Optimal Income Taxation: Mirrlees Meets Ramsey*

Jonathan Heathcote
FRB of Minneapolis and CEPR

Hitoshi Tsujiyama
Goethe University Frankfurt

January 2, 2015

Abstract

What structure of income taxation maximizes the social benefits of redistribution while minimizing the social harm associated with distorting the allocation of labor input? Many authors have advocated scrapping the current tax system, which redistributes primarily via marginal tax rates that rise with income, and replacing it with a flat tax system, in which marginal tax rates are constant and redistribution is achieved via non-means-tested transfers. In this paper we compare alternative tax systems in an environment with distinct roles for public and private insurance. We evaluate alternative policies using a social welfare function designed to capture the taste for redistribution reflected in the current tax system. In our preferred specification, moving to the optimal flat tax policy reduces welfare, whereas moving to the optimal fully nonlinear Mirrlees policy generates only tiny welfare gains. These findings suggest that proposals for dramatic tax reform should be viewed with caution.

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

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

1 Introduction

In this paper we revisit a classic question in public finance: what structure of income taxation can maximize the social benefits of redistribution while minimizing the social harm associated with distorting the allocation of labor input?

The tax and transfer systems currently in place in many countries, including the United States, achieve a measure of redistribution through a combination of tax rates that increase with income, coupled with means-tested transfers. However, many authors have argued that a more efficient way to redistribute would be to move to a flat tax system, in which marginal tax rates are constant across the income distribution, and redistribution is achieved via universal transfers. For example, Friedman (1962) advocated "negative income tax," which effectively combines a lump-sum transfer with a constant marginal tax rate. Mirrlees (1971) pioneered the study of optimal tax design in environments with unobservable heterogeneity. He did not impose any constraints on the shape of the tax schedule, but still found that the optimal schedule is in fact close to linear, a finding that extends to many other papers in the subsequent literature; see Mankiw, Weinzierl, and Yagan (2009) for a recent survey.

In this paper we explore the welfare consequences of replacing the current U.S. income tax system – which redistributes primarily via increasing marginal tax rates – with two alternatives tax schemes: the optimal one and the best policy in the affine class. In our preferred model specification, we find that moving to the best affine policy is welfare reducing, whereas moving to the optimal fully nonlinear Mirrlees policy generates only tiny welfare gains. These findings suggest that proposals for dramatic tax reform should be viewed with caution.

A natural starting point for characterizing the optimal structure of taxation is the Mirrleesian approach, 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. We will contrast the optimal Mirrlees policy with two particular 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 specification is borrowed from Heathcote, Storesletten, and Violante (2014b), who show that the following parametric function for taxes net of transfers closely approximates the current U.S. tax and transfer system: $T(y) = y - \lambda y^{1-\tau}$. In this HSV specification the parameter τ controls the progressivity of the tax and transfer system: for $\tau = 0$ taxes are proportional to income, whereas for $\tau > 0$ the marginal tax rate increases with income. Note that the HSV specification rules out pure lump-sum transfers, since $T(0) = 0$. By comparing welfare in the two cases, we will learn whether 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 affine or HSV tax systems can come close to decentralizing constrained efficient allocations.

We want to provide quantitative guidance on the welfare-maximizing shape of the tax function. With this goal in mind, we emphasize three dimensions of our analysis. First, rather than simply imposing a utilitarian welfare criterion, we evaluate alternative tax systems using a social welfare function that is designed to be consistent with the amount of redistribution embedded in the U.S. tax code. Second, 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. Third, we use the observed distribution for labor earnings to calibrate the model distribution for labor productivity, which we approximate for computational purposes using a very fine grid. We now discuss these three issues in more detail.

The shape of the optimal tax schedule in any social insurance problem is sensitive to the social welfare function that the planner is assumed to be maximizing.[1] 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: $\theta = 0$ corresponds to the utilitarian case. What is the taste for redistribution in the United States? We argue that the degree of progressivity built into the

[1] For example, in the simplest static Mirrlees problem, one can construct a social welfare function in which Pareto weights increase with productivity at a rate such that the planner has no desire to redistribute and thus (absent public expenditure) no desire to tax. Alternatively, a Rawlsian welfare objective that puts weight only on the least well-off agent in the economy will typically call for a highly progressive tax schedule.

actual U.S. tax and transfer system is informative about the preferences of U.S. voters and policymakers. We are able to 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^*. This *empirically motivated social welfare function* will serve as our baseline objective function.

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. For the purposes of practical tax design, the more risk agents are able to insure privately, the smaller is the role of the government in providing social insurance, and the less redistributive will be the resulting tax schedule.

The form of the distribution of uninsurable risk is known to be critical for the shape of the optimal tax function. Saez (2001) has forcefully argued that the heavy Pareto-like right tail for the distribution of labor earnings suggests that marginal tax rates should be increasing, at least over some portion of the earnings distribution. In our calibration we are careful to replicate observed dispersion in U.S. wages. We assume that log labor productivity follows an exponentially modified Gaussian (EMG) distribution, and we estimate the exponential parameter defining the weight of the right tail using cross-sectional data on the distribution of household earnings from the Survey of Consumer Finances. We decompose the overall variance of wages into the uninsurable and insurable components described above by adapting estimates from Heathcote, Storesletten, and Violante (2014a) on the fraction of idiosyncratic wage variation that is privately insured.

Our key findings are as follows.

First, in our baseline model, the welfare gains of moving from the current tax system – which we approximate using the HSV functional form – to the tax system that decentralizes the Mirrlees solution are very small: 0.1 percentage points of consumption. Moving to the optimal policy in the affine class would reduce welfare by around 0.6 percentage points.

Second, our baseline empirically motivated social welfare function underweights low productivity workers, and a utilitarian planner would want to make the tax system more progressive. However, the utilitarian welfare gain from optimally increasing progressivity within the HSV class of tax functions still exceeds the gain from moving to the optimal policy in the affine class.

Third, similar results extend to a specification in which there is no private insurance against idiosyncratic wage risk. Assuming away private insurance while retaining our baseline social welfare function leads to a larger role for government redistribution and thus more progressive taxation. But the optimal policy in the HSV class again delivers higher welfare than the optimal policy in the affine class.

We conclude from these experiments that the presence of lump-sum transfers is not a critical component of optimal fiscal policy, but that it is very important that marginal tax rates increase with income. This finding should give pause to proponents of affine tax schedules.

Why is it important to have marginal tax rates that increase with income? Saez (2001) emphasized that the distribution for labor productivity is a critical determinant of the shape of the optimal tax schedule. With a heavy right tail to the wage distribution, there is a strong incentive to set high marginal tax rates at high income levels because the associated extra revenue from all higher-income individuals is large. When we counterfactually assume a log-normal distribution for wages, we find that an affine tax very nearly implements efficient allocations, whereas the best policy in the HSV class does less well.

Other elements also play a role in explaining why the efficient tax system involves relatively small transfers in our baseline model. First, our social welfare function puts relatively low weight on the utility of low productivity agents. Second, a portion of low wage draws reflect shocks that are privately insured and do not transmit to consumption. This reduces the role for public transfers in establishing a consumption floor.

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 it 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, Troshkin, and Tsyvinski (2013) are prominent examples. 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, Storesletten, and Violante (2014b), who explore the function used by Feldstein (1969), Persson (1983), and Benabou (2000).

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.[2]

Recent papers by Bourguignon and Spadaro (2012), Saez and Stantcheva (2013), Brendon (2013), and Lockwood and Weinzierl (2014) address the inverse of the optimal taxation problem, which is to characterize the profile for marginal social welfare weights that rationalize a particular observed tax system: given these weights, the observed tax system is optimal by construction. Our approach is similar in spirit, in that it uses the progressivity built into the observed tax system to infer a single parameter in the planner's social welfare function that summarizes the planner's taste for redistribution. In contrast to the inverse-optimum approach, however, this one-parameter function only allows for a simple tilt in social preferences toward relatively high or low productivity workers and requires that Pareto weights vary smoothly with productivity. We find this parametric assumption attractive because it is flexible enough to nest most of the standard social welfare functions used in the literature, but not so flexible that it can rationalize any observed tax system, so that we can still ask how the current tax system could be improved. In addition, our specification 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.

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

[2] 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 familiar zero-tax-at-the-top result.

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.

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.

Although the model we describe here is static, it will become clear that it has 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.[3] A more challenging extension to the framework would be to allow for persistent shocks to the unobservable noninsurable component of productivity α. However, Heathcote, Storesletten, and Violante (2014a) estimate that life-cycle uninsurable shocks account for only 17% of the observed cross-sectional variance of log wages.

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

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

The 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. In Appendix 8.1, 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.

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. In Section 3.4, we argue that allowing the planner to observe and tax income (after within-family transfers) at the individual level would not change the solution to the tax authority's problem.

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

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

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).

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.[4] In Section 4, we develop a methodology for using the degree of progressivity built into the actual tax system to reverse engineer the nature of this variation.

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

[4] We assume symmetric weights with respect to ε to focus attention 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.

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

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 characterize the efficient allocation of labor supply within the family more sharply for the tax functions in which we are particularly interested.

Affine Taxes Suppose taxes are an affine function of income, $T(y) = \tau_0 + \tau_1 y$.[5] Then we have the following explicit solution for hours worked as a function of productivity $\exp(\alpha+\varepsilon)$ and family income $y(\alpha)$:

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

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

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

HSV Taxes Suppose income taxes are in the HSV class, $T(y) = y - \lambda y^{1-\tau}$.[6] 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 head solves

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

[6] Then condition (9) becomes

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

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.

subject to

$$
\begin{aligned}
\int c(\alpha,\tilde{\alpha},\varepsilon)dF_{\varepsilon}(\varepsilon) &= c(\tilde{\alpha}), \\
\int \exp(\alpha+\varepsilon)h(\alpha,\tilde{\alpha},\varepsilon)dF_{\varepsilon}(\varepsilon) &= y(\tilde{\alpha}).
\end{aligned}
$$

The first-order conditions to this problem give

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

Let $U(\alpha,\tilde{\alpha})$ denote expected family utility conditional on a draw α and a report $\tilde{\alpha}=\hat{\alpha}(\alpha)$. Substituting in the allocations above, we get

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

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

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:

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

subject to

$$
\int c(\alpha)dF_{\alpha}(\alpha) + G = \int y(\alpha)dF_{\alpha}(\alpha) \qquad \text{for all } \alpha, \tag{13}
$$

and

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

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.[7] We will solve this problem numerically.

[7]Note, however, that the period utility function for each agent in this problem is not identical to the

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.[8]

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, equations (7) and (8). Substituting in the first-order condition for hours from the Mirrlees problem (11) and letting $c^*(\alpha)$ and $y^*(\alpha)$ denote the values for family income and solution that solve the Mirrlees problem, we can describe how optimal marginal tax rates vary with income:

$$\begin{aligned} 1 - T'\left(y^*(\alpha)\right) &= \frac{h(\alpha,\varepsilon)^{\sigma}}{c^*(\alpha)^{-\gamma}\exp(\alpha+\varepsilon)} \\ &= \frac{\Omega}{c^*(\alpha)^{-\gamma}\exp(\alpha)}\left(\frac{y^*(\alpha)}{\exp(\alpha)}\right)^{\sigma}. \end{aligned} \tag{15}$$

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).

3.4 Individual- versus Family-Level Taxation

To this point, we have assumed 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. In Appendix 8.2 we prove that if the tax function for individual income satisfies condition (9), then optimal consumption and income are independent of ε, as in the version when taxes apply to total family income.

There is no benefit to the Mirrlees planner from being able to observe individual-level income (after within-family transfers) because for the family head it is equally costly for any

utility function we started with in the underlying problem, since the weight on hours worked is now Ω rather than 1.

[8]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 decentralization with income taxes as long as income at those values is taxed sufficiently heavily.

family member to deliver income to the planner and equally valuable for any family member to receive additional consumption. Thus, there is no way for the planner to tailor allocations to actual values for ε and therefore no reason for the planner to ask agents to report ε.

4 Estimating Social Preferences

In this section we describe our methodology for using the degree of progressivity built into the actual tax system to infer social preferences.

What social welfare function is the government trying to maximize? Absent knowledge of the government's objective function, it is difficult to compare alternative tax systems unless one Pareto dominates the other. One approach is to argue that a particular objective, such as a utilitarian criterion, should be preferred a priori on theoretical grounds. In contrast, our approach will be to assume that the U.S. government has a specific but unknown social welfare function and to use the degree of progressivity built into the actual tax and transfer system to shed light on the nature of that function.

More precisely, we will characterize in closed form the mapping between the taste for redistribution parameter in a one-parameter class of social welfare functions and the progressivity parameter that maximizes welfare within the HSV class of tax and transfer systems. This mapping can be inverted to infer the taste for redistribution that would lead a planner to choose precisely the observed degree of tax progressivity, if the planner were restricted to choosing a policy within the HSV class. Holding the social welfare function fixed, we will then assess the welfare gains or losses from replacing the HSV policy (which by construction is the best in the HSV class) with a best-in-class affine policy or with the best fully nonlinear Mirrlees policy.

4.1 An Empirically Motivated Social Welfare Function

We assume the social welfare function takes the form

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

The 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. This one-parameter specification is flexible enough to nest several standard social preference specifications in the literature.

- 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]

- The case $\theta = 0$ corresponds to the utilitarian case, with equal Pareto weights on all agents. Given $\theta = 0$ and our assumption of separable preferences, a planner that could observe α and apply $\alpha-$specific lump-sum taxes would give all agents equal consumption.

- 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]

An alternative way to motivate an objective function of the form (16) is to appeal to a positive political economic model of electoral competition. In the probabilistic voting model, 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 of the form described by eq. (16), with higher weights on more productive (and more tax sensitive) households.[11]

Tax System Heathcote, Storesletten, and Violante (2014b) argue that the following income tax function closely approximates the actual U.S. tax and transfer system (see Section 5 for more details):

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

Thus, we adopt this specification as our baseline tax function. The marginal tax rate on individual income is given by $T'(y) = 1 - \lambda(1-\tau)y^{-\tau}$.

[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 $\theta \to \theta$ 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 $\alpha-$specific lump-sum taxes would choose: (i) consumption proportional to productivity, $c(\alpha) \propto \exp(\alpha)$, and (ii) hours worked independent of α.

[11]Persson and Tabellini (2000) offer a detailed explanation of the probabilistic voting model.

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:

$$\frac{1 - T'(y)}{1 - \frac{T(y)}{y}} = 1 - \tau \qquad \forall y.$$

Planner's Problem 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.

Empirically Motivated Social Welfare Function Although we do not directly observe θ, we do observe the degree of progressivity chosen by the U.S. political system. Denote this degree by τ^*. Our baseline *empirically motivated social welfare function* will be $W(\alpha; \theta^*)$ where the social preference for redistribution θ^* satisfies $\hat{\tau}(\theta^*) = \tau^*$. In our view this social welfare function is an attractive benchmark for the purposes of comparing alternative tax policies because it rules out the possibility of generating welfare gains simply by increasing or reducing the degree of progressivity embedded in the current functional form for taxes (17). In this sense, any alternative policy that delivers higher welfare than the current policy must feature a more efficient shape for the income tax function.

The next subsection describes the operational details of how we reverse engineer θ^* given the observed value τ^*.

4.2 Closed-Form Implementation in Baseline Model

The methodology for constructing a mapping between θ and τ is as follows. First, combining the budget constraint, $c(\alpha) = \lambda y(\alpha)^{1-\tau}$, and the expression for hours, eq. (10), gives the following expression for equilibrium household income for a family of type α under the HSV tax scheme:

$$y(\alpha) = \left[\lambda^{1-\gamma}(1-\tau)\exp\left((1+\sigma)\,\alpha\right)\Omega^{-1}\right]^{\frac{1}{\sigma+\tau+\gamma(1-\tau)}}. \tag{18}$$

Given a value for G, eq. (18) can be substituted into the government budget constraint (5) to solve for λ for any possible value for τ. Individual utility, for any pair (α, ε), can then be evaluated as a function of τ (note that these utilities are independent of θ). The last step is to evaluate social welfare for all combinations of θ and τ and to ask for what value(s) for

θ the social-welfare-maximizing value for τ is equal to the value for progressivity estimated from tax data.

In our baseline calibration, utility is assumed to be logarithmic in consumption, so $\gamma = 1$. We also assume that F_α is exponentially modified Gaussian, $EMG(\mu_\alpha, \sigma_\alpha^2, \lambda_\alpha)$, and that F_ε is Gaussian, $N(-v_\varepsilon/2, v_\varepsilon)$. Given these functional form assumptions (which we discuss in the next section), we can derive closed-form expressions for λ and for social welfare and an implicit closed-form expression mapping from τ to θ^*.[12] In particular, we have the following.

Proposition 1 *The social preference parameter θ^* is a solution to the following quadratic equation:*

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

where g is the ratio of government purchases to output.

Proof. See Appendix 8.3. ■

Equation (19) 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 as $\lambda_\alpha \to \infty$ in (19) gives the following explicit solution for θ^*[13]

$$\theta^* = -(1-\tau) + \frac{1}{\sigma_\alpha^2}\frac{1}{(1+\sigma)}\left[\frac{1}{(1-g)(1-\tau)} - 1\right]. \qquad (20)$$

In eqs. (19) and (20), g denotes the observed value for the ratio G/Y where Y denotes aggregate output. For the purpose of inferring θ^*, we can treat g as exogenous. 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.[14]

[12] 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.

[13] This special case provides numerical guidance about which is the relevant root among the two solutions to the quadratic equation (19). One could also consider the case in which α is exponentially distributed ($\sigma_\alpha^2 = 0$), in which case θ^* solves

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

[14] Given eq. (10), we have

$$Y(\tau) = \int\int \exp(\alpha + \varepsilon) h(\varepsilon;\tau) dF_\alpha dF_\varepsilon = (1-\tau)^{\frac{1}{1+\sigma}} \exp\left(\frac{1}{\sigma}\frac{v_\varepsilon}{2}\right).$$

4.3 Comparative Statics

From eqs. (19) or (20), it is straightforward to derive comparative statics on the mapping from structural policy and distributional parameters to θ^*.

Comparative statics with respect to τ **:** First, θ^* is increasing in τ, as expected. Thus if we observe more progressive taxation, all else constant, we can infer that the policymaker puts less weight on higher wage individuals.

The values for τ that signal a laissez-faire or a utilitarian planner, which we denote τ^{LF} and τ^U, can be derived from (19) by substituting in, respectively, $\theta^* = -1$ and $\theta^* = 0$ and solving for the relevant root. Equation (19) 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 (20) is quadratic in τ. In that case, τ^{LF} and τ^U are, respectively,

$$\tau^{LF} = 1 + \frac{1-(1+\sigma)\sigma_\alpha^2 - \sqrt{1+(1+\sigma)^2\left(\sigma_\alpha^2\right)^2 - 2(1+\sigma)\sigma_\alpha^2 + 4\frac{1+\sigma}{1-g}\sigma_\alpha^2}}{2(1+\sigma)\sigma_\alpha^2}, \tag{21}$$

$$\tau^{U} = 1 + \frac{1-\sqrt{1+4\frac{1+\sigma}{1-g}\sigma_\alpha^2}}{2\left(1+\sigma\right)\sigma_\alpha^2}. \tag{22}$$

Note that $\tau^U > \tau^{LF}$, as expected.[15]

The clearest signal of a Rawlsian welfare objective ($\theta^* \to \infty$) is a ratio of expenditure to output $g = G/Y(\tau)$ approaching one.[16] 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

$$\tau^R = 1 - G^{1+\sigma}\exp\left(-\frac{1+\sigma}{\sigma}\frac{v_\varepsilon}{2}\right). \tag{23}$$

Note that with $G = 0$, $\tau^R = 1$, but for $G > 0$, $\tau^R < 1$. Indeed, if $G \geq \exp\left(\frac{v_\varepsilon}{2\sigma}\right)$, 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.[17]

---

[15]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. (19).

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

[17]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. (29) in Appendix 8.3). A Rawlsian planner's desire to minimize the

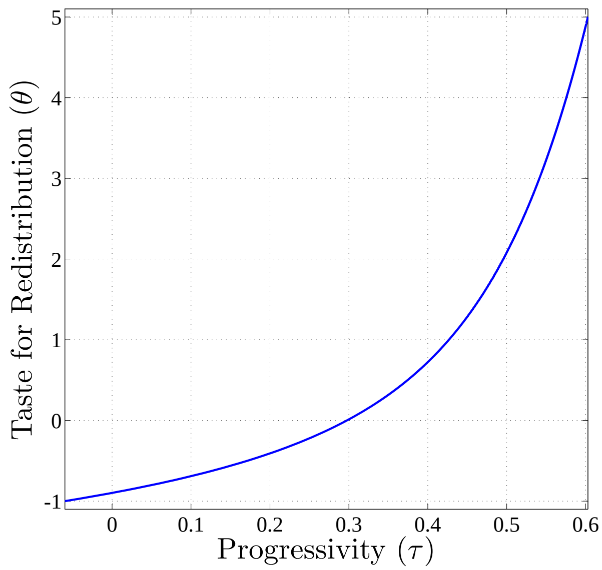

Figure 1: Mapping from τ to θ. This figure is plotted from the closed-form expression in Proposition 1 for the economy with a continuous unbounded EMG distribution for α. We use the version of the expression involving G (eq. (33)).

Figure 1 plots 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 (see Section 5 for more details). The value for progressivity that would signal a utilitarian ($\theta = 0$) social planner is $\tau^U = 0.3$, implying an average effective marginal tax rate of 43%. A laissez-faire social planner ($\theta = -1$) would choose a regressive scheme, with $\tau^{LF} = -0.06$. As we will see, the actual tax and transfer system in the United States lies in between these two values: $\tau = 0.151$ and the average marginal tax rate is 31%. Thus, observed policy appears inconsistent with the existence of either a utilitarian planner or a laissez-faire planner in the United States.

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

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.

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 (21) and (22)). Similarly, the larger is G, the smaller is the value τ^{R} consistent with a Rawlsian objective (see (23)).

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, which we will revisit later, is that a heavier right tail in the distribution for α implies higher optimal progressivity (holding fixed θ).

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, Storesletten, and Violante

(2014b). 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, Storesletten, and Violante (2014a). 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 (footnote 14) with respect to one minus the average income-weighted marginal tax rate is also equal to $1/(1+\sigma)$.[18] According to Saez, Slemrod, and Giertz (2012) the best available estimates for the long run version of this elasticity range from 0.12 to 0.40, 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 or public redistribution.

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, Storesletten, and Violante (2014b) show that this functional form closely mimics actual effective tax rates in the United States. They 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, Storesletten, and Violante (2014b) estimate $\tau = 0.151$. Figure 2, 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 income, ranked from lowest to highest pre-government income. The x-axis shows log of average pre-government 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 $(1 - 0.151)$.

The remaining fiscal policy parameter, λ, is set such that aggregate revenue net of transfers is equal to 18.8% of GDP, which was the ratio of government purchases to output in the United States in 2005 ($g = 0.188$).

Wage Variances We need to characterize individual productivity dispersion, and to decompose this dispersion into an uninsurable component α, and an orthogonal insurable com-

[18] The average income-weighted marginal tax rate is $1-(1-g)(1-\tau)$ (see eq. 4 in Heathcote et al., 2014).

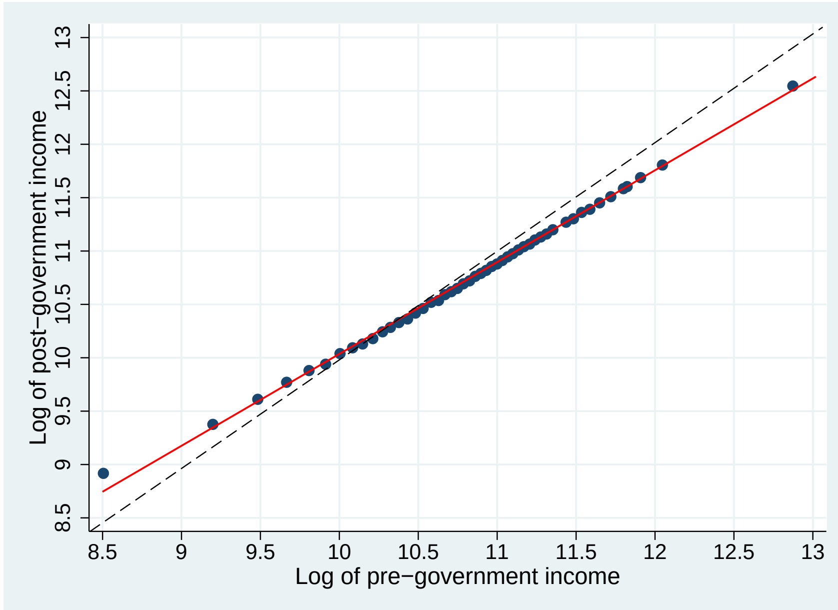

Figure 2: Fit of HSV Tax Function. The figure, from Heathcote, Storesletten, and Violante (2014b), shows the relationship between log household income before and after taxes and transfers for working age households in the PSID. The red line shows the least squares best fit through the underlying micro data, with estimated slope (1-0.151).

ponent ε. For the variances of the two wage components, we adapt estimates from Heathcote, Storesletten, and Violante (2014a). They estimate a richer version of the model considered in this paper using micro data from the PSID and the Consumer Expenditure Survey (CEX), assuming that an individual's labor productivity is equal to their reported earnings per hour. 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. They find a variance for insurable shocks of $v_\varepsilon^* = 0.193$ in 2004 (adding the two solid lines in panels B and C of their Figure 3), which we adopt directly.[19] Our target for variance of the uninsurable component is $v_\alpha^* = 0.273$, implying a total cross-sectional wage variance of $v_\alpha^* + v_\varepsilon^* = 0.466$ (the variance of log hourly wages for men in 2005 reported in Heathcote, Perri, and Violante (2010)).[20] Given these estimates, 41% of the variance of log

[19] As in this paper, the model in Heathcote, Storesletten, and Violante (2014a) features two independent sources of wage risk – one that is explicitly insurable and one that remains uninsured in equilibrium. In the background the government runs the same type of tax-transfer system considered in this paper. The environment in Heathcote, Storesletten, and Violante (2014a) features explicit dynamics from which the present paper abstracts. In addition, individuals in that model differ with respect to preferences as well as labor productivity.

[20] Heathcote, Storesletten, and Violante (2014a) report a slightly lower total wage variance of 0.419 for 2004 (Figure 3). However, their estimate is net of wage variation reflecting differences in age. Rather than excluding this wage variation, we treat it as part of the uninsurable component of wages. In Section 6.4.2,

earnings is directly privately insurable.

Wage Distribution We now have estimates for the variances of the two wage components, but it is well known that higher moments of the shape of the productivity distribution have a large impact on the shape of the optimal tax schedule (see, e.g. Saez (2001)). In fact, the importance of the shape of the right tail of the productivity distribution is explicit in the mapping we have derived between social preferences and desired progressivity: see eq. (19). Saez (2001) argued that there is more mass in the right tail than would be implied by a log-normal wage distribution and that the right tail of the log wage distribution is well approximated by an exponential distribution (so the right tail of the level wage distribution is Pareto). Unfortunately, as Mankiw, Weinzierl, and Yagan (2009) emphasize, it is difficult to sharply estimate the shape of the 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.[21]

Recall that we need distributions for both the insurable and uninsurable components of the wage. We will assume that the insurable component ε is normally distributed, $\varepsilon \sim N(-v_\varepsilon^*/2, v_\varepsilon^*)$, 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. It is natural to attribute the heavy right tail in the log wage distribution to the uninsurable component for wages, since it is unlikely that people are insured against the risk of becoming extremely rich.[22] 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.

We will use the SCF data to estimate λ_α. We do this by maximum likelihood, searching for the values of the three parameters in the EMG distribution that maximize the likelihood of drawing the observed distribution of log earnings. The resulting estimate is $\lambda_\alpha = 2.2$. Figure 3 plots the empirical density against a normal distribution with the same mean and

we will explicitly consider allowing the planner to condition taxes and transfers on observables such as age.

[21]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 underreport income to SCF interviewers.

[22]For example, many individuals far in the right tail of the earnings distribution are entrepreneurs, and it is notoriously difficult to diversify entrepreneurial risk.

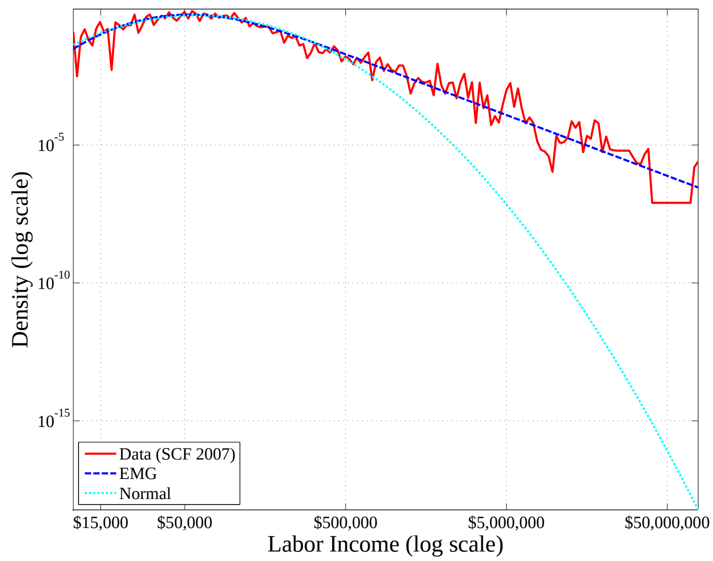

Figure 3: Fit of EMG Distribution. The figure plots the empirical density from the SCF against the estimated EMG distribution and against a normal distribution.

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 excellent fit, substantially improving on the normal specification.[23]

One might be concerned that we are estimating λ_α using an empirical distribution for earnings rather than wages (the SCF does not contain data on hours worked, so we cannot construct a wage measure in the usual way as earnings divided by hours worked). Fortunately, given our specification for the tax system and the fact that preferences in the model have the balanced growth property, we know that hours worked are independent of the uninsurable shock α, and thus the distributions for earnings and wages will both be EMG, with the same exponential (right tail) parameter.[24]

Discretization For computational purposes, as noted above, 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.

[23]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).

[24]Hours in our model do respond (positively) to insurable shocks. This implies that the variance of model earnings is larger than the variance of wages. But the distribution of the insurable component of log earnings remains normal, so the total distribution of earnings remains EMG. Similarly, one could introduce normally distributed heterogeneity in preferences, as in Heathcote, Storesletten, and Violante (2014a), without changing the form of the earnings distribution.

Thus, there is no need to approximate the distribution for ε, and we can assume these shocks are drawn from a continuous unbounded normal distribution with mean $-v_\varepsilon^*/2$ and variance v_ε^*.

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. First, we make the endpoints of the grid, α_1 and α_I, sufficiently widely spaced that only a tiny fraction of individuals lie outside these bounds in the true continuous distribution. In particular we set α_1 such that $\exp(\alpha_1) = 0.12$. This is the ratio of the lower bound on labor income we imposed in constructing our sample ($\$10,000$) to average labor income in our sample.[25] We set α_I such that $\exp(\alpha_I) = 74$, which corresponds to household labor income at the 99.99th percentile of the SCF labor income distribution relative to average income ($6.17 million). We read corresponding probabilities π_i directly from the continuous EMG distribution, rescaling to ensure that $\sum_i \pi_i = 1$. In doing so, we use exponential parameter $\lambda_\alpha = 2.2$ (the SCF estimate) and calibrate the other two parameters in the EMG distribution – the normal variance σ_α^2 and the normal mean μ_α – so that (i) the variance of α (which in the continuous distribution case is exactly $\sigma_\alpha^2 + \lambda_\alpha^{-2}$) is equal to v_α^* (the HSV-based estimate) and (ii) $E[\exp(\alpha)] = 1$. For our baseline set of numerical results we set $I = 10,000$. In Section 6.3.3 we repo rt how the results change when we increase or reduce I.

The resulting model wage distribution $\exp(\alpha + \varepsilon)$ is plotted in Figure 4. The distribution appears continuous, even though it is not, because our discretization is very fine.

Social Welfare Given our fiscal policy parameter estimates and the productivity distribution parameters just described, we apply the procedure described in Section 4.2 to infer the taste for redistribution parameter θ^*. The implied estimate is $\theta^* = -0.547$, indicating that the U.S. social planner wants more redistribution than a laissez-faire planner ($\theta = -1$) but less than a utilitarian one ($\theta = 0$). The relative Pareto weights implied by $\theta^* = -0.547$ are illustrated in Figure 5. Pareto weights are increasing in the uninsurable shock α.

Why does the model interpret the existing tax system as signaling a weak taste for redistribution, relative to the utilitarian case? The logic is that the U.S. tax and transfer system is not particularly progressive, even though U.S. households 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

[25] Assuming $2,000$ household hours worked, this in turn corresponds to $5, which is less than the federal minimum wage in 2007 ($5.85). Reducing the bound further would not materially affect any of our results, since given the parameters for the EMG distribution, the probability of drawing $\alpha < \log(0.12)$ is vanishingly small.

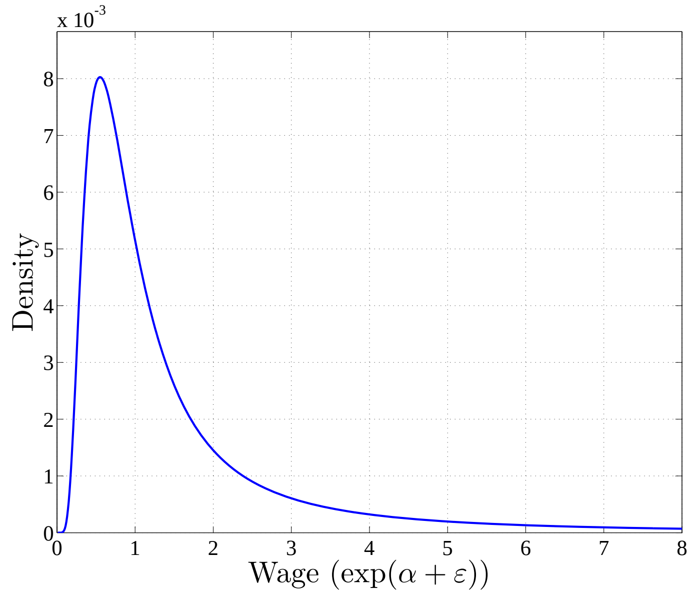

Figure 4: Model Wage Distribution. The average wage is normalized to one. The plot is truncated at eight times the average wage.

finance public expenditure. Thus, the model interprets the absence of a more progressive tax system as reflecting the absence of a strong desire to redistribute on the part of U.S. policymakers. As mentioned previously, 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.

6 Quantitative Analysis

We now explore the structure of the optimal tax and transfer system, given the model specification described above, and given our baseline empirically motivated social welfare function. Specifically, we compute the optimal tax and transfer systems in the affine class and in the fully nonlinear Mirrlees framework and compare allocations and welfare in those cases with their counterparts under our baseline HSV tax and transfer system.[26] Our key findings are (i) moving from the baseline HSV tax system to the best affine policy is welfare reducing, and (ii) moving from the baseline system to the optimal Mirrlees policy generates only tiny welfare gains.

We then explore how sensitive these conclusions are to the choice of taste for redistribution parameter θ. We find that a utilitarian planner ($\theta = 0$) still prefers the best policy in the HSV class to the best affine tax function.

[26] In Appendix 8.4 we explain how we numerically solve the Mirrlees optimal tax problem.

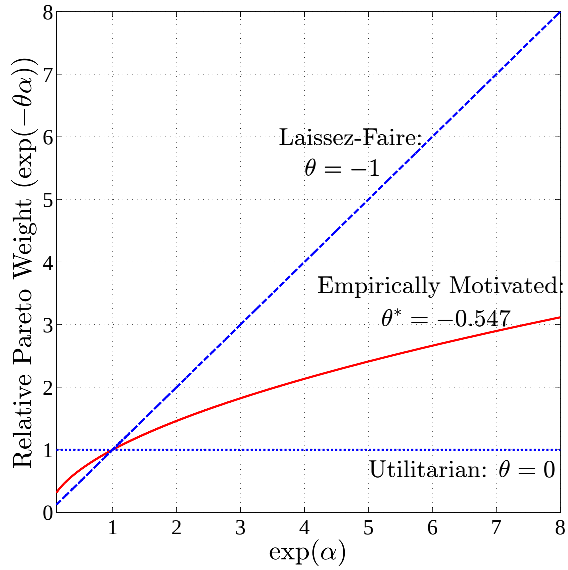

Figure 5: Social Welfare Functions. The figure plots our empirically motivated social welfare function (red solid line) and the utilitarian and laissez-faire alternatives (blue dashed lines).

Next, we explore the implications of assuming away all private insurance. The planner then favors more progressive taxation, but still prefers the best policy in the HSV class to the best affine tax function.

We then replace our baseline EMG distribution for the uninsurable shock α with a normal distribution. This counterfactual changes policy prescriptions significantly, and an affine tax schedule is now close to optimal.

Finally, 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 1 presents outcomes for each tax function. The best tax system in the HSV class is precisely the function estimated for the United States, by virtue of the way our empirically motivated social welfare function was constructed. In particular, the progressivity parameter is the U.S. estimate ($\tau = 0.151$). The outcomes reported for each tax system, 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

Table 1: Optimal Tax and Transfer System in the Baseline Model

Tax System	Tax Parameters		Outcomes			
			ω (%)	ΔY (%)	$\overline{T'}$	Tr/Y
HSV^{US}	$\lambda : 0.836$	$\tau : 0.151$	–	–	0.311	0.018
Affine	$\tau_0 : -0.116$	$\tau_1 : 0.303$	−0.58	0.41	0.303	0.089
Mirrlees			0.11	0.82	0.287	0.003

lowest α type household, relative to average income, Tr/Y.[27]

The best system in the affine class combines lump-sum transfers (i.e., $\tau_0 < 0$) with a proportional tax rate of 30.3%. A first key finding is that moving to the best affine tax system is welfare reducing. 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 allows for. The best affine policy reduces welfare – by 0.6% of consumption – even though it is associated with a lower average marginal tax rate and 0.4 percent higher output.

A second key finding is that moving to the fully optimal Mirrlees system generates welfare gains equal to only 0.1% of consumption. This indicates that the current tax system – more precisely, our HSV approximation to it – is close to efficient and that proposals for tax reform should therefore be treated with caution. 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 the prescriptions that marginal rates should be zero at the top and bottom ends of the productivity distribution.[28]

Note that the average marginal tax rate under the optimal policy is 2.4 percentage points

[27] 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 1, the baseline policy T is the current HSV tax system, and allocations are valued using the empirically motivated value for θ.

[28] Our specification satisfies the conditions provided by Seade (1977), and hence the optimal marginal tax rate at the bottom is zero, in contrast to Saez (2001). Also the familiar no-distortion-at-the-top result applies because the productivity distribution is bounded above.

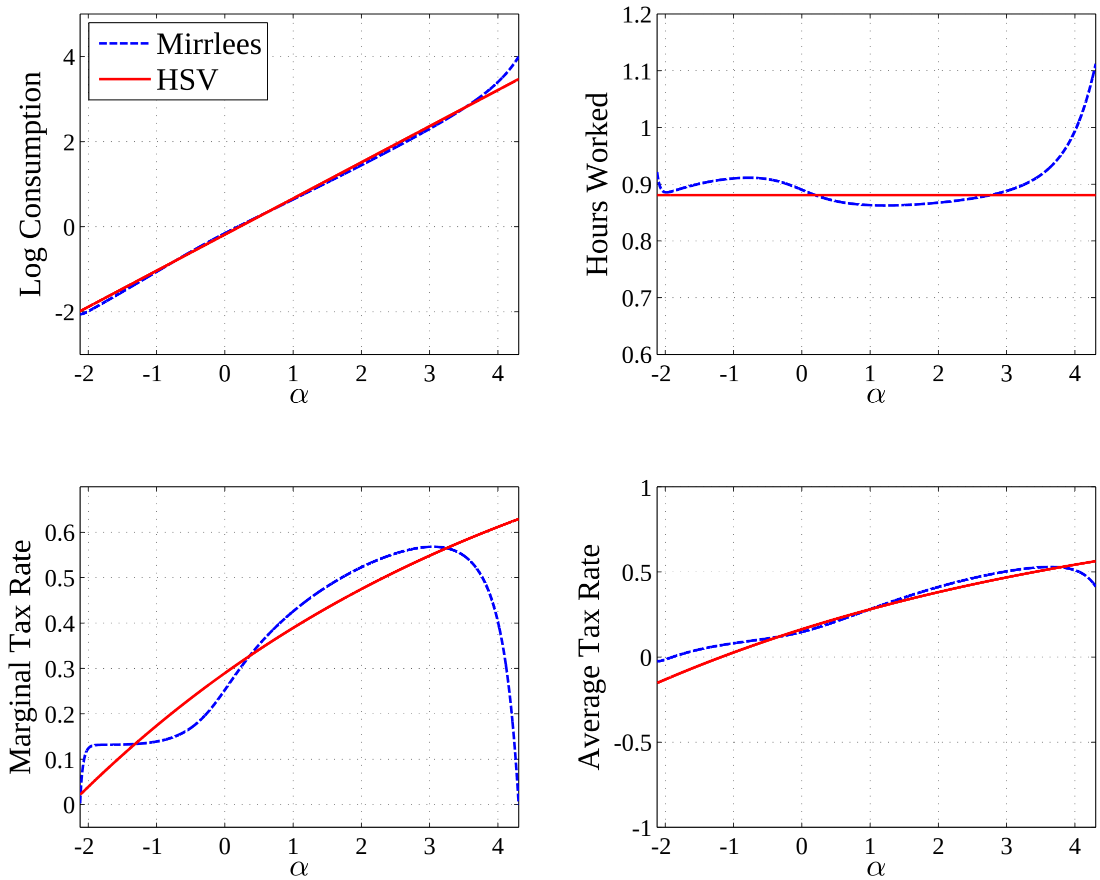

Figure 6: HSV Tax Function. The figure contrasts allocations under the HSV and Mirrlees tax systems. The top panels plot decision rules for consumption and hours worked, and the bottom panels plot marginal and average tax schedules. The plot for hours worked is for an agent with average ε.

lower than under the current system, and the optimal policy generates output gains as well as welfare gains. Interestingly, net transfers received by the least productive agent under the optimal policy are very small.

To develop intuition for these results, Figures 6 and 7 plot decision rules for consumption and hours (top panels) and marginal and average tax schedules (bottom panels) for each best-in-class tax and transfer scheme.[29] Each figure compares a particular third-best Ramsey-style tax function (HSV in Figure 6, affine in Figure 7) to the second-best Mirrlees case.

Allocations under the HSV policy are very similar to those in the constrained efficient Mirrlees case, except for those at the very top of the productivity distribution. Allocations are similar because the HSV marginal and average tax schedules are broadly similar to those

[29]Note that under each tax function, consumption, marginal tax rates, and average tax rates are independent of ε. Hours worked does depend on ε. We plot hours for an individual with the average value for ε.

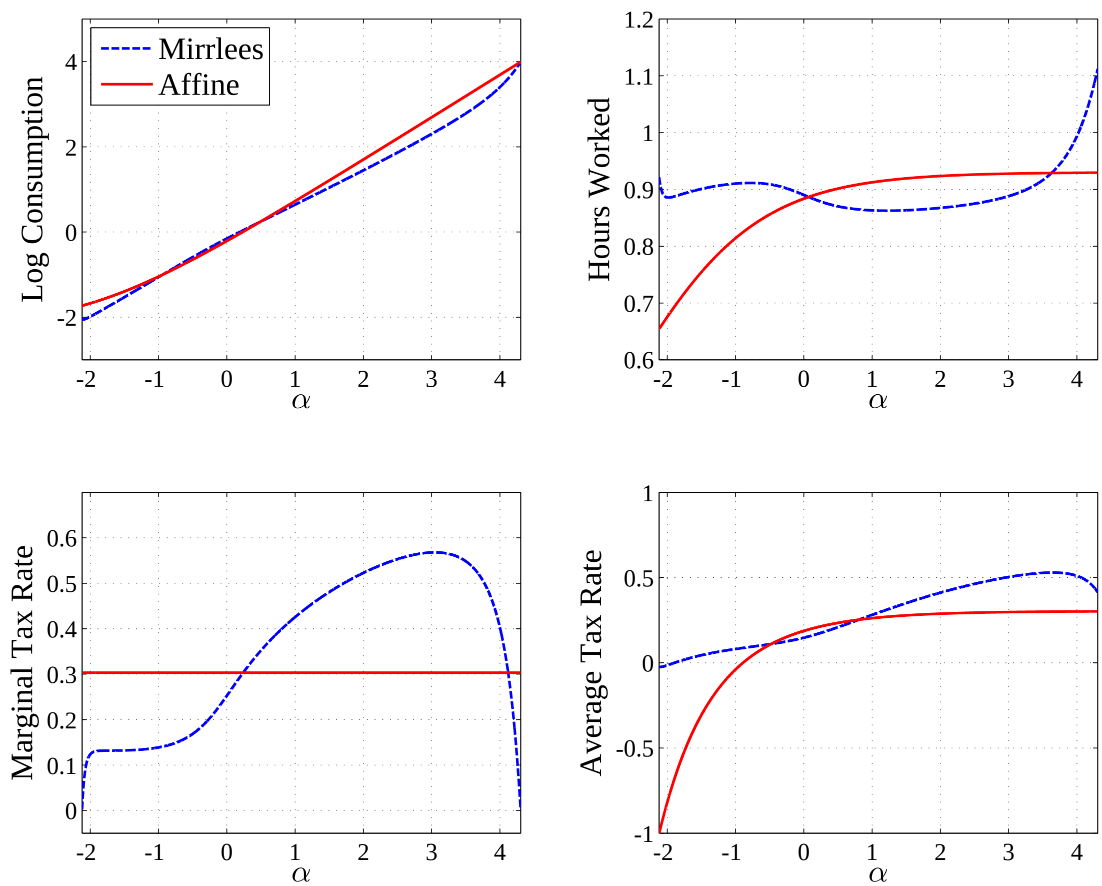

Figure 7: Affine Tax Function. The figure contrasts allocations under the best-in-class affine and Mirrlees tax systems. The top panels plot decision rules for consumption and hours worked, and the bottom panels plot marginal and average tax schedules. The plot for hours worked is for an agent with average ε.

under the optimal policy. 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, although marginal tax rates increase smoothly under the HSV specification, the optimal schedule has a more complicated shape. The optimal marginal rate starts at zero for the least productive households and is fairly flat and low (between 13 and 15%) up to half of average productivity. The optimal marginal rate then rises rapidly to peak at 56.8% at 21 times average productivity before dropping to zero at the very top. The fact that the HSV function does not replicate the decline to zero at the top accounts for the observed differences in allocations at the top. These differences are not very costly in welfare terms because the density of households in this range is very small.

Figure 7 offers a straightforward visualization of why switching to an affine tax is welfare reducing. Because the best affine tax function necessarily features a constant marginal rate, it cannot come close to replicating the optimal schedule. 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.

The bottom right panels of Figures 6 and 7 plot 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. Figure 6 shows that moving from the HSV schedule to the optimal one generates lower average tax rates and thus welfare gains for middle wage earners; the red solid line lies above the blue dashed line from around $\alpha = -0.2$ to $\alpha = 0.8$. Figure 7 shows a similar but more extreme pattern for affine taxes: the optimal scheme implies welfare gains in the middle of the distribution, but much higher average taxes – and lower welfare – for households at both tails of the distribution.

Figure 8 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% of average income to highlight how the different tax systems treat the poor. The line labeled "Mirrlees" 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 1, this household receives a very 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 $(y^*(\alpha_1), c^*(\alpha_1))$. If net transfers at any income value below $y^*(\alpha_1)$ were to lie above this curve, the α_1 type would reject this intended allocation $(y^*(\alpha_1), c^*(\alpha_1))$ in favor of this lower income alternative. Note, however, that any consumption schedule (and associated net tax schedule) that lies everywhere below this indifference curve will also decentralize the Mirrlees solution; the set of possible such schedules is shaded light grey in the figure.[30]

[30] Recall that the marginal tax rate is zero at $(y^*(\alpha_1), c^*(\alpha_1))$, and thus to the left of $y^*(\alpha_1)$ the indifference curve for the α_1 type is flatter than a 45° line. Thus, one example of a tax schedule that ensures the α_1 type will not choose $y < y^*(\alpha_1)$ is a zero marginal tax rate between $y = 0$ and $y = y^*(\alpha_1)$.

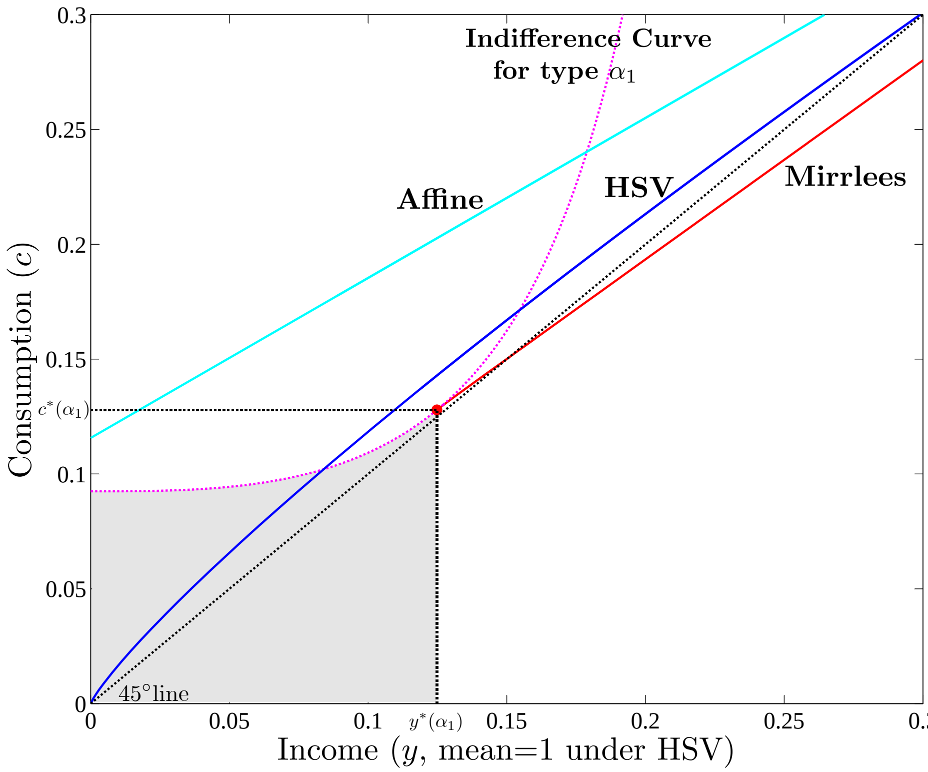

Figure 8: Allocations for Low Income Households. The figure plots household consumption against household income at the bottom of the income distribution under the Mirrlees (red), HSV (dark blue), and best-in-class affine (light blue) tax systems.

Figure 8 also plots the best income tax schedules in the affine and HSV classes. It is clear from the plot that the HSV schedule is much closer than the affine one to the optimal Mirrlees schedule. The affine schedule implies net transfers that are much too generous at the bottom of the distribution. In fact, the pure lump-sum transfer τ_0 under the affine scheme is larger than the upper bound on net transfers implied by the indifference curve for the α_1 type. Why does the affine schedule impose lump-sum transfers that appear to be too large? 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.

One might have anticipated that the HSV tax function would generate suboptimally small net transfers for the least productive households, given that the function rules out pure lump-sum transfers. Instead, however, the HSV scheme actually implies slightly too much redistribution at the bottom of the income distribution.[31] One reason transfers to the least productive households are small under the optimal Mirrlees policy is that our planner

[31] Recall that the HSV scheme generates net transfers at low income levels with $\tau > 0$ because marginal tax rates are negative at low income levels.

puts relatively low weight on such households. A second reason transfers are small is that there are relatively few very low income households in the baseline parameterization, in part because our calibration implies that a portion of wage dispersion is privately insurable.

6.2 Alternative Social Welfare Functions

Recall the two key findings from our baseline parameterization: the welfare gain from adopting the best affine tax schedule is negative, and the welfare gain from moving to the fully optimal policy is very small. What is the role of our baseline social welfare function in delivering these results?

To address this question, we now experiment with alternative values for the taste for redistribution θ. We start with the utilitarian case ($\theta = 0$), since that is a benchmark in the literature. We then consider the laissez-faire ($\theta = -1$), and Rawlsian ($\theta \to \infty$) cases.[32]

Welfare Decomposition To better understand the sensitivity of our welfare results to θ, we will decompose our welfare numbers in two different ways. First, we follow a standard approach to decompose total welfare gains into an aggregate efficiency gain component and a change in inequality component (see, e.g. Heathcote, Storesletten, and Violante (2008)).

Recall that the welfare gain of moving from a generic policy T to an alternative policy $\hat{T}$, expressed as an equivalent percentage change in consumption, is equal to $\omega(T, \hat{T}) = V(\hat{T}, \theta) - V(T, \theta) - 1$, where $V(T, \theta)$ denotes the planner's realized value under policy T given a taste for redistribution θ (see footnote 27). Let $C(T)$ and $H(T)$ denote per capita consumption and per capita hours worked under policy T, and let $V^{RA}(T)$ denote utility for a hypothetical agent enjoying per capita consumption and hours: $V^{RA}(T) = \log C(T) - H(T)^{1+\sigma}/(1+\sigma)$.

Define the *aggregate efficiency gain* of moving from policy T to $\hat{T}$, $\omega^E(T, \hat{T})$, as the percentage increase in consumption under policy T that makes the hypothetical representative agent indifferent about switching to $\hat{T}$:

$$\omega^E(T, \hat{T}) = V^{RA}(\hat{T}) - V^{RA}(T) - 1.$$

Define the *cost of inequality* under policy T, $\omega^I(T)$, as the percentage increase in consumption of all agents required to deliver equal value to the planner as an allocation with zero inequality:

$$\omega^I(T) = V^{RA}(T) - V(T, \theta) - 1.$$

[32] 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.

Table 2: Optimal Tax and Transfer System with Utilitarian Social Welfare

Tax System	Tax Parameters		Outcomes				
			ω (%)	ω^E (%)	ΔY (%)	$\overline{T'}$	Tr/Y
HSV^{US}	$\lambda : 0.836$	$\tau : 0.151$	–	–	–	0.311	0.018
HSV*	$\lambda : 0.821$	$\tau : 0.295$	1.38	−3.76	−6.02	0.436	0.068
Affine	$\tau_0 : -0.233$	$\tau_1 : 0.303$	0.45	−2.02	−6.43	0.452	0.220
Mirrlees			1.53	−3.72	−6.15	0.440	0.122

From these two definitions, it follows that the welfare gain of moving from policy T to policy $\hat{T}$ can be expressed as the sum of the efficiency gain plus the *gain from reduced inequality*:

$$\omega(T,\hat{T}) = \omega^E(T,\hat{T}) + \left[\omega^I(T) - \omega^I(\hat{T})\right].$$

Our second welfare decomposition divides the welfare gains of moving from the U.S. tax system to the optimal tax system, ω(HSV^{US}, Mirrlees), into (i) the maximum welfare gains achievable from changing progressivity τ while retaining the same HSV functional form ω(HSV^{US},HSV*), and (ii) the additional gains from introducing a more flexible non-parametric tax system. We label the first component the *gain from changing progressivity* and the second the *gain from increasing flexibility*.

Utilitarian Social Welfare Table 2 presents results when we compare alternative tax systems using a utilitarian social welfare function ($\theta = 0$) instead of our baseline empirically motivated one. Note that the HSV tax function with the empirical estimate for $\tau = 0.151$ is no longer optimal within the HSV class, and the table therefore has a new row labeled "HSV*" corresponding to the optimal policy within the HSV tax class given $\theta = 0$.

The utilitarian social welfare function features a stronger taste for redistribution than the baseline. It therefore delivers more progressive tax policies. For example, the best policy in the HSV class tax function has higher progressivity ($\tau = 0.295$). The average effective marginal tax rate under the optimal policy is now 44%, compared with 31% under the baseline welfare function. The maximal welfare gain from tax reform is now large: 1.53% of consumption. This welfare gain arises even though higher marginal tax rates under the optimal policy discourage labor supply, reducing output by 6.15% and efficiency by 3.72%.

However, some key results from our baseline specification survive here: the best policy in the HSV class is still preferred to the best affine tax system, and the best HSV class policy still closely approximates – in welfare terms – the constrained efficient Mirrlees policy, capturing more than 90% of the maximum feasible welfare gains (1.38% out of 1.53%). Thus,

Table 3: Alternative Social Welfare Functions

SWF		Mirrlees Allocations			ω and ω^E (in parentheses) (%)			Prog. Share
	θ	$\overline{T'}$	Tr/Y	ΔY	HSV*	Affine	Mirrlees	$\frac{\omega(\mathrm{HSV}^{\mathrm{US}},\mathrm{HSV}^*)}{\omega(\mathrm{HSV}^{\mathrm{US}},\mathrm{Mirrlees})}$
Laissez-Faire	-1	0.070	-0.116	9.78	2.65 (3.48)	2.85 (2.82)	2.86 (2.60)	93%
Baseline	-0.547	0.287	0.003	0.82	– –	-0.58 (1.38)	0.11 (-0.26)	0%
Utilitarian	0	0.440	0.122	-6.15	1.38 (-3.76)	0.45 (-2.02)	1.53 (-3.72)	90%
Rawlsian	∞	0.697	0.535	-21.87	160.24 (-36.39)	268.99 (-21.77)	301.30 (-14.65)	53%

our baseline procedure for inferring the taste for redistribution is not an essential ingredient for finding that it is preferable to redistribute via increasing marginal tax rates rather than via lump-sum transfers.

Alternative Social Welfare Functions Table 3 shows results for a wider range of social welfare functions. 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. A second takeaway is that the choice of social welfare function has a huge impact on the potential welfare gains from policy reform. Had we chosen a Rawlsian welfare function as our baseline, we would have concluded that tax reform could raise welfare by 300% rather than by 0.11%!

The efficiency gain associated with tax reforms, ω^E, is reported in parentheses underneath the baseline welfare number. The welfare gains from tax reform in the laissez-faire case are mostly due to increased efficiency. For the social welfare functions that embed a stronger taste for redistribution than the baseline, in contrast, the welfare gains from more progressive tax schemes entirely reflect the value the planner attaches to reducing inequality, which dominate large efficiency losses in the overall welfare calculus.

The last column of Table 3 reports the fraction of the maximum possible welfare gains that can be achieved merely by changing progressivity within the HSV functional form. This fraction is very large in most cases: in the utilitarian case ($\theta = 0$), increased progressivity accounts for 90% of the total gain, whereas in the laissez-faire case ($\theta = -1$), it captures 93%.

We conclude that the sizable total welfare gains that emerge once we deviate from our baseline empirically motivated value for θ are almost entirely due to the fact that the planner wants a more or less redistributive system than the one currently in place – and not because the current system redistributes in a very inefficient way. In our view, a similar decomposition should be part of any Mirrleesian policy analysis, as a way to elucidate the extent to which welfare numbers are driven by the choice of objective function versus all the other model elements.

Although an HSV-class tax function is better than an affine one under our baseline welfare function and in the utilitarian case, Table 3 indicates that this result does not extend to all possible social welfare functions. The laissez-faire planner prefers an affine tax because he wants to use lump-sum taxes to raise revenue; in fact, this planner chooses negative marginal tax rates. The Rawlsian planners prefer an affine tax because they overweight the least productive agents and value a high consumption floor. 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.

An Alternative Strategy for Estimating θ We have claimed that the maximal welfare gain from tax reform in our baseline specification is small because the HSV functional form is already close to optimal and much closer to optimal than the best-in-class affine system. We have been asked whether these results are baked in the cake as a result of having approximated the actual U.S. tax and transfer system using the HSV functional form and then having searched for a taste for redistribution parameter θ that rationalizes the observed degree of progressivity. To answer this question, we have experimented with taking an affine approximation to the U.S. tax and transfer system and then recomputing the value for θ that rationalizes the estimated (constant) marginal tax rate.

The best fit affine system has a marginal tax rate of 37.7%, coupled with a lump-sum transfer equal to 18.9% of GDP.[33] The implied taste for redistribution is $\theta = -0.314$. Given this value for θ, and taking the estimated affine schedule as the baseline, the welfare gains of moving to the best policy in the HSV class and the best unconstrained policy are 0.86% and 0.95%. Thus, the maximum possible welfare gains from tax reform are quite large here – even though we chose θ so that the baseline tax system is the best possible one in the affine class. We conclude that restricting the tax and transfer schedule to be affine is a costly restriction from a welfare standpoint, whereas restricting the schedule to be in the HSV class is not.

[33] We impose the same ratio of government purchases G to output in the affine case as in the baseline. Thus, with affine taxation, tax revenue is split equally between G and lump-sum transfers. More details about these estimates are available upon request.

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

Tax System	Tax Parameters		Outcomes			
			ω (%)	ΔY (%)	$\overline{T'}$	Tr/Y
HSV^{US}	$\lambda : 0.836$	$\tau : 0.151$	–	–	0.311	0.018
HSV*	$\lambda : 0.839$	$\tau : 0.192$	0.12	−1.64	0.346	0.033
Affine	$\tau_0 : -0.156$	$\tau_1 : 0.360$	−0.21	−1.97	0.360	0.139
Mirrlees			0.23	−2.11	0.361	0.081

6.3 Sensitivity Analysis

We now conduct a sensitivity analysis to better understand why the HSV approximation to the current tax and transfer scheme is close to optimal and preferable to an affine tax scheme. First, we explore the role of the existence of private insurance against a subset of idiosyncratic shocks. Second, we explore the role of the shape of the distribution for the uninsurable component of productivity.[34]

To preview our findings, assuming zero private insurance against idiosyncratic wage risk implies a larger role for public insurance – and thus a more progressive tax system – but does not change the conclusion that the best policy in the HSV class is preferred to the best affine tax function. In contrast, if the earnings distribution in the United States was log-normal, it would be welfare improving to replace the current tax system with an affine schedule.

6.3.1 No Private Insurance

Table 4 presents results when we shut down private insurance by setting $v_\varepsilon = 0$ 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 including the taste for redistribution θ are set to their values in the baseline calibration.[35]

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 the U.S. policy: $\tau = 0.192$. The second-best policy now features an average marginal tax rate of 36% and larger transfers. The maximal welfare gains from tax reform, though still small, are now twice as large as in the baseline model and are associated with an output decline of

[34] 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. Results are available upon request.

[35] We could have reapplied our procedure to infer a new empirically motivated value for θ, but we chose to hold θ fixed so as to isolate the impact of the insurability of shocks for optimal taxation.

Table 5: Optimal Tax and Transfer System with Log-Normal Wage Distribution

Tax System	Tax Parameters		Outcomes			
			ω (%)	ΔY (%)	$\overline{T'}$	Tr/Y
HSV^{US}	$\lambda: 0.836$	$\tau: 0.151$	–	–	0.311	0.018
HSV*	$\lambda: 0.826$	$\tau: 0.070$	0.27	3.10	0.239	−0.005
Affine	$\tau_0: -0.068$	$\tau_1: 0.250$	0.34	2.67	0.250	0.042
Mirrlees			0.35	2.71	0.249	0.042

2%. The best affine tax and transfer system is dominated not only by the best HSV system but also by the current (HSV-approximated) U.S. system.

Recall that imposing a utilitarian social welfare function also implied larger optimal transfers. If we simultaneously assume a utilitarian social welfare and also assume zero private insurance, then the best policy in the HSV class delivers slightly lower welfare than the best affine class policy.[36]

6.3.2 Log-Normal Wage Distribution

Table 5 presents results when we counterfactually assume a log-normal wage distribution. Specifically, we assume $\alpha \sim N(-v_\alpha^*/2, v_\alpha^*)$, holding fixed all other parameter values including θ. Relative to the baseline (Pareto log-normal) case, the distribution for the uninsurable component of wages now has a much thinner right tail and a heavier left tail.

The best HSV-class tax function now features less progressivity than that in the U.S. ($\tau = 0.070$). The second-best policy implies a lower average marginal tax rate of 25% (compared with 29% in the baseline calibration), 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 moreover 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.

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

[36]Detailed results are available upon request.

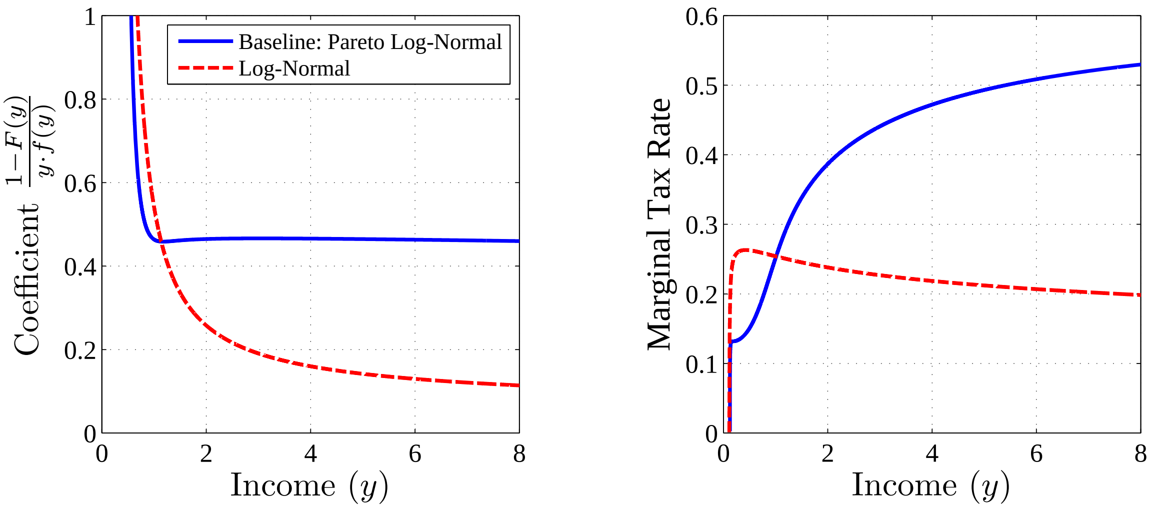

Figure 9: Log-Normal versus Pareto Log-Normal Wage Distribution. The left panel plots the ratios of the complementary CDF for household income relative to the income-weighted density. The right panel plots the profiles of optimal marginal tax rates. The plots are truncated at eight times average income.

rates generate significant additional revenue.

The left panel of Figure 9 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 9. The baseline specification exhibits a general upward-sloping profile for marginal rates that obviously cannot be replicated by an affine tax system.[37] 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. 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.

[37]Imposing a utilitarian social welfare function shifts the entire marginal tax rate profile upward, but does not change its basic shape.

Table 6: Grid points

# of grid points I	ω (%, relative to HSV)		
	Affine	Mirrlees	First Best
25	-0.58	3.82	8.80
100	-0.58	1.17	8.79
1,000	-0.58	0.21	8.80
10,000	-0.58	0.11	8.80
100,000	-0.58	0.10	8.80

6.3.3 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 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 6 we report welfare gains from tax reform (relative to the HSV baseline) as we increase the number of grid points from $I = 25$ to $I = 100,000$.[38] 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 $I = 25$ these gains are 3.8% of consumption, compared with 0.11% 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 familiar 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 10 points. The findings in Table 6 cast some doubt on the quantitative robustness of such analyses if the true underlying productivity distribution is continuous.

[38] It is difficult to reduce I much below 25 while still retaining a realistic shape for the wage distribution.

6.4 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. This additional nonlinearity gives the Ramsey planner more flexibility, allowing it to more nearly replicate the Mirrlees policy. 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% of consumption.[39]

6.4.1 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.[40] Thus,

$$T(y) = \begin{cases} T_n(y) & \text{for } y \leq \bar{y}, \\ T_n(\bar{y}) + T_n'(\bar{y})(y - \bar{y}) & \text{for } y > \bar{y}. \end{cases}$$

We focus on the cases $n = 2$ and $n = 3$ (i.e., quadratic and cubic tax functions).[41] Table 7 and Figure 10 in Appendix 8.5 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.[42] With the best quadratic function, marginal tax rates are increasing in income ($\tau_2 > 0$) – the key property of the optimal tax schedule – whereas lump-sum transfers are smaller than under the best affine scheme. Under the best cubic system, lump-sum transfers are reduced still further, whereas the cubic coefficient τ_3 is negative, which ensures that marginal rates flatten out before

[39] 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.

[40] Thus, $\bar{y} \equiv 10 \times \int \int \exp(\alpha+\varepsilon) h(\alpha,\varepsilon) dF_\alpha(\alpha) dF_\varepsilon(\varepsilon) = 10 \times (1-\tau)^{\frac{1}{1+\sigma}} \exp\left(\frac{1}{\sigma}\frac{v_\varepsilon}{2}\right)$ where the second equality follows from the expression for $h(\alpha,\varepsilon)$ in eq. (10).

[41] The agents' first-order conditions are now not sufficient in general. However, it is possible to prove that the marginal utility is decreasing in income at sufficiently high income levels. Hence, the solution must exist and can be found numerically if we evaluate all the roots of the first-order necessary conditions in the range $[0, y]$ with y sufficiently large.

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

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 half of the (small) maximum potential welfare gains from tax reform.

6.4.2 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, v_κ, equal to the variance of wage dispersion that can be accounted for by standard observables in a Mincer regression. Heathcote, Perri, and Violante (2010) estimate the variance of cross-sectional wage dispersion attributable to observables to be 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 and therefore reduce the variance of α to $v_\alpha^* - v_\kappa = 0.165$. The three parameters $\mu_\alpha, \sigma_\alpha^2$, and λ_α characterizing the EMG distribution for α are now recalibrated so that (i) the variance of α is equal to 0.165, (ii) $E[\exp(\alpha)] = 1$, and (iii) the value of the shape parameter $\sigma_\alpha \lambda_\alpha$ is the same as that in the baseline model (1.129).[43]

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 8 in Appendix 8.5 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 1.46% of consumption, compared with only 0.11% in the baseline analysis. This large welfare gain arises in part because the average effective marginal tax rate drops to only 20%, which translates into

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

a 5.2% output gain. 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 paying lump-sum taxes rather than receiving 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, e.g., Heathcote, Storesletten, and Violante (2014b)). In particular, relatively high taxation of high κ households would discourage education investment. Thus, we regard our 1.46% welfare gain as an upper bound on the feasible welfare gains from tagging.

7 Conclusions

All governments must decide how taxes and transfers should vary with income. The character of the optimal policy depends on the answers to two questions. First, how important is income redistribution as a public policy objective? Second, holding fixed the desire to redistribute, what is the most efficient way to do so?

In this paper we answered the first question by using the progressivity of the current tax and transfer system, interpreted through the lens of a structural model, to impute a taste for redistribution to the U.S. government. Given this taste for redistribution, we then characterized the optimal tax schedule and explored how closely it can be approximated by similar parametric functions.

Our key findings are, first, that it is difficult to improve substantially on the current system and, second, that moving to the best affine tax and transfer scheme would be welfare reducing. Optimal marginal rates are rising in income at high income levels because the estimated underlying distribution for wages has a Pareto right tail. Large lump-sum transfers are not desirable, in part because the estimated taste for redistribution is modest and in part because a portion of wage dispersion is privately insured.

We now highlight four lessons from our analysis that should be useful for future work geared toward providing practical quantitative advice on tax and transfer design.

First, in any heterogeneous agent environment, 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 value for this parameter is the value that rationalizes the progressivity embedded in the current tax and transfer system.

Second, for the purposes of practical policy advice, the role of the tax and transfer system should be limited to offering a degree of public insurance against risks that cannot be insured privately. Thus, for quantitative policy guidance, it is important to model both public and private insurance.

Third, optimal policy is sensitive to the shape of the underlying productivity distribution. This finding is not new, but we have shown that a three-parameter exponentially modified Gaussian distribution fits the empirical earnings distribution remarkably well.

Fourth, 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, Storesletten, and Violante (2014b) 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.

Our model environment could be enriched along various dimensions. First, the only decision margin that taxes distort in our model is labor supply. Although this has been the focus of the optimal tax literature, skill investment and entrepreneurial activity are additional decision margins that are likely sensitive to the tax system. Second, our model is static. In dynamic environments, the Mirrlees planner can do better by making taxes history dependent, rewarding higher current income and tax payments with lower future taxes. However, the welfare gains from this added sophistication may be modest, given that uninsurable life-cycle shocks to productivity are small relative to permanent differences at the time of labor market entry. Third, a sharper characterization of the bottom of the distribution for productivity would be desirable. The challenge is that wages are not observed for nonparticipants, who are likely less productive, on average, than workers.

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(1), 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(2), 85–114.

Conesa, J. C., and D. Krueger (2006): "On the Optimal Progressivity of the Income Tax Code," *Journal of Monetary Economics*, 53(7), 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(1), 83–95.

Farhi, E., and I. Werning (2013): "Insurance and Taxation over the Life Cycle," *Review of Economic Studies*, 80(2), 596–635.

Feldstein, M. S. (1969): "The Effects on Taxation on Risk Taking," *Journal of Political Economy*, 77(5), 755–764.

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

Golosov, M., M. Troshkin, and A. Tsyvinski (2013): "Redistribution and Social Insurance," mimeo, Princeton University.

Gouveia, M., and R. P. Strauss (1994): "Effective Federal Individual Income Tax Functions: An Exploratory Empirical Analysis," *National Tax Journal*, 47(2), 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(1), 15–51.

Heathcote, J., K. Storesletten, and G. L. Violante (2008): "Insurance and Opportunities: A Welfare Analysis of Labor Market Risk," *Journal of Monetary Economics*, 55(3), 501–525.

——— (2014a): "Consumption and Labor Supply with Partial Insurance: An Analytical Framework," *American Economic Review*, 104(7), 2075–2126.

——— (2014b): "Optimal Tax Progressivity: An Analytical Framework," mimeo, Federal Reserve Bank of Minneapolis.

Keane, M. P. (2011): "Labor Supply and Taxes: A Survey," *Journal of Economic Literature*, 49(4), 961–1075.

Lockwood, B. B., and M. Weinzierl (2014): "Positive and Normative Judgments Implicit in U.S. Tax Policy and the Costs of Unequal Growth and Recessions," mimeo, Harvard Business School.

Mankiw, N. G., M. Weinzierl, and D. Yagan (2009): "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, 23(4), 147–74.

Mirrlees, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(2), 175–208.

Persson, M. (1983): "The Distribution of Abilities and the Progressive Income Tax," *Journal of Public Economics*, 22(1), 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(1), 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(1), 3–50.

Saez, E., and S. Stantcheva (2013): "Generalized Social Marginal Welfare Weights for Optimal Tax Theory," Working paper 18835, NBER.

Seade, J. K. (1977): "On the Shape of Optimal Tax Schedules," *Journal of Public Economics*, 7(2), 203–235.

Weinzierl, M. (2011): "The Surprising Power of Age-Dependent Taxes," *Review of Economic Studies*, 78(4), 1490–1518.

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

8 Appendices

8.1 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

$$\int B(\alpha,\varepsilon)Q(\varepsilon)d\varepsilon = 0, \tag{24}$$

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

$$c(\alpha,\varepsilon) = y(\alpha,\varepsilon) - T\left(y(\alpha,\varepsilon)\right) \qquad \text{for all } \varepsilon, \tag{25}$$

where individual income before taxes and transfers is given by

$$y(\alpha,\varepsilon) = \exp(\alpha+\varepsilon)h(\alpha,\varepsilon) + B(\alpha,\varepsilon) \qquad \text{for all } \varepsilon. \tag{26}$$

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 equations (24), (25), and (26). 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.[44]

Finally, note that if insurance against ε is achieved via decentralized financial markets, the planner could conceivably ask agents to report ε after the ε shock is drawn and offer allocations for consumption $c(\tilde{\alpha}, \tilde{\varepsilon})$ and income $y(\tilde{\alpha}, \tilde{\varepsilon})$ indexed to reports of both α and ε. With decentralized insurance, the planner might be able to offer contracts that separate agents with different values for ε (recall that under the family model for insurance, this was not possible). One might think there would be no possible welfare gain to doing so, since private insurance already appears to deliver an efficient allocation of hours and consumption within any group of agents sharing the same α. However, it is possible that by inducing agents to sacrifice perfect insurance with respect to ε, the planner can potentially loosen incentive constraints and thereby provide better insurance with respect to α.[45] 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.

8.2 Individual versus Family Taxation

Proposition 2 *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, equations (7) and (8).

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

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

[45] 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 6.4.2).

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

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

subject to

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

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

The FOCs are

$$\left[y(\alpha,\varepsilon) - T\left(y(\alpha,\varepsilon)\right)\right]^{-\gamma} \left[1 - T'\left(y(\alpha,\varepsilon)\right)\right] = \mu(\alpha), \tag{27}$$

$$h(\alpha,\varepsilon)^{\sigma} = \mu(\alpha) \exp(\alpha+\varepsilon), \tag{28}$$

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

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 (27) with respect to ε. The right-hand side is independent of the insurable shock ε, and hence its derivative with respect to ε is zero. The derivative of the left-hand side of this equation with respect to ε is, by the chain rule,

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

The first term is nonzero by condition (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 (27) and (28) combine to deliver exactly the original intratemporal FOC with family-level taxation, namely, eq. (8). $\mathcal{Q.E.D.}$

8.3 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:

$$c(\alpha;\lambda,\tau) = \lambda(1-\tau)^{\frac{1-\tau}{1+\sigma}} \exp\left[(1-\tau)\alpha\right] \exp\left(\frac{1-\tau}{\sigma}\frac{v_\varepsilon}{2}\right), \tag{29}$$

$$h(\varepsilon;\tau) = (1-\tau)^{\frac{1}{1+\sigma}} \exp\left(\frac{-1}{\sigma^2}\frac{v_\varepsilon}{2}\right) \exp\left(\frac{\varepsilon}{\sigma}\right). \tag{30}$$

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

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

We can substitute these expressions into the planner's objective function in order to get an unconstrained optimization problem with one choice variable, τ. Differentiating the objective function with respect to τ then gives a first-order condition that maps θ and G into the optimal choice for 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

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

Substituting equations (29) and (30) into these expressions, the optimization problem can be rewritten as

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

where

$$G = g(1-\tau)^{\frac{1}{1+\sigma}} \exp\left(\frac{1}{\sigma}\frac{v_\varepsilon}{2}\right). \tag{31}$$

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 unconstrained optimization problem becomes[46]

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

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

$$\max_{\tau} \quad \frac{(1-\tau)}{\frac{\lambda_\alpha}{\lambda_\alpha+\theta}\exp\left[-\mu_\alpha\theta+\frac{\sigma_\alpha^2\theta^2}{2}\right]}\int \alpha\exp(-\theta\alpha)dF_\alpha - \log\left(\frac{\lambda_\alpha}{\lambda_\alpha-1+\tau}\right) - \mu_\alpha(1-\tau) - \frac{\sigma_\alpha^2(1-\tau)^2}{2} \\ + \log\left[(1-\tau)^{\frac{1}{1+\sigma}}\exp\left(\frac{1}{\sigma}\frac{v_\varepsilon}{2}\right) - G\right] - \frac{1-\tau}{1+\sigma}. \tag{32}$$

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 have

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

Lemma 3 *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 Ramsey 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

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

Substituting this expression into (32), the unconstrained optimization problem becomes

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

The first-order condition with respect to τ is

$$\begin{aligned}0 &= -\frac{1}{\lambda_\alpha+\theta}+\sigma_\alpha^2\theta+\frac{1}{1+\sigma}+\frac{1}{\lambda_\alpha-1+\tau}+\sigma_\alpha^2(1-\tau)+\frac{\frac{\partial}{\partial\tau}\left\{(1-\tau)^{\frac{1}{1+\sigma}}\exp\left(\frac{1}{\sigma}\frac{v_\varepsilon}{2}\right)\right\}}{(1-\tau)^{\frac{1}{1+\sigma}}\exp\left(\frac{1}{\sigma}\frac{v_\varepsilon}{2}\right)-G}\\ &= -\frac{1}{\lambda_\alpha+\theta}+\sigma_\alpha^2\theta+\frac{1}{1+\sigma}+\frac{1}{\lambda_\alpha-1+\tau}+\sigma_\alpha^2(1-\tau)-\frac{\left[1-\frac{G}{\exp\left(\frac{1}{\sigma}\frac{v_\varepsilon}{2}\right)(1-\tau)^{\frac{1}{1+\sigma}}}\right]^{-1}}{(1-\tau)(1+\sigma)}\end{aligned}\tag{33}$$

Substituting (31) 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 (19). $\mathcal{Q.E.D.}$

8.4 Computational Method

In this section, we briefly describe how we compute the optimal allocation in the baseline economy.

We numerically solve the Mirrlees planner's problem (12) for our discretized economy. We

Table 7: Polynomial Tax Functions

Tax System and Tax Parameters					Outcomes			
					ω (%)	ΔY (%)	$\overline{T'}$	Tr/Y
Baseline HSV								
			$\lambda: 0.836$	$\tau: 0.151$	–	–	0.311	0.018
Polynomial								
	τ_0	τ_1	τ_2	τ_3				
Proportional	–	0.178	–	–	−1.31	5.61	0.178	−0.021
Affine	−0.116	0.303	–	–	−0.58	0.41	0.303	0.089
Quadratic	−0.074	0.221	0.021	–	−0.10	0.61	0.295	0.051
Cubic	−0.032	0.126	0.064	−0.003	0.05	0.79	0.289	0.017
Mirrlees					0.11	0.82	0.287	0.003

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).[47] 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.

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, Weinzierl, and Yagan (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.

8.5 Results from Extensions to Richer Tax Structures

Table 7 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 10 is the analogue to Figure 6 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 8 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 wage multiplicatively), whereas the wealth effect associated with lump-sum taxes ensures that high κ households still work relatively hard.

[47]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.

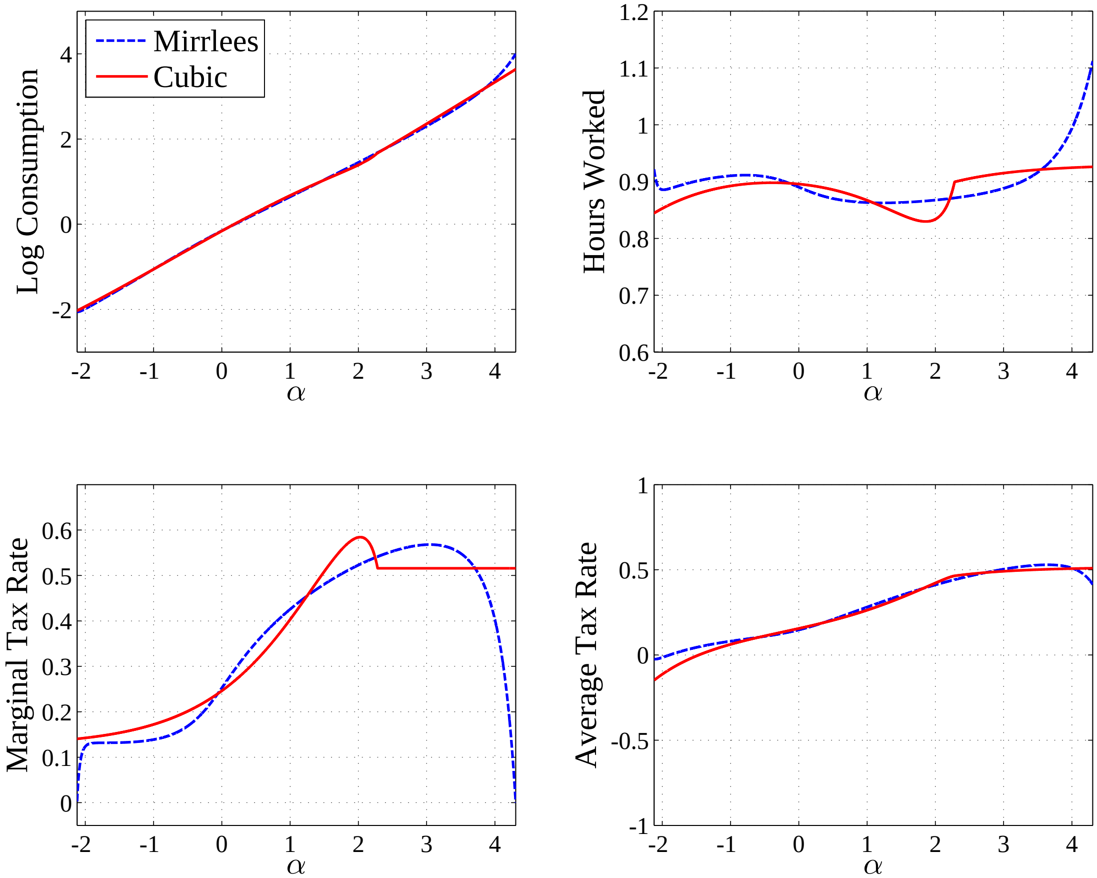

Figure 10: Cubic Tax Function. The figure contrasts allocations under the best-in-class cubic and Mirrlees tax systems. The top panels plot decision rules for consumption and hours worked, and the bottom panels plot marginal and average tax schedules. The plot for hours worked is for an agent with average ε.

Table 8: Type-Contingent Taxes

Tax System			Outcomes			
			ω (%)	ΔY (%)	$\overline{T'}$	Tr/Y
HSV^{US}	$\lambda : 0.827$	$\tau : 0.151$	–	–	0.311	0.017 0.015
HSV*	$\lambda^L : 0.988$ $\lambda^H : 0.694$	$\tau^L : 0.180$ $\tau^H : -0.059$	1.34	4.64	0.212	0.043 -0.061
Affine	$\tau_0^L : -0.140$ $\tau_0^H : 0.095$	$\tau_1^L : 0.151$ $\tau_1^H : 0.224$	1.39	5.20	0.199	0.126 -0.137
Mirrlees			1.46	5.20	0.200	0.103 -0.136