is that it offers 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.[7]

Suppose the family head maximizes family welfare, taking as given a tax on family income. We have already discussed the first-order conditions to this problem, eqs. (7) and (8). Substituting the expression for hours from eq. (11) into eq. (8) and letting $c^*(\alpha)$ and $y^*(\alpha)$ denote the values for family consumption and income that solve the Mirrlees problem (12), we can recover how optimal marginal tax rates vary with income:

$$1 - T'\left(y^*(\alpha)\right) = \frac{\Omega}{c^*(\alpha)^{-\gamma}\exp(\alpha)}\left(\frac{y^*(\alpha)}{\exp(\alpha)}\right)^{\sigma}. \tag{15}$$

3.3 First Best

If the planner can observe α directly, the welfare maximization problem is identical to the one described above, except that there are no incentive compatibility constraints (14). Formally, the planner's problem is to maximize (12) subject to (13).

4 Estimating Social Preferences

Absent knowledge of the government's objective function, it is difficult to compare alternative tax systems unless one Pareto dominates the other. As a baseline, we will compare alternative tax systems assuming the planner is utilitarian, since this is the most common approach in the literature. However, we will also be interested in comparing tax systems under alternative social welfare functions that embed a stronger or weaker taste for redistribution.

---

[7]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.

Throughout we will assume the social welfare function takes the form

$$W(\alpha; \theta) = \frac{\exp(-\theta\alpha)}{\int \exp(-\theta\alpha) dF_\alpha(\alpha)}. \qquad (16)$$

Here the single parameter θ controls the extent to which the planner puts relatively more or less weight on low relative to high productivity workers. With a negative θ, the planner puts relatively high weight on the more productive agents, whereas with a positive θ the planner overweights the less productive agents. One way to motivate an objective function of the form (16) is to appeal to a positive political economic model of electoral competition.[8]

This one-parameter specification is flexible enough to nest several standard social preference specifications that have been advocated in the literature. First, the case $\theta = 0$ corresponds to the baseline utilitarian case, with equal Pareto weights on all agents. Second, the case $\theta \to \infty$ corresponds to the maximal desire for redistribution. We label this the Rawlsian case, because in the environments we will consider (with elastic labor supply and unobservable uninsurable productivity) a planner with this objective function will seek to maximize the minimum level of welfare in the economy.[9] Third, the case $\theta = -1$ corresponds to a laissez-faire planner. The logic is that given preferences that are logarithmic in consumption (our baseline assumption), these planner weights are the inverse of equilibrium marginal utility absent any taxation.[10]

Empirically Motivated Social Welfare Function In addition to these special cases just described, there is one value for θ in which we will be especially interested, which is the value for θ that rationalizes the extent of redistribution embedded in the actual U.S. tax and transfer system. Heathcote et al. (2016) argue that the following income tax function closely approximates

---

[8]In the probabilistic voting model (see Persson and Tabellini 2000), 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 a weighted social welfare function similar to eq. (16) with $\theta < 0$, i.e., a function that puts more weight on more productive (and more tax sensitive) households.

[9]With elastic labor supply and unobservable shocks, the rankings of productivity and welfare will always be aligned. So maximizing minimum welfare is equivalent to maximizing welfare for the least productive household. With inelastic labor supply or observable shocks, a planner with $\theta > 0$ could and would deliver higher utility for low α households relative to high α households, so in such cases it would be wrong to label the case $\theta \to \infty$ Rawlsian.

[10]If the government needs to levy taxes to finance expenditure $G > 0$, then given $\theta = -1$, a planner that could observe α and apply α−specific lump-sum taxes would choose: (i) consumption proportional to productivity, $c(\alpha) \propto \exp(\alpha)$, and (ii) hours worked independent of α.

the actual U.S. tax and transfer system (see Section 5 for more details):

$$T\left(y\right) = y - \lambda y^{1-\tau}. \tag{17}$$

Thus, we adopt this specification as our baseline tax function. The marginal tax rate on individual income is given by $T'(y) = 1 - \lambda(1 - \tau)y^{-\tau}$. For $\tau > 0$, the tax system embeds the following properties: (i) marginal tax rates are increasing in income, with $T'(y) \to -\infty$ as $y \to 0$, and $T'(y) \to 1$ as $y \to \infty$, (ii) taxes net of transfers are negative for $y \in \left(0, \lambda^{\frac{1}{\tau}}\right)$, and (iii) marginal and average tax rates are related as follows: $(1 - T'(y)) / \left(1 - \frac{T(y)}{y}\right) = 1 - \tau$ for all y.

Because a higher value for τ corresponds to a higher ratio of marginal to average tax rates, τ is a natural index of tax progressivity. We let τ^* denote the degree of progressivity of the actual U.S. tax and transfer system.

Now, consider a Ramsey problem of the form (6) where the planner uses a social welfare function characterized by (16) and is restricted to choosing a tax-transfer policy within the parametric class described by (17). Although in principle the planner chooses two tax parameters, λ and τ, it 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 social welfare function indexed by θ. We define an *empirically motivated social welfare function* $W(\alpha; \theta^*)$ as the special case of the function defined in eq. (16) in which the taste for redistribution θ^* satisfies $\hat{\tau}(\theta^*) = \tau^*$. This approach to estimating a social welfare function can be generalized to apply to alternative tax function specifications.[11]

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

---

[11] In particular, for any representation of the actual tax and transfer scheme $T(y)$, one can always compute the value for θ that maximizes the social welfare associated with $W(\alpha; \theta)$, given the equilibrium allocations corresponding to $T(y)$.

At the same time, assuming that the social welfare function is in the class described by eq. (16) is an *ad hoc* restriction, and there likely exist alternative social welfare functions that make the maximum potential welfare gains from tax reform (relative to the HSV function with $\tau = \tau^*$) even smaller.[12] Thus, the welfare gains from optimal tax reform that we will find assuming the social welfare function is given by $W(\alpha; \theta^*)$ offer only an upper bound estimate for the inefficiency of the current HSV system. Still, this upper bound will turn out to be informative. Anticipating some of our quantitative results, we will find that moving to the best fully nonlinear Mirrlees policy generates very large welfare gains assuming $\theta = 0$ (a utilitarian objective) but very small welfare gains when $\theta = \theta^*$. The large gains in the former case simply reflect the fact that a utilitarian planner prefers much more redistribution that the current tax and transfer system delivers, while the small gain in the latter case indicates that the current tax system cannot be grossly inefficient.

A Closed-Form Link between Tax Progressivity and the Taste for Redistribution We now describe the operational details of how we reverse engineer an empirically motivated θ^* given the observed value for progressivity τ^*.

Our baseline calibration will assume that utility is logarithmic in consumption ($\gamma = 1$), that F_α is Exponentially-Modified Gaussian, $EMG(\mu_\alpha, \sigma_\alpha^2, \lambda_\alpha)$, and that F_ε is Gaussian, $N(-\sigma_\varepsilon^2/2, \sigma_\varepsilon^2)$. Given these functional form assumptions, we can use the government budget constraint (5) to solve in closed form for λ for any possible values for τ and G.[13] Given this expression for λ, we can derive a closed-form expression for social welfare for any possible taste for redistribution θ. This expression offers an implicit closed-form mapping between τ and θ. We use this mapping to ask for what value θ^* the social-welfare-maximizing value for τ is equal to the value for progressivity τ^* estimated from tax data.

Proposition 1 *The social preference parameter θ^* consistent with the observed choice for progressivity τ^* is a solution to the following quadratic equation:*

$$\sigma_\alpha^2 \theta^* - \frac{1}{\lambda_\alpha + \theta^*} = -\sigma_\alpha^2(1-\tau) - \frac{1}{\lambda_\alpha - 1 + \tau} + \frac{1}{1+\sigma}\left[\frac{1}{(1-g)(1-\tau)} - 1\right], \qquad (18)$$

---

[12]In Heathcote and Tsujiyama (2016) we characterize the (non-parametric) social welfare weights such that given those weights the observed tax system is fully optimal.

[13]With logarithmic consumption, we can solve in closed form for λ as a function of G and other structural parameters. For $\gamma > 1$, we must solve for λ numerically.

where g is the observed ratio of government purchases to output.

Proof. See Appendix A. ∎

Equation (18) is novel and very useful. Given observed choices for g and τ, 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^*. This is especially simple in the special case in which F_α is normal, since taking the limit $\lambda_\alpha \to \infty$ in (18) gives the following explicit solution for θ^{*}[14]

$$\theta^* = -(1-\tau) + \frac{1}{\sigma_\alpha^2}\frac{1}{(1+\sigma)}\left[\frac{1}{(1-g)(1-\tau)} - 1\right]. \tag{19}$$

For the purpose of inferring θ^*, we can treat g as exogenous.[15]

From eq. (18) it is straightforward to derive comparative statics on the mapping from structural policy and distributional parameters to θ^*, which we now briefly discuss (see Appendix B.3 for more details).

First, θ^{*} is increasing in τ. Thus, if we observe more progressive taxation, all else constant, we can infer that the policymaker puts less relative weight on high wage individuals. Second, θ^* is increasing in g. The logic is that tax progressivity reduces labor supply, making it more difficult to finance public spending. Thus, governments with high revenue requirements will tend to choose a less progressive system – unless they have a strong desire to redistribute. Third, θ^{*} is decreasing in σ_α^2. More uninsurable risk (holding fixed tax progressivity) suggests that the planner has less desire to redistribute. Fourth, θ^{*} is decreasing in σ. The less elastic is labor supply (and thus the smaller the distortions associated with progressive taxation), the less desire to redistribute we should attribute to the planner. Finally, θ^* is increasing in λ_α, holding fixed the total variance of the uninsurable component (namely, $\sigma_\alpha^2 + \lambda_\alpha^{-2}$). Thus, a more right-skewed distribution for α (a smaller λ_α) suggests a weaker taste for redistribution.

---

[14]This special case provides numerical guidance about which is the relevant root among the two solutions to the quadratic equation (18).

[15]If we were to contemplate the welfare effects of varying τ (holding fixed θ^* and G), it would be important to recognize that output and thus the ratio $G/Y(\tau)$ would change with different values for τ.

5 Calibration

Preferences We assume preferences are separable between consumption and labor effort and logarithmic in consumption:

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

This specification is the same one adopted by Heathcote et al. (2016). We choose $\sigma = 2$ so that the Frisch elasticity ($1/\sigma$) is 0.5. This value is broadly consistent with the microeconomic evidence (see, e.g., Keane 2011) and is also very close to the value estimated by Heathcote et al. (2014). The compensated (Hicks) elasticity of hours with respect to the marginal net-of-tax wage is approximately equal to $1/(1 + \sigma)$ (see Keane 2011, eq. 11) which, given $\sigma = 2$, is equal to $1/3$. Again this value is consistent with empirical estimates: Keane reports an average estimate across 22 studies of 0.31. Given our model for taxation, the elasticity of average income with respect to one minus the average income-weighted marginal tax rate is also equal to $1/(1+\sigma)$.[16] According to Saez et al. (2012), the best available estimates for the long run version of this elasticity range from 0.12 to 0.40, so again our calibration is consistent with existing empirical estimates. Note that because our logarithmic consumption preference specification is consistent with balanced growth, high and low wage workers will work equally hard in the absence of private insurance and redistributive taxation.

Tax and Transfer System The class of tax functions described by eq. (17) and that we label "HSV" was perhaps first used by Feldstein (1969) and introduced into dynamic heterogeneous agent models by Persson (1983) and Benabou (2000).

Heathcote et al. (2016) begin by noting that the functional form in (17) implies a linear relationship between $\log(y)$ and $\log(y - T(y))$, with a slope equal to $(1 - \tau)$. Thus, given micro data on household income before taxes and transfers and income net of taxes and transfers, it is straightforward to estimate τ by ordinary least squares. Using micro data from the Panel Study of Income Dynamics (PSID) for working-age households over the period 2000 to 2006, Heathcote et al. (2016) estimate $\tau = 0.161$. Figure 1, borrowed from that paper, shows the relationship between income before and after taxes and transfers for fifty equal-size bins of the distribution of household taxable

---

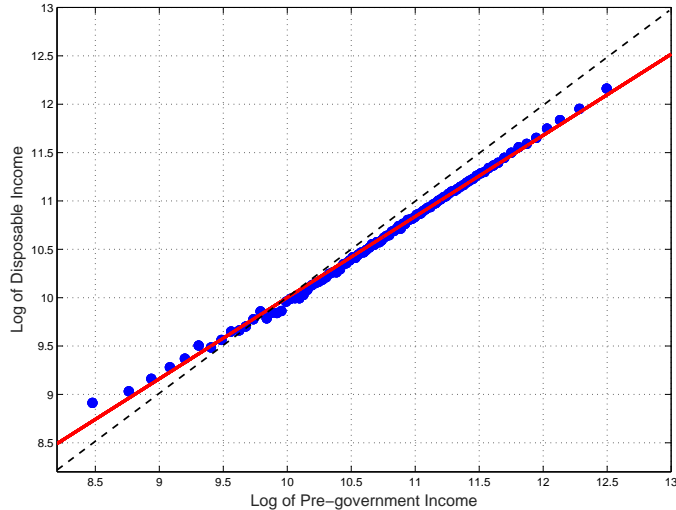[16]The average income-weighted marginal tax rate is $1 - (1 - g)(1 - \tau)$ (see Heathcote et al. 2016, eq. 4).

Figure 1: Fit of HSV tax function. Figure 1 in Heathcote et al. (2016), shows the relationship between log household income before and after taxes and transfers for working age households in the PSID. The red line shows the least squares best fit through the underlying micro data, with slope (1-0.161).

income, ranked from lowest to highest. The x-axis shows log of average pre-government taxable income for each bin, while the y-axis shows log of average income after taxes and transfers. The red line shows the least squares best fit through the underlying PSID micro data, with estimated slope $(1 - 0.161)$.

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

Wage Distribution We need to characterize individual productivity dispersion and to decompose this dispersion into orthogonal uninsurable and insurable components.

We assume that the insurable component of productivity, ε, is normally distributed, $\varepsilon \sim N(-\sigma_\varepsilon^2/2, \sigma_\varepsilon^2)$, and that the uninsurable component, α, follows an exponentially modified Gaussian (EMG) distribution: $\alpha = \alpha_N + \alpha_E$, where $\alpha_N \sim N(\mu_\alpha, \sigma_\alpha^2)$ and $\alpha_E \sim Exp(\lambda_\alpha)$ so that $\alpha \sim EMG(\mu_\alpha, \sigma_\alpha^2, \lambda_\alpha)$. This distributional assumption allows for a heavy right tail in the distribution for the uninsurable component of the log wage, which is heavier the smaller is the value for λ_α. Saez (2001) argued that there is more mass in the right tail of the log wage distribution than would be implied by a log-normal wage distribution and that this right tail is well approximated by

an exponential distribution. By attributing the heavy right tail in the log wage distribution to the uninsurable component of wages we are implicitly assuming that there is limited insurance against the risk of becoming extremely rich.[17]

Note that given these assumptions on the distributions for α and ε, the distribution of the log wage $(\alpha + \varepsilon)$ is itself EMG (the sum of the independent normally distributed random variables α_N and ε is normal) so the level wage distribution is Pareto log-normal. Furthermore, given our specifications for preferences and the baseline tax system, the distribution for log *earnings* is also EMG. Because preferences have the balanced growth property, hours worked are independent of the uninsurable shock α, and the exponential coefficient in the EMG distribution for log earnings is again λ_α, as for log wages. Hours do respond (positively) to insurable shocks, and the implied normal variance coefficient in the EMG distribution for log earnings is given by

$$\sigma_y^2 = \left(\frac{1+\sigma}{\sigma+\tau}\right)^2 \sigma_\varepsilon^2 + \sigma_\alpha^2. \tag{20}$$

As Mankiw et al. (2009) emphasize, it is difficult to sharply estimate the shape of the productivity distribution given typical household surveys, such as the Current Population Survey, in part because high income households tend to be under-represented 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.[18] We estimate λ_α and σ_y^2 by maximum likelihood, searching for the values of the three parameters in the EMG distribution that maximize the likelihood of drawing the observed 2007 distribution of log labor income.[19] The resulting estimates are $\lambda_\alpha = 2.2$ and $\sigma_y^2 = 0.4117$. Figure 2 plots the empirical density against a normal distribution with the same mean and variance and against the estimated EMG distribution. 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

---

[17]This assumption is consistent with the fact that a large fraction of individuals in the far right tail of the earnings distribution are entrepreneurs, and entrepreneurial risk is notoriously difficult to diversify.

[18]The SCF has some advantages over the IRS data used by Saez (2001). First, the unit of observation is the household, rather than the tax unit. Second, the IRS data exclude those who do not file tax returns or who file late. Third, people in principle have no incentive to under-report income to SCF interviewers.

[19]The empirical distribution for labor income in 2007 is constructed as follows. We define labor income as wage income plus two-thirds of income from business, sole proprietorship, and farm. We then restrict our sample to households with at least one member aged 25-60 and with household labor income of at least $10,000 (mean household labor income is $77,325).
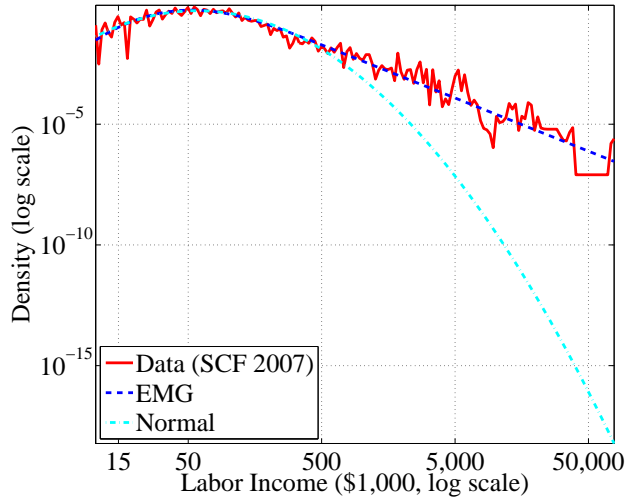
Figure 2: Fit of EMG distribution. The figure plots the empirical earnings density from the SCF against the estimated EMG distribution and against a normal distribution.

excellent fit, substantially improving on the normal specification.

Given values for σ and τ, and an estimate for σ_y^2, it remains only to partition the normal component of earnings dispersion, σ_y^2, into the components due to insurable versus uninsurable shocks (see eq. 20). Heathcote et al. (2016) estimate a richer version of the model considered in this paper using micro data from the PSID and the Consumer Expenditure Survey (CEX). They are able to identify the relative variances of the two wage components by exploiting two key implications of the theory: a larger variance for insurable shocks will imply a smaller cross-sectional variance for consumption and a larger covariance between wages and hours worked. Depending on how they model the right tail of the earnings distribution, their estimate for the variance of insurable shocks is either $\sigma_\varepsilon^2 = 0.139$ or $\sigma_\varepsilon^2 = 0.164$. In light of this evidence, we simply assume $\sigma_\varepsilon^2 = \sigma_\alpha^2$, which implies, via eq. (20), that $\sigma_\varepsilon^2 = \sigma_\alpha^2 = 0.1407$. Thus, the total model variance for log wages is $\sigma_\varepsilon^2 + \sigma_\alpha^2 + \lambda_\alpha^{-2} = 0.488$. For comparison, Heathcote et al. (2010, Figure5) report a log wage variance for men of 0.499 in the Current Population Survey in 2005.

Given these parameter values, 28.8 percent of the model variance of log wages and 43.8 percent of the variance of log earnings reflects insurable shocks.[20] One way to assess whether our decomposition of wage risk into uninsurable and insurable components is reasonable is to compare the extent of consumption inequality implied by the model to its empirical counterpart. Given the

---

[20]These shares are computed as $\sigma_\varepsilon^2/(\sigma_\varepsilon^2 + \sigma_\alpha^2 + \lambda_\alpha^{-2})$ and $\left(\left(\frac{1+\sigma}{\sigma+\tau}\right)^2 \sigma_\varepsilon^2\right) / \left(\left(\left(\frac{1+\sigma}{\sigma+\tau}\right)^2 \sigma_\varepsilon^2\right) + \sigma_\alpha^2 + \lambda_\alpha^{-2}\right)$.

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

Percentile Ratios	Model	LP (2015)
P5/P1	1.48	1.48
P10/P5	1.24	1.20
P25/P10	1.44	1.40

Note: P1, P5, P10 and P25 denote the 1st, 5th
10th, and 25th percentiles respectively.

calibration described above, the variance of log consumption in the model is 0.246. Heathcote et al. (2010, Figure 13) report a corresponding variance in the Consumer Expenditure Survey in 2006 of 0.332. However, Heathcote et al. (2014, Table 3) estimate that 29.6 percent of the variance of measured consumption reflects measurement error. Thus, we conclude that the model implies a realistic level of consumption inequality. In Section 6.4, we will explore how changing the relative magnitudes of insurable and uninsurable wage risk changes the optimal tax schedule.

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. In order to characterize optimal transfers and the optimal profile for marginal tax rates at the bottom of the earnings distribution, it is important to assess whether our wage distribution also accurately captures 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 the estimated model in Low and Pistaferri.[21] Reassuringly, the two sets of statistics are very similar.

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

---

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

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 $\{\alpha_1, \alpha_2, ..., \alpha_I\}$ with corresponding probabilities $\{\pi_1, \pi_2, ..., \pi_I\}$ as follows. We make the endpoints of the grid, α_1 and α_I, sufficiently extreme that only a tiny fraction of individuals lie outside these bounds in the true continuous distribution. In particular, we set α_1 such that $\exp(\alpha_1)/\sum_i(\pi_i \exp(\alpha_i)) = 0.05$, and set α_I such that $\exp(\alpha_I)/\sum_i(\pi_i \exp(\alpha_i)) = 74$, which corresponds to household labor income at the 99.99th percentile of the SCF labor income distribution (\$6.17 million) relative to average income.[22] We read corresponding probabilities π_i directly from the continuous EMG distribution, rescaling to ensure that (i) $\sum_i \pi_i = 1$, (ii) $\sum_i \pi_i \exp(\alpha_i) = 1$, and (iii) the variance of (discretized) α is equal to $\sigma_\alpha^2 + \lambda_\alpha^{-2}$. For our baseline set of numerical results we set $I = 10,000$. The resulting model wage distribution $\exp(\alpha + \varepsilon)$ is plotted in Figure 3. The distribution appears continuous, even though it is not, because our discretization is very fine. In Section 6.7 we report how the results change when we increase or reduce I. Note that when we compare the optimal Mirrlees tax schedule to various parametric Ramsey alternatives, we always use the same discrete grid for α, thereby ensuring that the environments differ only with respect to the tax and transfer system.

6   Quantitative Analysis

We now explore the structure of the optimal tax and transfer system, given the model specification described above.[23] We start in Section 6.1 by comparing welfare under alternative tax and transfer schemes, assuming a utilitarian social welfare function. Specifically, we compute the optimal tax and transfer systems in (i) the HSV class, (ii) the affine class, and (iii) the fully nonlinear Mirrlees framework, and compare allocations and welfare in each of those three cases with their counterparts under our baseline HSV approximation to the current U.S. tax and transfer system.

Section 6.2 explores alternative values for the taste for redistribution parameter θ. In Section 6.3 we work to understand the shape of the optimal tax schedule at the bottom of the income

---

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

[23]In Appendix B.4 we explain how we numerically solve the Mirrlees optimal tax problem.
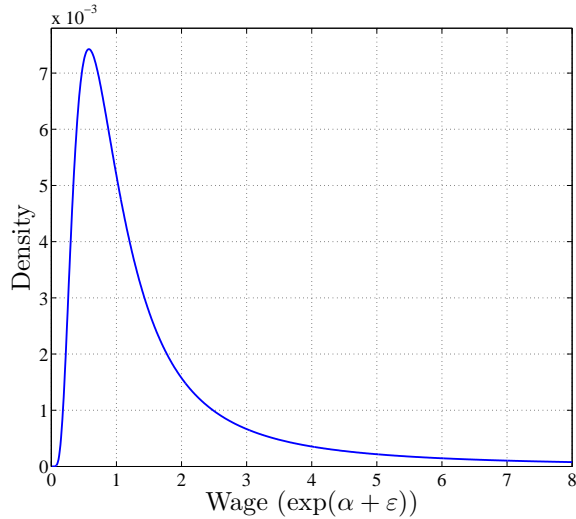
Figure 3: Model wage distribution. The plot is truncated at eight times the average wage, which is normalized to one.

distribution. We use sensitivity analysis with respect to government purchases and the taste for redistribution in combination with a simple approximation to the Diamond-Saez formula to shed new light on whether marginal tax rates should optimally be declining or increasing with income at low income levels. Section 6.4 considers the implications of assuming away all private insurance. Section 6.5 experiments with replacing our baseline EMG distribution for the uninsurable shock α with a normal distribution.

In Section 6.6 we explore richer tax structures. First, we consider polynomial tax functions that add quadratic and cubic terms to the affine functional form, thereby giving the Ramsey planner more flexibility. Second, we introduce a new component to individual labor productivity that cannot be insured privately but which is observed by the planner. This allows us to quantify the potential welfare gains from tagging: indexing taxes and transfers to characteristics such as age, education, gender, and marital status that are observable and correlated with wages.

6.1 Optimal Taxation in the Baseline Model

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

Table 2: Optimal Tax and Transfer System in the Baseline Model

Tax System	Tax Parameters		Outcomes			
			ω (%)	ΔY (%)	$\overline{T'}$	Tr/Y
HSV^{US}	$\lambda : 0.839$	$\tau : 0.161$	$-$	$-$	0.319	0.018
HSV	$\lambda : 0.817$	$\tau : 0.330$	2.08	-7.22	0.466	0.063
Affine	$\tau_0 : -0.259$	$\tau_1 : 0.492$	1.77	-8.00	0.492	0.279
Mirrlees			2.48	-7.99	0.491	0.213

to average income, Tr/Y.[24]

The first thing to note is that there are large potential welfare gains from tax reform here, and the nature of optimal reform is to make the tax system much more progressive. The best policy in the HSV class, for example, dictates an increase in the progressivity parameter τ from 0.161 to 0.330. This increases the average effective marginal tax rate from 31.9 percent to 46.6 percent. The associated additional disincentive to work is large, and reduces output by 7.22 percent. Nonetheless, the welfare gains to the utilitarian planner from larger net transfers to low income households are large, and overall the reform generates a welfare gain equivalent to giving all households 2.08 percent more consumption. The optimal non-linear tax system generates only a slightly larger welfare gain of 2.48 percent, and thus the best policy in the HSV class delivers 84 percent of the maximum possible welfare gains from tax reform. The best policy in the affine class does less well, delivering only 71 percent of the welfare gains from the optimal Mirrlees reform. This indicates that for welfare it is more important that marginal tax rates increase with income – which the HSV functional form accommodates but which the affine scheme rules out – than that the government provides universal lump-sum transfers – which only the affine scheme admits.

To develop intuition for these results, Figure 4 plots decision rules for consumption and hours (Panels A and B) and marginal and average tax schedules (Panels C and D) for each best-in-class

---

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

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

For the welfare numbers in Table 2, the baseline policy T is the current HSV tax system, and allocations are valued using $\theta = 0$.
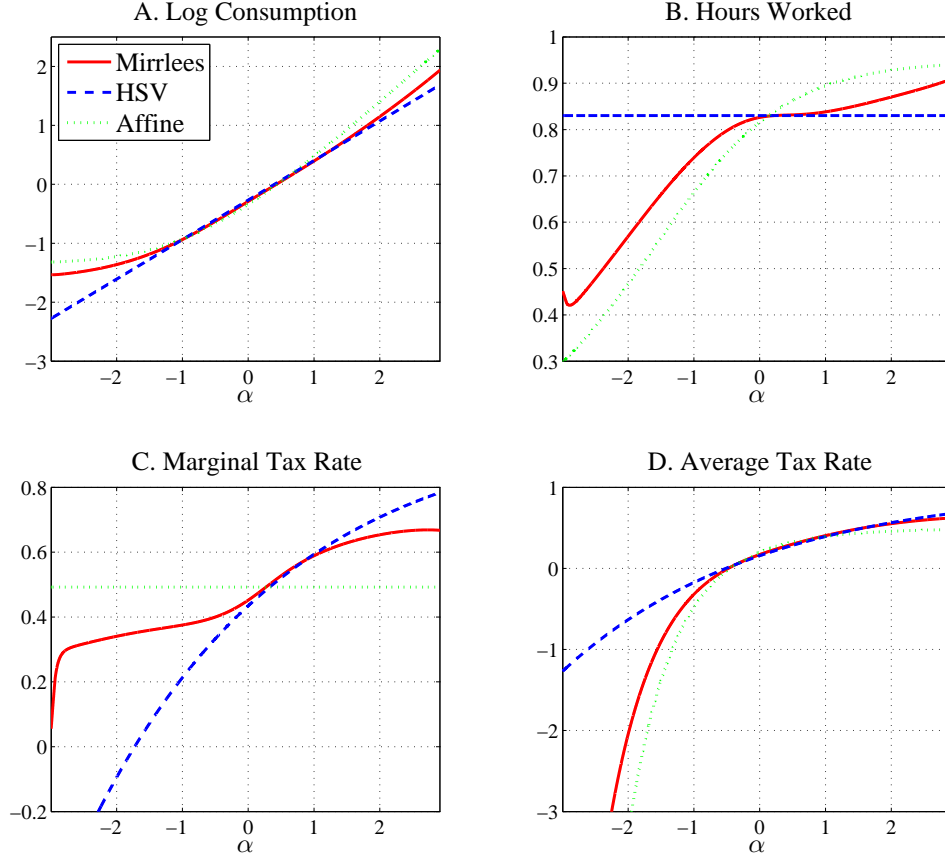
Figure 4: Mirrlees, HSV, and affine tax functions. The figure contrasts allocations under the HSV tax system (blue dashed line), the affine system (green dotted), and the Mirrlees system (red solid). Panels A and B plot decision rules for consumption and hours worked, while Panels C and D plot marginal and average tax schedules. The plot for hours worked is for an agent with average ε.

tax and transfer scheme. The figure compares particular third-best Ramsey-style tax functions (i.e., HSV and affine) to the second-best Mirrlees case.[25]

Allocations under the HSV policy are very similar to those in the constrained efficient Mirrlees case in the middle of the distribution for α, with larger differences in the tails of the distribution, especially for hours worked. Allocations are similar because the HSV marginal and average tax schedules are broadly similar to those under the optimal policy, especially for α between zero and one, corresponding to wages between the average wage and 2.7 times the average. In particular, the profile for marginal tax rates that decentralizes the constrained efficient allocation is generally increasing in productivity, and the HSV schedule captures this. However, while marginal tax rates

---

[25]At the very top of the distribution for uninsurable productivity α, the Mirrleesian marginal tax rate drops to zero. However, this happens only very close to the upper bound for α, the choice for which is somewhat arbitrary. To avoid being visually distracted by this property, we have truncated the visible range for productivity α at the 99.95^{th} percentile of the model distribution for α in this Figure and in subsequent similar ones.

increase smoothly under the HSV specification, the optimal schedule has a more complicated shape. The optimal marginal rate starts at 6 percent for the least productive households and is fairly flat (between 30 and 40%) up to half of average productivity.[26] The optimal marginal rate then rises rapidly to peak at 66.9 percent at 15 times average productivity. Because marginal rates are too high at the top under the HSV scheme, very productive agents work too little. At the same time, because transfers are too small, very unproductive agents work too much. Recall, however, that the mass of agents in these tails is small.

Panel C of Figure 4 offers a straightforward visualization of why an affine tax schedule is welfare inferior to the HSV form. Because the best affine tax function necessarily features a constant marginal rate, it cannot replicate the optimal marginal tax schedule, which rises rapidly in the middle of the productivity distribution. Under the affine scheme, low wage households face marginal tax rates that are too high relative to the optimal tax schedule, and in addition they receive relatively large lump-sum transfers. Thus, low productivity workers end up working 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 workers end up consuming too much.

Panel D of Figure 4 plots average equilibrium tax rates by household productivity. The difference between the average tax rates a household of a particular type faces under alternative tax schemes is closely tied to the difference in conditional welfare the household can expect: a higher average tax rate under one scheme translates into lower welfare. Thus, we can use the distribution of average tax rate differences across alternative tax schemes as a proxy for the distribution of relative welfare differences. Moving from the HSV schedule to the optimal one generates lower average tax rates and thus welfare gains for households in the tails, but not for the bulk of households who are in the middle of the productivity distribution.

To summarize, the optimal Mirrlees scheme redistributes via both lump-sum transfers and increasing marginal tax rates. The best affine schedule (which does not admit increasing marginal rates) does too much redistribution via lump-sum transfers, while the best HSV schedule (which does not admit lump-sum transfers) does too much redistribution via increasing marginal rates. Overall, having marginal rates that increase with income is a more important component of redis-

---

[26]At the very bottom of the productivity distribution the optimal allocation exhibits bunching: consumption and income are independent of α. This implies that hours are decreasing in α, while the marginal tax rate is strictly positive (see Ebert 1992) and increasing in α (see eq. 15).

tribution than lump-sum transfers, in the sense that the best HSV schedule is closer to the Mirrlees solution, in terms of welfare, than the best affine schedule.

6.2 Alternative Social Welfare Functions

We now consider alternative social welfare functions. There are two reasons to do so.

First, the fact that a utilitarian planner prefers much more redistribution that is embedded in the current U.S. tax and transfer scheme suggests that the U.S. planner is not in fact utilitarian and in fact has a weaker taste for redistribution. Thus, we want to explore tax reforms for planners with smaller values for θ than the utilitarian $\theta = 0$ case. We are particularly interested in our empirically motivated value θ^*, given which the observed progressivity parameter τ^* is welfare-maximizing within the HSV class of tax systems.

Second, we would like to explore the robustness with respect to alternative objective functions of our two key findings, first that the best policy in the HSV class delivers most of the feasible welfare gains from tax reform, and second that the best policy in the HSV class is preferred to the best affine schedule. As we will see, these findings extend to a wide range of alternative welfare functions with an intermediate taste for redistribution, but not to objective functions that would dictate either much more or much less redistribution that is currently observed.

Given our fiscal policy parameter estimates and the productivity distribution parameters described, we apply the procedure described in Section 4 to infer the taste for redistribution parameter θ^*. The implied estimate is $\theta^* = -0.566$, indicating that the U.S. social planner wants more redistribution than a laissez-faire planner ($\theta = -1$) but less than a utilitarian one ($\theta = 0$).[27] The relative Pareto weights implied by $\theta^* = -0.566$ are illustrated in Panel A of Figure 5. Pareto weights are increasing in the uninsurable shock α.

The logic for why the model interprets the U.S. planner as having a weaker taste for redistribution than a utilitarian planner is that the U.S. tax and transfer system is not particularly progressive, even though Americans face a lot of uninsurable wage risk. At the same time, the theory implies quantitatively relatively minor roles for the factors that would cut against high progressivity: elastic labor supply and the need to finance public expenditure. As we discussed

---

[27]Moser and de Souza e Silva (2015) adopt our functional form for the social welfare function and estimate the taste for redistribution parameter to be -0.60.
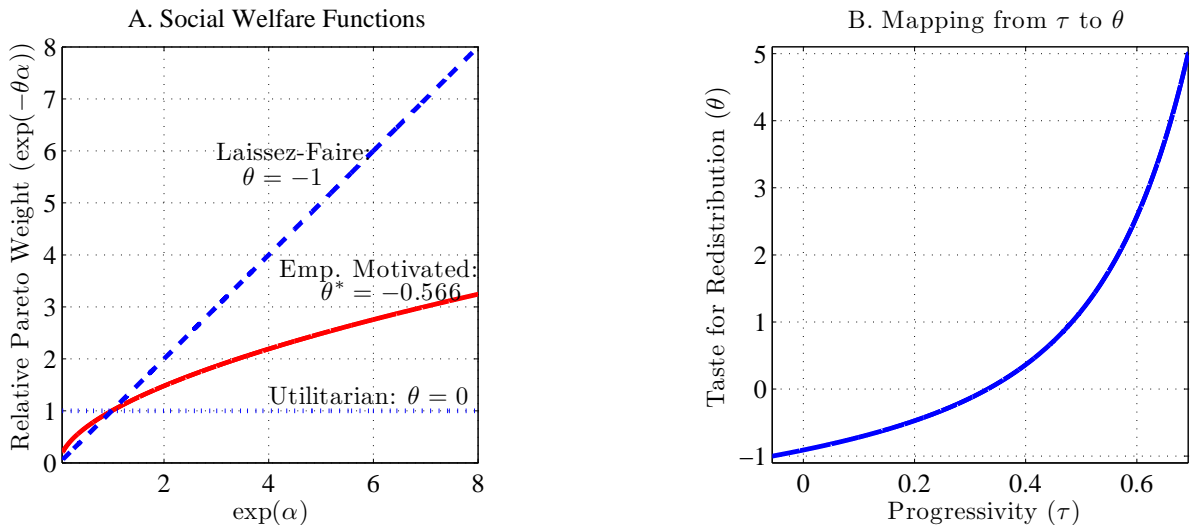
Figure 5: Social welfare functions. Panel A plots our empirically-motivated social welfare function (red solid line) and the utilitarian and laissez-faire alternatives (blue dashed lines). Panel B plots the mapping from τ to θ obtained from the expression in Proposition 1. We use the version of the expression involving G in eq. (28).

in Section 4, a possible political economic interpretation for this weak taste for redistribution is that politicians view high wage workers as more pivotal in elections and put more weight on their preferences in crafting tax policy.

How sensitive is our estimate for θ^* to our estimate for τ^*, the index of progressivity of the U.S. tax system? Panel B of Figure 5 uses Proposition 1 to plot the mapping from progressivity τ to the taste for redistribution θ^*, holding fixed our baseline values for the structural parameters $(\sigma_\alpha^2, \lambda_\alpha, \sigma)$ and for the level of government purchases G. The value for progressivity that would signal a utilitarian ($\theta = 0$) social planner is $\tau^U = 0.332$, which implies an average effective marginal tax rate of 47 percent, much higher than we see in the United States. A laissez-faire social planner ($\theta = -1$) would choose a regressive scheme, with $\tau^{LF} = -0.06$. The actual tax and transfer system in the United States lies in between these two values: $\tau = 0.161$ and the average marginal tax rate is 32 percent. Thus, observed policy appears inconsistent with the U.S. planner having either a utilitarian or a laissez-faire objective.

Alternative Social Welfare Functions   Table 3 shows results for all the social welfare functions we have discussed so far, moving downwards from the weakest to the strongest taste for

Table 3: Alternative Social Welfare Functions

| SWF | Mirrlees Allocations | | | | Welfare Gain ω (%) | | | |
	θ	$\overline{T'}$	Tr/Y	ΔY	HSV*	Affine	Mirrlees	$\frac{\omega(\text{HSV}^{\text{US}},\text{HSV}^*)}{\omega(\text{HSV}^{\text{US}},\text{Mirrlees})}$
Laissez-Faire	-1	0.083	-0.082	9.72	2.98	3.14	3.15	95%
Emp. Motivated	-0.566	0.314	0.051	0.16	$-$	-0.48	0.05	0%
Utilitarian	0	0.491	0.213	-7.99	2.08	1.77	2.48	84%
Rawlsian	∞	0.711	0.538	-22.55	354.9	649.1	708.3	50%

redistribution.[28] The line labelled "Utilitarian" repeats the findings from Table 2. The first set of columns describes some properties of the optimal Mirrlees tax schedule for each social welfare function. The second set of columns describes the welfare gains of moving from the current tax system (HSV with $\tau = 0.161$) to the Mirrlees policy and to the best-in-class affine and HSV policies.

The first takeaway from the table is that the optimal policy prescription is enormously sensitive to the choice for θ. This is worth emphasizing given the explosion of policy research in heterogeneous agent environments. Here, the stronger the planner's desire to redistribute, the higher the marginal tax rates the planner chooses. Moving from the laissez-faire to the Rawlsian social welfare function, the average income-weighted marginal tax rate rises from 8.3 percent to 71.1 percent.[29]

A second takeaway is that the choice of social welfare function has a huge impact on the potential welfare gains from policy reform. Recall that the moving to the optimal tax schedule in the utilitarian case increases welfare by 2.48 percent. If we measure welfare gains using a Rawlsian welfare function as our baseline, we would conclude that tax reform could raise welfare by 708 percent. Given the empirically motivated social welfare function, in contrast, the maximum welfare gain from tax reform is only 0.05 percent! This indicates that the current tax system – more precisely, our HSV approximation to it – must be close to efficient. The small size of the maximum welfare gain from tax reform is perhaps surprising given that the HSV schedule violates some established theoretical properties of optimal tax schedules. In particular, it violates the prescriptions that marginal rates should be everywhere non-negative, and that the rate should be zero at the upper bound of the productivity distribution.

---

[28]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.

[29]Recall that public consumption G is fixed exogenously, and is thus invariant to θ.
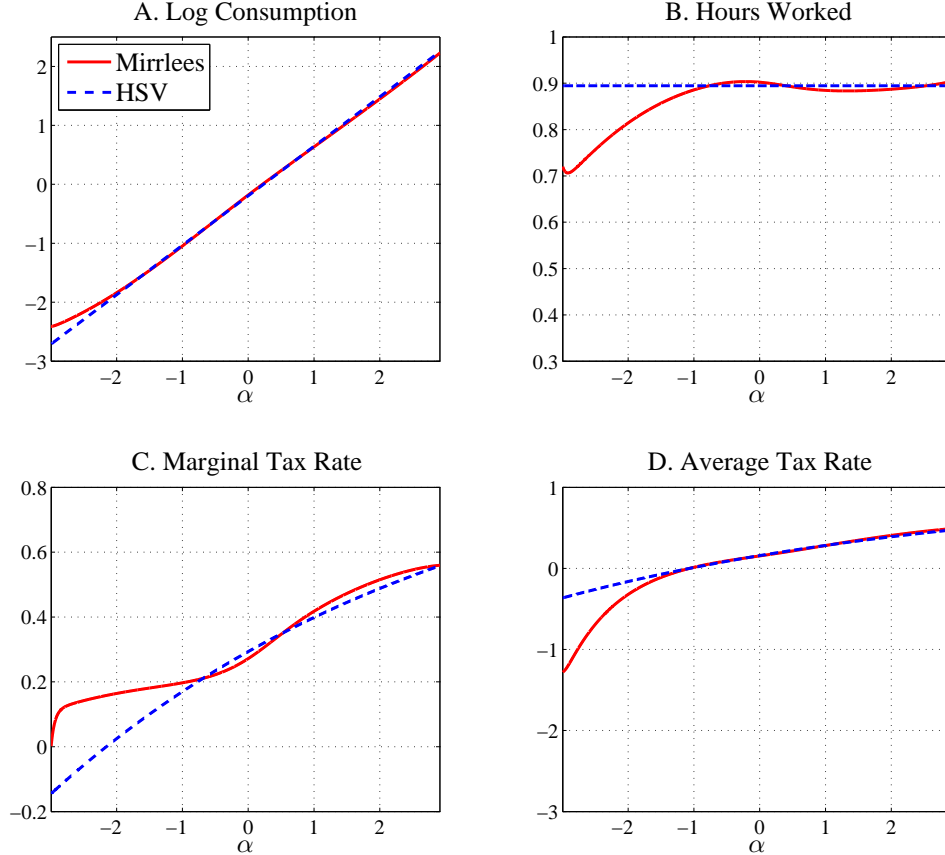
Figure 6: HSV versus Mirrlees tax functions with $\theta = \theta^*$. The figure contrasts allocations and tax rates under the current HSV tax system to those under the Mirrlees policy using our empirically-motivated social welfare function.

Our third takeaway from Table 3 is that assuming an empirically motivated social welfare function does not change our finding from the utilitarian case that the best-in-class HSV function is preferred to the best affine policy. In fact, given $\theta = \theta^*$ moving from the current HSV system to the best possible affine tax scheme reduces welfare by 0.48 percent.

Why, under the empirically motivated social welfare function, are the maximum welfare gains from tax reform so small? Figure 6 plots allocations under the current HSV tax schedule against those under the Mirrlees policy, given $\theta = \theta^*$. It is clear that consumption and hours allocations are very similar across most of the distribution for α under the two schemes, which is consistent with welfare being very similar. While allocations are more different at the extremes of the distribution, the population density in those ranges is very small. We conclude that the fact that the HSV schedule does not satisfy theoretical prescriptions for efficiency at the bounds of the α distribution is quantitatively largely irrelevant.
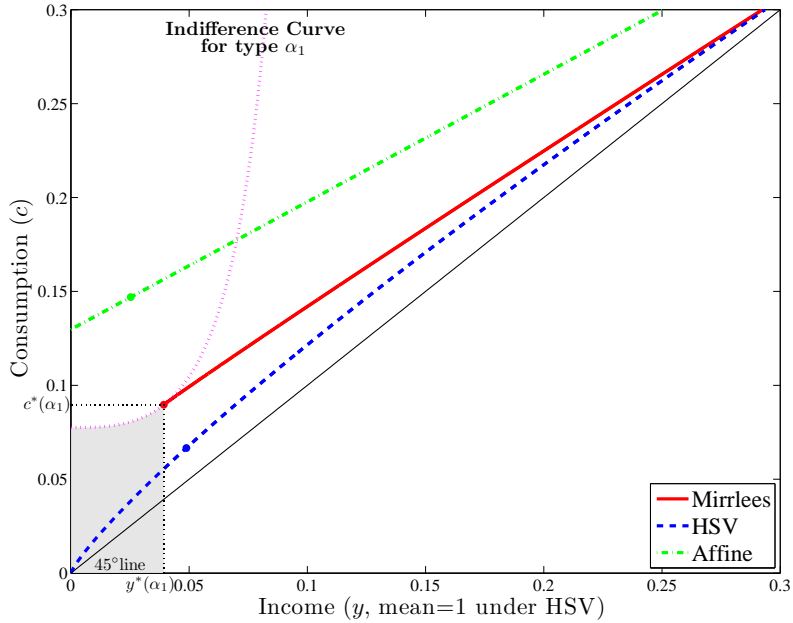
Figure 7: Allocations for low income households. The figure plots household consumption against household income at the bottom of the income distribution under the Mirrlees (red solid), HSV (blue dash), and best-in-class affine (green dot-dash) tax systems. Each tax scheme is best-in-class given the empirically motivated social welfare function.

Figure 7 offers another perspective on the properties of optimal allocations at the bottom end of the income distribution. Here we plot the level of household consumption against the level of household income: net transfers is the difference between the two. We truncate the plot at 30 percent of average income to highlight how the different tax systems treat the poor. The red solid line traces out the budget set associated with constrained efficient allocations. The line stops at the red dot, which corresponds to the level of household income that the planner asks the least productive household to produce, $y^*(\alpha_1)$. As reported in Table 3, this household receives a small net transfer. What does the Mirrlees tax schedule look like for lower income levels? An upper bound on net transfers is given by the indifference curve for the α_1 type that is tangent to the Mirrlees budget set at the point $(y^*(\alpha_1), c^*(\alpha_1))$. Any consumption schedule (and associated net tax schedule) that lies everywhere below this indifference curve will decentralize the Mirrlees solution; the set of possible such schedules is shaded light grey in the figure.

Figure 7 also plots the best income tax schedules in the affine and HSV classes. It is clear from the plot that the HSV schedule is closer than the affine one to the optimal Mirrlees schedule. The
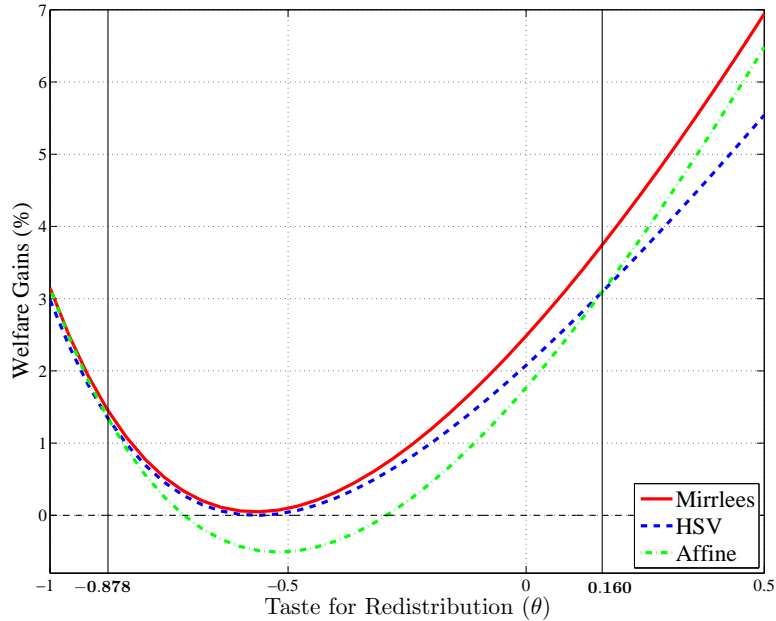
Figure 8: Maximum welfare gains from tax reform. The figure plots the maximum possible gains from tax reform for a range of values for the taste for redistribution parameter θ. Three lines are plotted, corresponding to the best policies in the unrestricted Mirrlees class (red solid), the HSV class (blue dash), and the affine class (green dot-dash).

affine schedule implies net transfers that are much too generous at the bottom of the distribution. Part of the explanation is that the optimal Mirrlees allocation dictates high marginal tax rates at higher income levels, but under an affine scheme, imposing high marginal rates on the rich necessitates high marginal rates across the distribution – and thus large lump-sum transfers. At the same time, transfers to the least productive households are small under the optimal Mirrlees policy. Transfers are optimally small in part because the planner puts relatively low weight on the least productive households, and in part because the fact that a portion of wage dispersion is privately insurable reduces the need for public insurance.[30]

Figure 8 plots welfare gains under alternative tax systems, for a range of values for θ. The red solid line is the welfare gain associated with moving from the current HSV tax system to the optimal Mirrlees scheme, and the blue and green dashed lines are the gains moving from to the best-in-class HSV and affine schemes.

The first message from Figure 8 is that for most intermediate values for θ, the red solid and

---

[30]In Section 6.4 we shut down private insurance and find larger optimal lump-sum transfers.

blue dashed lines are not far apart, indicating that the lion's share of potential welfare gains from tax reform can be achieved by adjusting progressivity while retaining the HSV functional form. For example, the sizable welfare gains from tax reform that are possible under the utilitarian social welfare function ($\theta = 0$) almost entirely reflect the fact that a utilitarian planner wants a more redistributive tax system – and do not signal that the current system redistributes in a very inefficient way.

Second, the optimal HSV scheme outperforms the optimal affine scheme for a wide range of intermediate values for θ between -0.878 and 0.160.

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

6.3 Increasing versus U-Shaped Marginal Rates

In our baseline model specification, optimal marginal tax rates are generally increasing in income, and the optimal scheme can therefore be closely approximated by the HSV functional form. However, in influential quantitative exercises Saez (2001) found a U-shaped pattern for optimal marginal tax rates. His schedules are U-shaped because they feature high marginal rates for low income households, and he writes "the government should apply high rates at the bottom in order to target welfare only to low incomes" (Saez 2001, p.223). We do not find high rates at the bottom to be optimal, and thus we do not find a U-shaped profile for optimal marginal rates to be optimal.

We now show that alternative counter-factual calibrations of our model can give rise to much higher optimal marginal rates at the bottom, and thus to U-shaped optimal marginal rate schedules. We use these alternative calibrations in conjunction with a version of the Diamond-Saez formula for optimal marginal rates in order to help illuminate what determines the optimal level of marginal rates at low income levels.

Figure 9 shows how the optimal allocations and marginal rate profile change as we vary the level of public expenditure that must be financed. The red solid lines are the baseline model, while
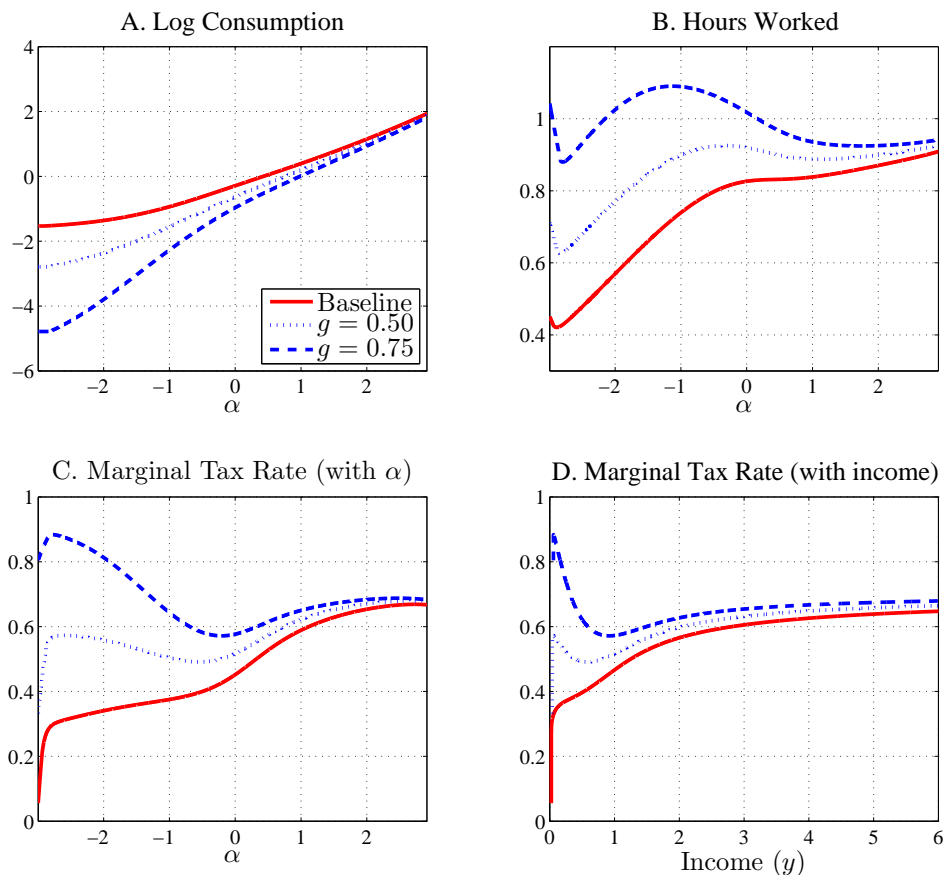
34

Figure 9: Alternative government expenditure levels. The figure plots allocations and tax rates under the optimal Mirrlees tax policy for different values for government purchases. For example, the line labelled $g = 0.5$ corresponds to allocations when G is equal to 50% of GDP in the baseline economy with the HSV tax scheme and $\tau = 0.161$. Panel D is truncated at six times average income.

the blue dashed lines correspond to higher expenditure levels. It is clear that raising the amount of expenditure that must be financed raises optimal marginal tax rates by the most at the bottom of the income distribution, and by much less at the high income levels. The result is that as we increase g, the schedule eventually becomes U-shaped (with the qualification that rates still decline to zero at the very top).

Given our utility function, the Diamond-Saez equation that is widely used to interpret optimal

tax schedules can be written as

$$
\frac{T'(y(\alpha))}{1 - T'(y(\alpha))} = A(\alpha) \times B(\alpha), \tag{21}
$$

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

$$
B(\alpha) = \int_\alpha^\infty \left(1 - \frac{\exp(-s\theta) \cdot C}{c(s)}\right) \frac{c(s)}{c(\alpha)} \frac{dF_\alpha(s)}{1 - F_\alpha(\alpha)}.
$$

Here $f_\alpha(\alpha)$ is the population density at α, and C denotes aggregate (and average) consumption. See Appendix B.5 for the derivation.

The two terms labelled $A(\alpha)$ and $B(\alpha)$ (as in Saez 2001) can be used to discuss the factors that determine the shape of the optimal marginal tax schedule. The first component of the $A(\alpha)$ term, $(1 + \sigma)$, indicates that the more elastic is labor supply, the lower are optimal marginal tax rates, all else equal. Since this term is constant, it plays no role in explaining how optimal marginal tax rates vary with income. The second component of the $A(\alpha)$ term is the ratio of fraction of households more productive than α relative to the density at α. Marginal rates should be high in regions of the productivity distribution where this ratio is high, so that there are lots of more productive agents who will pay extra taxes, but relatively few whose labor supply will be directly distorted by higher rates at the margin. Like the elasticity term, this term involves only structural primitives of the model.

The $B(\alpha)$ term is more involved, and involves the distribution of consumption, which is obviously endogenous to the tax system. However, this term is critical for interpreting the optimal tax schedule, since the $A(\alpha)$ term is identical across the different marginal tax schedules corresponding to different values for government purchases plotted in Panels C and D of Figure 9.

To dig deeper, we adopt our baseline social welfare function by setting $\theta = 0$ in eq. (21) and take an approximation to the integral term that is accurate for small α.[31] Then, for small α, we get

$$
\frac{T'(y(\alpha))}{1 - T'(y(\alpha))} \simeq (1 + \sigma) \times \frac{F_\alpha(\alpha)}{f_\alpha(\alpha)} \times \frac{C - c(\alpha)}{c(\alpha)}. \tag{22}
$$

To derive a simple expression for the slope of optimal marginal tax rates we differentiate the log of expression (21) and again set $\theta = 0$ and approximate the integral term. Then, for small α,

---

[31]Table 6 in Appendix B.5 reports the quality of this approximation at different points in the α distribution.

the slope of the marginal tax rate ratio is approximately given by

$$\frac{d}{d\alpha} \log\left(\frac{T'(y(\alpha))}{1 - T'(y(\alpha))}\right) \simeq -\frac{f'_\alpha(\alpha)}{f_\alpha(\alpha)} + \frac{f_\alpha(\alpha)}{F_\alpha(\alpha)} - \frac{c'(\alpha)}{c(\alpha)}. \tag{23}$$

We can use these two expressions to shed light on the level and slope of marginal tax rates at the bottom of the α distribution. It is clear from eq. (22) that marginal rates are lower, all else equal, the closer is consumption for a household with productivity level α to mean consumption. From eq. (23), the more slowly consumption increases with productivity (the smaller is $c'(\alpha)$), the more likely it is that optimal marginal rates are increasing in income. Looking at Panel A of Figure 9, it is clear that when government purchases are small, consumption at low values for α is relatively close to the mean, and the consumption schedule is relatively flat. Equations (22) and (23) make clear that these properties are consistent with marginal tax rates that are low but increasing with productivity. Conversely, when government purchases are high, the government achieves much less consumption compression, and marginal rates at the bottom are therefore high and declining with productivity.

While this discussion illustrates that the Diamond-Saez equation (21) and the plots in Figure 9 are mutually consistent, it does not quite get to the bottom of why the optimal consumption allocation looks the way it does. Our intuition here is that for low values for government purchases, raising revenue to redistribute is relatively easy, and the planner therefore delivers large lump-sum transfers to the poor. Given these transfers, the consumption level of the poor is not too far below average economy-wide consumption ($C - c(\alpha)$ in eq. (22) is small). In addition, given large lump-sum transfers and a utility function that features wealth effects, low productivity workers choose to work relatively few hours, and thus consumption increases only modestly with productivity ($c'(\alpha)$ is small in eq. 23).

Conversely, with high government purchases, high tax rates are required just to finance government purchases, and generous redistribution towards the poor would require very high marginal rates on middle and upper income households. Because such rates would be very distortionary, the planner chooses only modest lump-sum transfers. Thus, the least productive households have much lower consumption (implying, via eq. 22 high marginal rates). At the same time, low transfers also imply that the least productive households work long hours, and thus that consumption

increases rapidly with productivity. This implies, via eq. (23), that marginal tax rates decline with productivity.

Note that the presence of wealth effects on labor supply plays an important role in this discussion. If we had alternatively assumed a utility function with no wealth effects, as is quite common in the public finance literature (e.g., Diamond 1998), the differences in tax schedules across the high and low government expenditure parameterizations would be much smaller, and the tax schedule would more closely resemble the U-shaped one corresponding to high values for g.[32]

We take away several messages from this analysis. First, while the Diamond-Saez equation (21) is useful, it offers only an incomplete intuition for the shape of the optimal tax schedule at the bottom. Second, the U-shaped profile for marginal rates emphasized by Saez (2001) is not a general feature of an optimal tax system. Third, the common notion that marginal rates should be high at the bottom to rapidly tax away transfers intended only for the very poor is wrong: to the contrary, the larger such transfers are, the lower should be marginal tax rates at the bottom. Finally, if the 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. Raising rates at the top is not attractive, because the planner is already near the top of the Laffer curve in terms of how much revenue it can squeeze out of the very rich. Raising marginal rates at the bottom, on the other hand, might reduce net revenue from the least productive households, but will increase revenue from all more productive households.

We next consider a different set of experiments. In particular, we keep government purchases fixed at their baseline value, and consider how the optimal tax schedule changes as we vary the taste for redistribution θ. Figure 10 shows the results. The key finding here is that as we increase the taste for redistribution, the optimal marginal tax schedule moves from being upward sloping to becoming U-shaped. In this sense, a high taste for redistribution implies a similar marginal tax schedule to a high value for government purchases. However, the corresponding allocations in the two cases are quite different. With a high taste for redistribution, the government uses high marginal tax rates to finance very large lump-sum transfers, which – thanks to wealth effects –

---

[32]There is no upper bound to government purchases in our economy, because the utility function we use does not impose an upper bound on labor supply. The way to raise high revenue is to give zero (or negative) lump-sum transfers and to impose marginal tax rates of 100% up to a high income threshold. All households will then be willing to work up to the point that their income exceeds the threshold, thereby guaranteeing strictly positive consumption and bounded utility.
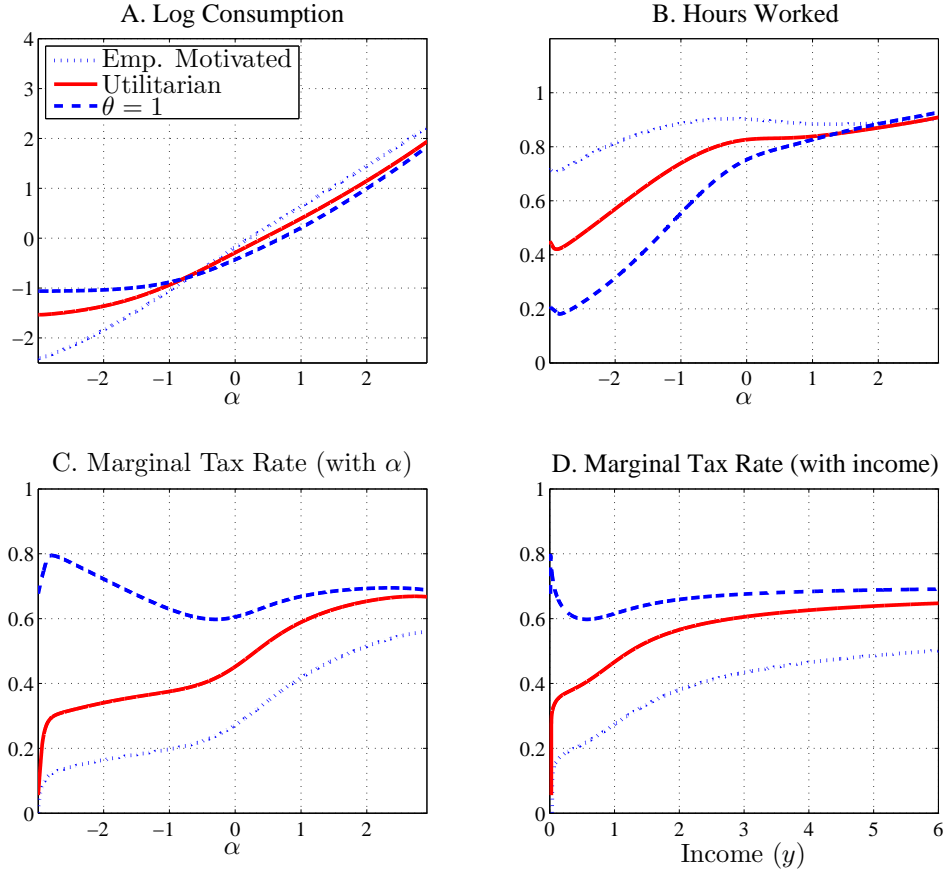
38

Figure 10: Alternative tastes for redistribution. The figure contrasts allocations and tax rates under the optimal Mirrlees tax policy for different values for the taste for redistribution parameter θ. Panel D is truncated at six times average income.

imply that the least productive households work very few hours while enjoying relatively healthy consumption. Still, at a basic level the intuition for the tax schedules in the high G and high θ cases are very similar. In both cases, the planner wants to raise a lot of revenue (for purchases in the first case, and lump-sum transfers in the second). In the $g = 0.75$ case, government purchases plus transfers are 63 percent of GDP, while in the $\theta = 1$ case the corresponding number is 72 percent. In both cases, the best way to raise this revenue is to have high marginal rates at the bottom. Higher rates in the middle of distribution are not desirable, because the density of households located there is high (implying a small value for the $A(\alpha)$ term in eq. 21). Higher rates at the top are unattractive because the rich are already close to the top of the Laffer curve.

Table 4: Optimal Tax and Transfer System with No Insurable Shocks

Tax System	Tax Parameters		Outcomes			
			ω (%)	ΔY (%)	$\overline{T'}$	Tr/Y
HSV^{US}	$\lambda : 0.842$	$\tau : 0.161$	$-$	$-$	0.319	0.019
HSV*	$\lambda : 0.804$	$\tau : 0.383$	4.17	-9.72	0.511	0.084
Affine	$\tau_0 : -0.283$	$\tau_1 : 0.545$	5.34	-10.45	0.545	0.326
Mirrlees			5.74	-10.64	0.550	0.284

6.4 Role of Private Insurance

We now explore how the existence of private insurance against a subset of idiosyncratic shocks affects the welfare implications of tax reform.

Table 4 presents results when we shut down private insurance by setting $\sigma_\varepsilon^2 = 0$ and increasing the variance of α_N, the normally distributed uninsurable component, so as to leave the total variance of log wages unchanged. All other parameter values are set to their values in the baseline calibration. Thus, the planner is utilitarian ($\theta = 0$).

Since the dispersion of uninsurable shocks is larger, there is now a larger role for public redistribution. Thus, the best tax policy in the HSV class is now more progressive than in the baseline parameterization: $\tau^* = 0.383$ instead of 0.330. The second-best policy now features an average marginal tax rate of 55 percent and larger transfers. The maximal welfare gains from tax reform are now more than twice as large as in the baseline model and are associated with an output decline of 10.6 percent.

One important difference relative to the baseline parameterization is that when all risk is privately uninsurable, an affine tax system is preferred to the best policy in the HSV class. One reason is that when the density of very low productivity workers is larger, the Mirrlees planner wants larger lump-sum transfers, which the affine scheme can accommodate. We conclude that to accurately characterize the qualitative nature of optimal taxation it is essential to explicitly account for the existence of private insurance.[33]

---

[33]If we were to instead adopt the empirically motivated value for θ at the same time as assuming zero private insurance, then the best policy in the HSV class would be preferred to the best policy in the affine class. Detailed results are available upon request.

6.5 Additional Sensitivity

We first counter-factually assume a log-normal wage distribution. Specifically, we assume that α is normally rather than EMG distributed, and adjust the variance so that the variance of α is identical to the baseline case. Relative to the baseline, the distribution for the uninsurable component of wages now has a much thinner right tail and a heavier left tail. We hold fixed all other parameter values and set $\theta = 0$.

The best HSV-class tax function now features less progressivity than under the baseline calibration ($\tau = 0.285$ versus 0.330). The second-best policy implies a lower average marginal tax rate of 44 percent, but also features larger net transfers for the least productive households. A key result is that the best affine tax and transfer system now dominates the best system in the HSV class and very closely approximates the second-best allocation. Thus, assuming a log-normal distribution for wages resurrects the original conclusion of Mirrlees (1971), namely, that the optimal nonlinear income tax is approximately linear. We conclude, like Saez (2001), that it is essential to carefully model the empirical productivity distribution for the purposes of providing quantitative guidance on the design of the tax and transfer system. See Appendix B.6 for more details.

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

6.6 Extensions to Richer Tax Structures

We now explore richer tax structures. First we consider polynomial tax functions that add quadratic and cubic terms to the affine functional form. Next we consider an economy in which there is a third component of idiosyncratic productivity, κ, that is privately uninsurable but observed by the planner. We find that the potential welfare gains from indexing taxes to κ are as large as 1.5 percent of consumption.[34]

---

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

6.6.1 Polynomial Tax Functions

In the baseline model we have learned that for welfare it is more important that marginal tax rates increase with income - which the affine scheme rules out – than that the government provides universal lump-sum transfers. Relative to the affine case, we now ask how much better the Ramsey planner can do if we introduce quadratic and cubic terms in the net tax function, thereby allowing marginal tax rates to increase with income.

Let $T_n(y)$ denote an n-th order polynomial tax function: $T_n(y) = \tau_0 + \tau_1 y + \cdots + \tau_n y^n$. We assume that the marginal tax rate becomes constant above an income threshold $\bar{y}$ equal to 10 times average income in the baseline HSV-tax economy. We focus on the cases $n = 2$ and $n = 3$ (i.e., quadratic and cubic tax functions). Table 8 and Figure 12 in Appendix B.7 contain detailed results, which we summarize here.

As we give the Ramsey planner access to increasingly flexible tax functions, outcomes and welfare converge to the Mirrlees solution, which is reassuring.[35] With the best quadratic function, marginal tax rates are increasing in income ($\tau_2 > 0$) – the key property of the optimal tax schedule – and lump-sum transfers are smaller than under the best affine scheme. Under the best cubic system, lump-sum transfers are reduced still further and the cubic coefficient τ_3 is negative, which ensures that marginal rates flatten out before income reaches the threshold $\bar{y}$. Because marginal and average tax rates under the best cubic policy are generally very similar to those implied by the Mirrlees solution, allocations are close to being constrained efficient. Thus, moving to the best cubic policy generates about 97 percent of the maximum potential welfare gains from tax reform.

6.6.2 Type-Contingent Taxes

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

We set the variance of this observable fixed effect, σ_κ^2, equal to the variance of wage dispersion

---

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

that can be accounted for by standard observables in a Mincer regression. Heathcote et al. (2010) estimate the variance of cross-sectional wage dispersion attributable to observables to be $\sigma^2_\kappa = 0.108$. For the sake of simplicity, we assume a two-point equal-weight distribution for κ. This gives $\exp(\kappa_{High})/\exp(\kappa_{Low}) = 1.93$.

The total variance of the privately uninsurable component of wages is unchanged relative to the baseline model, but we now attribute part of this variance to κ. The three parameters $\mu_\alpha, \sigma^2_\alpha$, and λ_α characterizing the EMG distribution for α are therefore recalibrated so that (i) the variance of (discretized) α is equal to 0.239 (i.e., $0.488 - \sigma^2_\varepsilon - \sigma^2_\kappa$), (ii) $\sum_i \pi_i \exp(\alpha_i) = 1$, and (iii) the value of the shape parameter $\sigma_\alpha \lambda_\alpha$ is the same as that in the baseline model (i.e., 0.827).[36]

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

Table 9 in Appendix B.7 describes optimal type-contingent tax functions and the associated outcomes. We find that if the planner can condition taxes on the observable component of labor productivity, it can generate large welfare gains relative to the current tax system, which does not discriminate by type. The maximum welfare gain is now 6.54 percent of consumption, compared with 2.48 percent in the baseline analysis. This large welfare gain arises in part because the average effective marginal tax rate drops to 42 percent, which translates into a smaller output loss. Recall that if productivity were entirely observable, the planner could implement the first best, with a zero marginal rate for all households.

If the Ramsey planner is allowed to impose a different tax schedule on each κ type, he can achieve welfare gains that nearly match those for the Mirrlees planner. Under the best affine system, the high κ type faces a double whammy, paying higher marginal tax rates than the low type and receiving only tiny transfers.

One important caveat to this analysis is that we have treated all the variation in κ as exogenous and have therefore ignored potential feedback from the tax system to the distribution for κ. However, an education-dependent tax system would likely affect agents' educational decisions (see,

---

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

Table 5: Grid points

# of grid points I	ω (%, relative to HSV)		
	Affine	Mirrlees	First Best
10	1.75	20.90	47.04
100	1.77	4.84	45.43
1,000	1.77	2.70	45.35
10,000	1.77	2.48	45.34
100,000	1.77	2.46	45.34

e.g., Heathcote et al. 2016). In particular, relatively high taxation of high κ households would discourage education investment. Thus, we regard our 6.54 percent welfare gain as an upper bound on the feasible welfare gains from tagging.

6.7 Coarser Grids

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

To make this point, in Table 5 we report welfare gains from tax reform (relative to the HSV baseline) as we increase the number of grid points from $I = 10$ to $I = 100,000$. Reassuringly, the number of grid points does not affect the results for the affine case or for the first best. However, as the number of grid points decreases, welfare gains for the Mirrlees planner increase substantially. For $I = 10$ these gains are 20.9 percent of consumption, compared with 2.48 percent with $I = 10,000$.

The intuition behind this result is that with a coarse grid, ensuring truthful reporting becomes easier for the Mirrlees planner. Consider a grid of size I. A common result is that only local downward incentive constraints bind at the solution to the planner's problem. Now suppose that we remove every other point from the grid, leaving all else unchanged. At the original conjectured solution, none of the incentive constraints are now binding. Thus, the planner can adjust allocations to compress the distribution of consumption or to strengthen the correlation between productivity and hours worked.

In our static model, using a very fine grid is not too costly from a computational standpoint. In dynamic Mirrleesian settings, however, the presence of additional state variables typically necessitates a very coarse discretization of types, often using fewer than 50 points. The findings in Table 5

cast some doubt on the quantitative robustness of such analyses if the true underlying productivity distribution is continuous.

7   Conclusions

The goal of this paper has been to characterize the optimal income tax schedule and to explore how closely it can be approximated by simple parametric functions. We found that optimal marginal tax rates are generally increasing in income, and are neither flat nor U-shaped. We also found that the optimal policy can be well approximated by a simple two parameter function. Our hope is that this result will help strengthen the bridge between the extensive theoretical literature on optimal taxation and the practical policy debate.

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

First, careful thought should be given to the specification of the planner's social welfare function, since this has an enormous impact on policy prescriptions. We have proposed a functional form for social welfare 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. Our approach could easily be generalized to construct social welfare functions in much richer models.

Second, it is important to recognize the existence of private insurance, if one views the role of the tax and transfer system as being limited to offering a degree of public insurance against risks that cannot be insured privately. In our environment, introducing privately-insurable risk has an important quantitative impact on the shape of the optimal tax schedule.

Third, although the fully optimal profile for marginal rates is quite complicated, it can be very well approximated by a simple two parameter power function of the form used by Benabou (2000) and Heathcote et al. (2016) among others. Thus, in terms of welfare, a simple parametric Ramsey-style policy that can be easily communicated to policymakers comes very close to replicating the constrained efficient Mirrlees allocation.

Fourth, our analysis indicates that an important characteristic of the optimal tax system is that marginal tax rates generally increase with income. Switching to a flat tax system that redistributes

via lump-sum transfers is likely to be welfare reducing.

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

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

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

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

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

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

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

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

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

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

FRIEDMAN, M. (1962): *Capitalism and Freedom*, Chicago: University of Chicago Press.

GOLOSOV, M., M. TROSHKIN, AND A. TSYVINSKI (forthcoming): "Redistribution and Social Insurance," *American Economic Review*.

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

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

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

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

——— (2016): "Optimal Tax Progressivity: An Analytical Framework," mimeo.

Heathcote, J. and H. Tsujiyama (2016): "An Analytical Approach to Estimating Social Preferences," mimeo.

Hendren, N. (2014): "The Inequality Deflator: Interpersonal Comparisons without a Social Welfare Function," Working Paper 20351, National Bureau of Economic Research.

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

Lockwood, B. B. and M. Weinzierl (forthcoming): "Positive and Normative Judgments Implicit in U.S. Tax Policy, and the Costs of Unequal Growth and Recessions," *Journal of Monetary Economics*.

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

Mankiw, N. G., M. Weinzierl, and D. Yagan (2009): "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, 23, 147–74.

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

Moser, C. and P. O. de Souza e Silva (2015): "Paternalism vs Redistribution: Designing Retirement Savings Policies with Behavioral Agents," mimeo.

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

Persson, T. and G. Tabellini (2000): *Political Economics: Explaining Economic Policy*, Cambridge, MA: MIT Press.

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

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

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

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

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

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

A   Appendix: Proof of Proposition 1

We provide the proof of Proposition 1.

Given the HSV tax function (17), decision rules as a function of τ are as follows:

$$c(\alpha; \lambda, \tau) = \lambda(1 - \tau)^{\frac{1-\tau}{1+\sigma}} \exp\left[(1 - \tau)\alpha\right] \exp\left(\frac{1 - \tau}{\sigma} \frac{\sigma_\varepsilon^2}{2}\right), \tag{24}$$

$$h(\varepsilon; \tau) = (1 - \tau)^{\frac{1}{1+\sigma}} \exp\left(\frac{-1}{\sigma^2} \frac{\sigma_\varepsilon^2}{2}\right) \exp\left(\frac{\varepsilon}{\sigma}\right). \tag{25}$$

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

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

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

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

and government expenditure is given by

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

Substituting eqs. (24) and (25) into these, the optimization problem can be rewritten as

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

where

$$G = g(1 - \tau)^{\frac{1}{1+\sigma}} \exp\left(\frac{1}{\sigma} \frac{\sigma_\varepsilon^2}{2}\right). \tag{26}$$

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

Given the social welfare function (16), the optimization problem becomes[37]

---

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

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

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

$$+ \log\left[(1-\tau)^{\frac{1}{1+\sigma}}\exp\left(\frac{1}{\sigma}\frac{\sigma_\varepsilon^2}{2}\right) - G\right] - \frac{1-\tau}{1+\sigma}. \tag{27}$$

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

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

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

Lemma 2 *Assume the support of θ is compact, $[\underline{\theta},\bar{\theta}]$. Then the integral and the derivative of V are interchangeable; that is,*

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

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

(i) is obvious since f_α is continuous.

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

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

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

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

Applying this lemma, we get

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

Substituting this expression into eq. (27), the optimization problem becomes

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

The first-order condition with respect to τ is

$$
0 = -\frac{1}{\lambda_\alpha+\theta}+\sigma_\alpha^2\theta+\frac{1}{1+\sigma}+\frac{1}{\lambda_\alpha-1+\tau}+\sigma_\alpha^2(1-\tau)-\frac{\left[1-\frac{G}{\exp\left(\frac{1}{\sigma}\frac{\sigma_\varepsilon^2}{2}\right)(1-\tau)^{\frac{1}{1+\sigma}}}\right]^{-1}}{(1-\tau)(1+\sigma)}. \tag{28}
$$

Substituting eq. (26) into this, we have

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

Therefore, the planner's weight θ^* must solve eq. (18). $\mathcal{Q.E.D.}$

B Appendix for Online Publication

B.1 Insurance via Family versus Insurance via Financial Markets

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

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

$$\int B(\alpha,\varepsilon)Q(\varepsilon)d\varepsilon = 0, \tag{29}$$

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

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

$$c(\alpha,\varepsilon) = y(\alpha,\varepsilon) - T\left(y(\alpha,\varepsilon)\right) \qquad \text{for all } \varepsilon, \tag{30}$$

where individual income before taxes and transfers is given by

$$y(\alpha,\varepsilon) = \exp(\alpha+\varepsilon)h(\alpha,\varepsilon) + B(\alpha,\varepsilon) \qquad \text{for all } \varepsilon. \tag{31}$$

The individual agent's problem is then to choose $c(\alpha,\cdot)$, $h(\alpha,\cdot)$, and $B(\alpha,\cdot)$ to maximize expected utility (4) subject to eqs. (29), (30), and (31). The equilibrium definition in this case is similar to that for the specification in which insurance takes place within the family.

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

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

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

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

subject to

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

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

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

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

The first-order condition for hours is

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

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

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

Now note that this expression is exactly the same as the one for the family planner decentralization (eq. (11)). Moreover, in both cases $c(\alpha,\varepsilon) = c(\tilde{\alpha})$. It follows that for any values for $(\alpha,\tilde{\alpha})$, expected utility for the agent in this decentralization with private insurance markets is identical to welfare for the family head in the decentralization with insurance within the family. Thus, the set of allocations that are incentive compatible when the social planner interacts with the family head are the same as those that are incentive compatible when the planner interacts agent by agent. It follows that the solution to the social planner's problem is the same under both models of ε insurance. Similarly, the income tax schedule that decentralizes the Mirrlees solution is also the same under both models of ε insurance, and marginal tax rates are given in both cases by eq. (15). Note that marginal tax rates do not vary with ε under either insurance model because income (including insurance payouts / family transfers) does not vary with ε.[38]

Finally, note that if insurance against ε is achieved via decentralized financial markets, the planner could conceivably ask agents to report ε after the ε shock is drawn and offer allocations for consumption $c(\tilde{\alpha}, \tilde{\varepsilon})$ and income $y(\tilde{\alpha}, \tilde{\varepsilon})$ indexed to reports of both α and ε. With decentralized

---

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

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

insurance, the planner might be able to offer contracts that separate agents with different values for ε (recall that under the family model for insurance, this was not possible). One might think there would be no possible welfare gain to doing so, since private insurance already appears to deliver an efficient allocation of hours and consumption within any group of agents sharing the same α. However, it is possible that by inducing agents to sacrifice perfect insurance with respect to ε, the planner can potentially loosen incentive constraints and thereby provide better insurance with respect to α.[39] We plan to explore this issue in future work. For now, we simply focus on the problem in which the planner offers contracts contingent only on α, which is the natural benchmark under our baseline interpretation that the family is the source of insurance against shocks to ε.

B.2 Individual- versus Family-Level Taxation

Our baseline model specification assumes that the planner only observes – and thus can only tax – total family income. However, taxing income at the individual level would have no impact on allocations. We now prove that if the tax function for individual income satisfies condition (9), then equilibrium consumption and income are independent of ε, as in the version when taxes apply to total family income.

Proposition 3 *If the tax schedule satisfies condition (9), then the solution to the family head's problem is the same irrespective of whether taxes apply at the family level or the individual level.*

Proof. We will show that given condition (9), the FOCs for the family head with individual-level taxation are identical to those with family-level taxation, namely, eqs. (7) and (8).

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

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

subject to

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

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

The FOCs are

$$
[y(\alpha,\varepsilon) - T(y(\alpha,\varepsilon))]^{-\gamma} [1 - T'(y(\alpha,\varepsilon))] = \mu(\alpha), \tag{32}
$$

$$
h(\alpha,\varepsilon)^\sigma = \mu(\alpha) \exp(\alpha + \varepsilon), \tag{33}
$$

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

---

[39] When we introduce publicly observable (but privately uninsurable) differences in productivity, we see that constrained efficient allocations typically have the property that agents with the same unobservable component α but different observable components of productivity κ are allocated different consumption (see Section 6.6.2).

If the tax schedule satisfies condition (9) (the condition that guarantees first-order conditions are sufficient for optimality) then we can show that optimal consumption and income are independent of ε, as in the version when taxes apply to total family income.

In particular, differentiate both sides of FOC (32) with respect to ε. The right-hand side is independent of the insurable shock ε, and hence its derivative with respect to ε is zero. The derivative of the left-hand side of this equation with respect to ε is, by the chain rule,

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

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

The first term is nonzero by condition (9), which immediately implies that $\frac{\partial y}{\partial \varepsilon} = 0$. Therefore, pre-tax income is independent of ε, and hence consumption is also independent of ε. Thus, the FOCs (32) and (33) combine to deliver exactly the original intratemporal FOC with family-level taxation, namely, eq. (8). $\mathcal{Q.E.D.}$

B.3 Comparative Statics for θ^*

Comparative statics with respect to τ : The values for τ that signal a laissez-faire or a utilitarian planner, which we denote τ^{LF} and τ^{U} respectively, can be derived from eq. (18) by substituting in, respectively, $\theta^* = -1$ and $\theta^* = 0$ and solving for the relevant root. Equation (18) is a cubic equation in τ, and the closed-form expressions for τ that correspond to these two baseline welfare functions are somewhat involved. The normal distribution case ($\lambda_\alpha \to \infty$) is simpler, since (19) is quadratic in τ. In that case, τ^{LF} and τ^{U} are, respectively,

$$\tau^{LF} = 1 + \frac{1 - (1+\sigma)\sigma_\alpha^2 - \sqrt{1 + (1+\sigma)^2 (\sigma_\alpha^2)^2 - 2(1+\sigma)\sigma_\alpha^2 + 4\frac{1+\sigma}{1-g}\sigma_\alpha^2}}{2(1+\sigma)\sigma_\alpha^2}, \tag{34}$$

$$\tau^{U} = 1 + \frac{1 - \sqrt{1 + 4\frac{1+\sigma}{1-g}\sigma_\alpha^2}}{2(1+\sigma)\sigma_\alpha^2}. \tag{35}$$

Note that $\tau^{U} > \tau^{LF}$, as expected. It is straightforward to verify that when $g = 0$, $\tau^{LF} = 0$. The same result also extends to the general EMG distribution for α, as can be readily verified from eq. (18).

The clearest signal of a Rawlsian welfare objective ($\theta^* \to \infty$) is a ratio of expenditure to output $g = G/Y(\tau)$ approaching one.[40] This signals that the planner has pushed τ to the maximum value that still allows the economy to finance required expenditure G. This limiting value for τ is

$$\tau^{R} = 1 - G^{1+\sigma} \exp\left(-\frac{1+\sigma}{\sigma} \frac{\sigma_\varepsilon^2}{2} \right). \tag{36}$$

---

[40]This implies that aggregate consumption and thus consumption of every agent must also converge to zero.

Note that with $G = 0$, $\tau^R = 1$, but for $G > 0$, $\tau^R < 1$. Indeed, if $G \geq \exp\left(\frac{\sigma_{\tilde{\varepsilon}}^2}{2\sigma}\right)$, then $\tau^R \leq 0$, since only a regressive scheme induces sufficient labor effort to finance expenditure. The Rawlsian planner pushes progressivity toward the maximum feasible level because under any less progressive system, households with sufficiently low uninsurable productivity α would gain from increasing progressivity. This result hinges on the distribution for α being unbounded below. The only component of welfare that varies with α (given utility that is logarithmic in consumption and the HSV tax schedule) is log consumption, which contains a term $(1 - \tau)\alpha$ (see eq. (24) in Appendix A). A Rawlsian planner's desire to minimize the negative welfare contribution of this term for unboundedly low α households leads it to choose the maximum feasible value for τ. If instead there was a lower bound α_1 in the productivity distribution (as in the numerical example we shall consider later), the Rawlsian planner would stop short of pushing progressivity to the maximum feasible level. The same would be true for any planner with a finite value for θ.

Comparative statics with respect to g : Now consider the comparative statics with respect to the observed ratio g. The implied taste for redistribution θ^* is increasing in g. Thus, if we saw two economies that shared the same progressivity parameter τ (and the same wage distribution), but one economy devoted a larger share of output to public expenditure, we would infer that the planner in the high spending country must have a stronger taste for redistribution. The logic is that tax progressivity reduces labor supply, making it more difficult to finance public spending. Thus, governments with high revenue requirements will tend to choose a less progressive system – unless they have a strong desire to redistribute.

A corollary of this comparative static is that the larger is g, the smaller are the values τ^{LF} and τ^U consistent with a planner being either laissez-faire or utilitarian (see (34) and (35)). Similarly, the larger is G, the smaller is the value τ^R consistent with a Rawlsian objective (see (36)).

Comparative statics with respect to σ_α^2 : Comparative statics with respect to the variance of uninsurable shocks σ_α^2 are straightforward. The parameter θ^* is decreasing in σ_α^2. Thus, more uninsurable risk (holding fixed tax progressivity) means we can infer the planner has less desire to redistribute.

Comparative statics with respect to σ : The implied redistribution preference parameter θ^* is decreasing in σ, meaning that the less elastic is labor supply (and thus the smaller the distortions associated with progressive taxation), the less desire to redistribute we should attribute to the planner. Consider the limit in which labor supply is inelastic $\sigma \to \infty$. Then output is independent of τ, and we get $\theta^* = \tau - 1$. Thus, in this case a utilitarian planner ($\theta^* = 0$) would set $\tau = 1$, thereby ensuring that all households receive the same after-tax income. A planner with a higher θ^* would actually choose $\tau > 1$, implying an inverse relationship between income before taxes and income after taxes.

With elastic labor supply, one would never observe $\tau \geq 1$, since in the limit as $\tau \to 1$, labor supply drops to zero (given that, at $\tau = 1$, all households receive after-tax income equal to λ, irrespective of pre-tax income).

Comparative statics with respect to λ_α : Finally, θ^* is increasing in λ_α, holding fixed the total variance of the uninsurable component (namely, $\sigma_\alpha^2 + \lambda_\alpha^{-2}$). Thus, if two economies were identical except that one had a more right-skewed distribution for α (a smaller λ_α), one would infer that the heavier right tail economy must have a weaker taste for redistribution. The mirror image of this finding is that a heavier right tail in the distribution for α implies higher optimal progressivity (holding fixed θ).

B.4 Computational Method

We briefly now describe how we compute the optimal allocation in the baseline economy. We solve the Mirrlees planner's problem (12) for our discretized economy numerically. We first note that the local downward and local upward incentive compatibility constraints are necessary and sufficient for the global incentive compatibility constraints (14) to be satisfied:

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

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

We then solve for the allocation exactly at each grid point. Specifically, we use forward iteration (forward from α_1 to α_I) to search for an allocation that satisfies all the first-order conditions, the incentive constraints above, and the resource constraint (13). Finally, we confirm that before-tax income is nondecreasing in wages, concluding that the resulting allocation is optimal given that our utility function exhibits the single-crossing property. Note that we never assume that the upward incentive constraints are slack, because their slackness is not guaranteed for any economy with $I > 2$. In our baseline economy, some upward incentive constraints are indeed binding at the bottom of the α distribution, which results in bunching.

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

B.5   Derivation of the Diamond-Saez Formula and Approximation

Derivation of the Diamond-Saez Formula   Reproducing the Mirrlees planner's problem from eqs. (12-14), we have

$$
\begin{cases}
\displaystyle \max_{\{c(\alpha),y(\alpha)\}} \quad \int W(\alpha) \left[\frac{c(\alpha)^{1-\gamma}}{1-\gamma} - \frac{\Omega}{1+\sigma} \left(\frac{y(\alpha)}{\exp(\alpha)} \right)^{1+\sigma} \right] dF_\alpha(\alpha) \\[2mm]
\text{s.t.} \quad \frac{c(\alpha)^{1-\gamma}}{1-\gamma} - \frac{\Omega}{1+\sigma} \left(\frac{y(\alpha)}{\exp(\alpha)} \right)^{1+\sigma} \geq \frac{c(\tilde\alpha)^{1-\gamma}}{1-\gamma} - \frac{\Omega}{1+\sigma} \left(\frac{y(\tilde\alpha)}{\exp(\alpha)} \right)^{1+\sigma} \quad \text{for all } \alpha \text{ and } \tilde\alpha, \\[2mm]
\int [y(\alpha) - c(\alpha)] \, dF_\alpha(\alpha) - G \geq 0.
\end{cases}
$$

The IC constraints state

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

Using the envelope condition:

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

we get

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

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

$$
\begin{cases}
\displaystyle \max_{\{U(\alpha),y(\alpha)\}} \quad \int W(\alpha) U(\alpha) dF_\alpha(\alpha) \\[2mm]
\text{s.t.} \quad U'(\alpha) = \frac{\Omega}{\exp([(1+\sigma)\alpha]} y(\alpha)^{1+\sigma} \quad \text{for all } \alpha, \\[2mm]
\int [y(\alpha) - c(\alpha; U, y)] \, dF_\alpha(\alpha) - G \geq 0,
\end{cases}
$$

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

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

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

$$
\begin{cases}
0 = \zeta \left[1 - c(\alpha)^\gamma \Omega \exp(-(1+\sigma)\alpha) y(\alpha)^\sigma \right] f_\alpha(\alpha) + \mu(\alpha) \frac{\Omega(1+\sigma)}{\exp([(1+\sigma)\alpha]} y(\alpha)^\sigma, \\[2mm]
-\mu'(\alpha) = [W(\alpha) - c(\alpha)^\gamma \zeta] f_\alpha(\alpha), \\[2mm]
\mu(0) = \mu(\infty) = 0.
\end{cases} \tag{37}
$$

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

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

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

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

We now consider the decentralization via income taxes (see Section 3.2). Using the FOC (15), the first equation in (37) can be written as

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

where T' is the marginal tax rate. Rearranging terms, we obtain the Diamond-Saez formula for our economy

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

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

Assuming the social welfare function in eq. (16) and imposing logarithmic preferences in consumption, we finally get eq. (21):

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

$$= (1 + \sigma) \frac{1}{f_\alpha(\alpha)} \frac{1}{c(\alpha)} \int_\alpha^\infty \left[c(s) - \exp(-s\theta) \cdot C\right] dF_\alpha(s) \tag{38}$$

where C is aggregate (and average) consumption in the economy.

Slope of Marginal Tax Rates To derive eq. (23) it is important to differentiate the true expression for marginal tax rates before approximating. Taking logs, we obtain

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

Differentiating this gives the slope of optimal marginal tax rates:

$$\frac{d}{d\alpha} \log\left(\frac{T'(y(\alpha))}{1 - T'(y(\alpha))}\right) = -\frac{f'_\alpha(\alpha)}{f_\alpha(\alpha)} - \frac{c'(\alpha)}{c(\alpha)} - \frac{\left[c(\alpha) - \exp(-\alpha\theta) \cdot C\right] f_\alpha(\alpha)}{\int_\alpha^\infty \left[c(s) - \exp(-s\theta) \cdot C\right] dF_\alpha(s)}. \tag{39}$$

This equation tells us that the slope of the optimal marginal tax rate depends on (i) the shape of the productivity density and the consumption schedule, (ii) the consumption level relative to mean

Table 6: Alternative Calculations for $\frac{T'(y(a))}{1-T'(y(\alpha))}$.

Productivity $\frac{\exp(\alpha)}{\text{med.}(\exp(\alpha))}$ (%)	(1) Heathcote Tsujiyama Euler equation wedge	(2) Saez formula $A(\alpha) \times B(\alpha)$	(3) Approximation $(1+\sigma)\frac{F_\alpha(\alpha)}{f_\alpha(\alpha)}\frac{C-c(\alpha)}{c(\alpha)}$
10	0.455	0.455	0.453
20	0.527	0.527	0.519
30	0.561	0.562	0.543
40	0.584	0.585	0.548
50	0.603	0.603	0.542

consumption, and (iii) the taste for redistribution.

Approximation We want to approximate the level and the slope of the marginal tax rate using eqs. (38) and (39). We demonstrate the approximation focusing on the baseline economy with the utilitarian planner, $\theta = 0$. Taking small α, the integrals in (38) or (39) can be approximated as follows:

$$
\begin{aligned}
\int_\alpha^\infty [c(s) - C]\, dF_\alpha(s) &= \int_{-\infty}^\infty c(s)dF_\alpha(s) - \int_{-\infty}^\alpha c(s)dF_\alpha(s) - C[1 - F_\alpha(\alpha)] \\
&= C \cdot F(\alpha) - \int_{-\infty}^\alpha c(s)dF_\alpha(s) \\
&\simeq [C - c(\alpha)]\, F(\alpha).
\end{aligned}
$$

Plugging this expression into eqs. (38) and (39), we obtain eqs. (22) and (23):

$$
\begin{aligned}
\frac{T'(y(\alpha))}{1 - T'(y(\alpha))} &\simeq (1+\sigma)\frac{F(\alpha)}{f_\alpha(\alpha)}\frac{C - c(\alpha)}{c(\alpha)}, \\
\frac{d}{d\alpha}\log\left(\frac{T'(y(\alpha))}{1 - T'(y(\alpha))}\right) &\simeq -\frac{f'_\alpha(\alpha)}{f_\alpha(\alpha)} - \frac{c'(\alpha)}{c(\alpha)} + \frac{f_\alpha(\alpha)}{F_\alpha(\alpha)}.
\end{aligned}
$$

Table 6 reports the quality of the eq. (22) approximation at different values for α. Column (1) reports the marginal tax rate ratio calculated by plugging the allocations that solve the Mirrlees problem into the household's intra-temporal first order condition, and recovering the implicit marginal tax wedge. Column (2) plugs the same allocations into the Diamond-Saez equation (eq. 21). A comparison of columns (1) and (2) indicates that our numerical solution satisfies the Diamond-Saez equation almost exactly, even though we have not used it directly for computation. Column (3) reports the marginal tax rate ratio implied by the simple approximate expression in eq. (22). A comparison of columns (2) and (3) indicates that this approximation is quite good for productivity levels up to around 30% of the median.

B.6 Log-Normal Wage Distribution

Table 7 presents results when we counter-factually assume a log-normal wage distribution.

Table 7: Optimal Tax and Transfer System with Log-Normal Wage Distribution

Tax System	Tax Parameters		Outcomes			
			ω (%)	ΔY (%)	$\overline{T'}$	Tr/Y
HSV^{US}	$\lambda : 0.836$	$\tau : 0.161$	$-$	$-$	0.319	0.017
HSV*	$\lambda : 0.813$	$\tau : 0.285$	0.88	-5.20	0.427	0.048
Affine	$\tau_0 : -0.230$	$\tau_1 : 0.451$	2.19	-6.01	0.451	0.242
Mirrlees			2.28	-5.74	0.443	0.254

Our finding that the shape of the empirical wage distribution is an important driver of the shape of the optimal tax function is not new. Saez (2001) notes that the government should apply high marginal rates at income levels where the density of taxpayers is low – so that the marginal labor supply choices of relatively few households are distorted – but where the fraction of income earned by higher income taxpayers is high – so that higher marginal rates generate significant additional revenue.

The left panel of Figure 11 plots the ratio of the complementary CDF for household income relative to the income-weighted density for (i) the baseline model (in which the α distribution is EMG) and (ii) the log-normal alternative. The ratio declines monotonically with income in the log-normal case, since the mass of income earned by higher income households declines rapidly with income. This weakens the planner's incentive to impose high marginal tax rates at high income levels. In the Pareto log-normal case, in contrast, the same ratio stabilizes as the Pareto tail kicks in. Thus, the temptation to impose high marginal tax rates to raise revenue at high income levels remains strong relative to the associated distortion.

This discussion illuminates the profiles of optimal marginal tax rates for the same two distributional assumptions, shown in the right panel of Figure 11. The baseline specification exhibits a general upward-sloping profile for marginal rates that obviously cannot be replicated by an affine tax system. With a normal distribution for α (given the baseline value for θ), the optimal tax schedule is much flatter, and not surprisingly, an affine schedule can now deliver nearly the same value to the planner.

B.7 Results from Extensions to Richer Tax Structures

With polynomial tax systems, the households' first-order conditions are not sufficient in general. However, it is possible to prove that marginal utility is decreasing in income at sufficiently high income levels. Hence, for a given tax system, equilibrium allocations can be found by evaluating all roots of the household first-order necessary conditions in the range $[0, y]$ with y sufficiently large. For both the quadratic and cubic polynomial cases we search for the optimal tax function coefficients using the Nelder-Mead simplex method. We check that the social welfare maximizing policy is independent of the initial set of tax parameters used to start the search process. Table 8
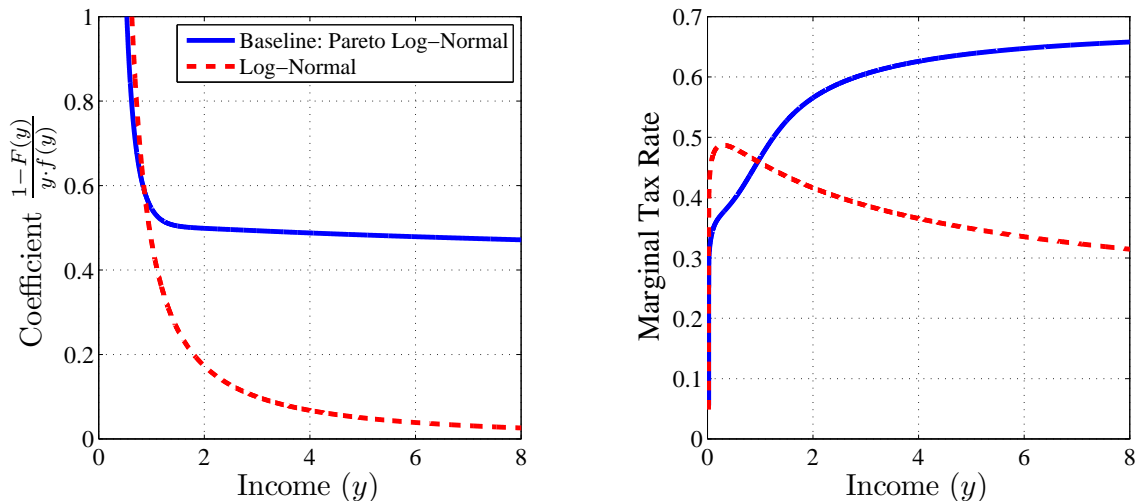
Figure 11: Log-Normal versus Pareto Log-Normal Wage Distribution. The left panel plots the ratios of the complementary CDF for household income relative to the income-weighted density. The right panel plots the profiles of optimal marginal tax rates. The plots are truncated at eight times average income.

presents outcomes for the best policies in the quadratic and cubic classes.

With the quadratic function, marginal tax rates are increasing in income ($\tau_2 > 0$) – the key property of the optimal tax schedule. Relative to the affine case, the linear coefficient τ_1 is reduced, and lump-sum transfers τ_0 are also smaller. Thus, the planner relies more heavily on increasing marginal tax rates, rather than lump-sum transfers, as the primary tool for redistribution. Under the cubic system, the linear coefficient and lump-sum transfers are reduced still further, whereas the quadratic coefficient τ_2 is larger, so that marginal tax rates now rise more rapidly at low income levels. The cubic coefficient τ_3 is negative.

Figure 12 is the analogue to Figure 4 for the cubic tax function. The top panels show that allocations under the cubic policy are generally close to the constrained efficient Mirrlees solution. In particular, for intermediate values for productivity (where the vast majority of households are concentrated), marginal and average tax rates are very similar to those implied by the Mirrlees solution. This explains why the cubic system comes very close, in welfare terms, to the Mirrlees solution.

Table 9 describes optimal type-contingent tax functions and the associated outcomes. The subscripts H and L correspond to tax schedule parameters for the κ_{High} and κ_{Low} types, respectively. By implementing type-contingent tax systems, the Ramsey planner achieves welfare gains that nearly match those under the Mirrlees planner. Under an affine system, the high κ type faces a double whammy, paying higher marginal tax rates than the low type ($\tau_1^H > \tau_1^L$) and paying lump-sum taxes rather than receiving transfers ($\tau_0^H > 0 > \tau_0^L$). Higher marginal rates are an effective way for the planner to redistribute from the high to the low type (recall that κ enters the level

Table 8: Polynomial Tax Functions

Tax System and Tax Parameters					Outcomes			
					ω (%)	ΔY (%)	$\overline{T'}$	Tr/Y
Baseline HSV								
			$\lambda : 0.839$	$\tau : 0.161$	$-$	$-$	0.319	0.018
Polynomial								
	τ_0	τ_1	τ_2	τ_3				
Proportional	$-$	0.177	$-$	$-$	-5.25	6.03	0.177	-0.009
Affine	-0.259	0.492	$-$	$-$	1.77	-8.00	0.492	0.279
Quadratic	-0.236	0.439	0.014	$-$	2.21	-8.01	0.492	0.254
Cubic	-0.212	0.370	0.049	-0.002	2.40	-8.01	0.491	0.228
Mirrlees					2.48	-7.99	0.491	0.213

wage multiplicatively), whereas the wealth effect associated with lump-sum taxes ensures that high κ households still work relatively hard.
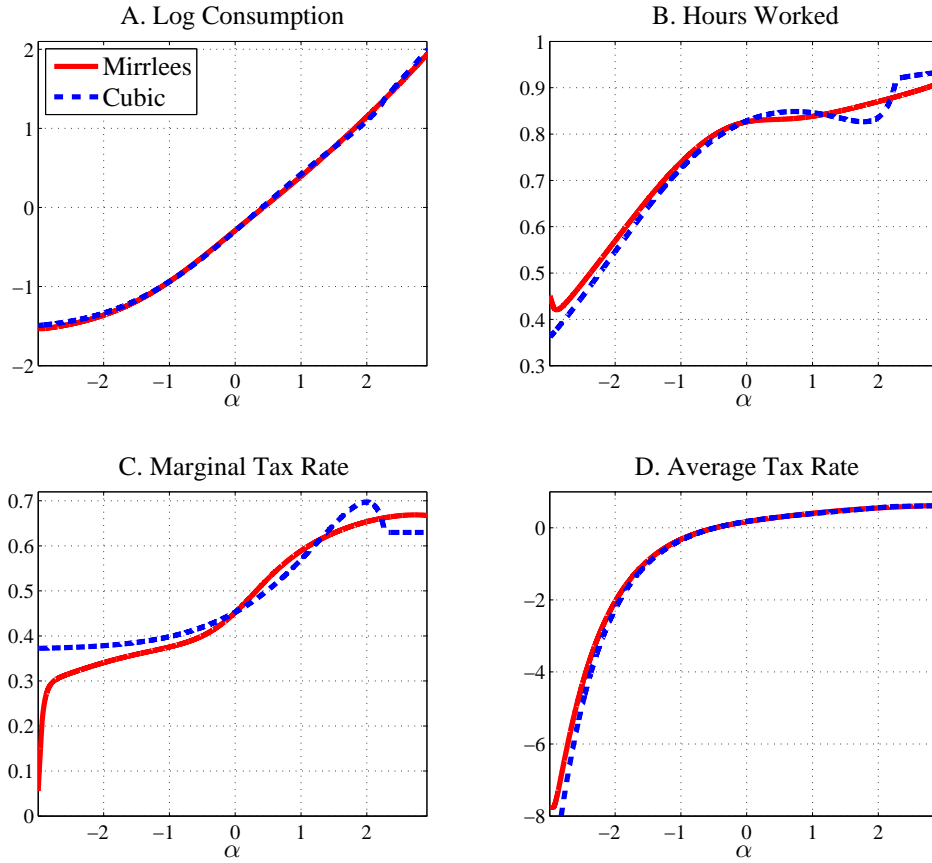
Figure 12: Cubic Tax Function. The figure contrasts allocations under the best-in-class cubic and Mirrlees tax systems. The top panels plot decision rules for consumption and hours worked, and the bottom panels plot marginal and average tax schedules. The plot for hours worked is for an agent with average ε.

Table 9: Type-Contingent Taxes

Tax System			Outcomes			
			ω (%)	ΔY (%)	$\overline{T'}$	Tr/Y
HSV^{US}	$\lambda : 0.834$	$\tau : 0.161$	–	–	0.319	0.015
						0.020
HSV*	$\lambda^L : 1.069$	$\tau^L : 0.480$	6.21	-2.80	0.416	0.147
	$\lambda^H : 0.595$	$\tau^H : 0.073$				-0.019
Affine	$\tau_0^L : -0.403$	$\tau_1^L : 0.345$	6.15	-2.53	0.421	0.420
	$\tau_0^H : -0.032$	$\tau_1^H : 0.452$				0.008
Mirrlees			6.54	-2.53	0.418	0.368
						0.007