Optimal Income Taxation: Mirrlees Meets Ramsey*

Jonathan Heathcote
Federal Reserve Bank of Minneapolis

Hitoshi Tsujiyama
Goethe University Frankfurt

Dec, 2013
Preliminary and Incomplete

1 Introduction

In this paper we revisit a classic question in public finance: what structure of labor earnings taxation can maximize the social benefits of redistribution and public insurance while minimizing the social harm associated with distorting the allocation of labor input. There are two approaches to this question in the literature. The Mirrlees approach is to look for tax systems that maximize social welfare subject to the constraint that the planner cannot observe the relative contributions of individual productivity versus individual labor input in producing observed individual earnings. 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 views expressed herein are those of the authors and not necessarily those of the institutions with whom we are affiliated.

The alternative Ramsey approach is to restrict the planner to choose a tax schedule within a parametric class – for example, to restrict taxes to be a linear function of earnings. While there are no theoretical foundations for imposing ad hoc restrictions on the design of the tax schedule, the practical advantage of doing so is that one can then consider tax design in richer models.

In this paper we do two things. First, we build on the Mirrlees tradition but extend the framework in three directions that, in our view, allow us to provide quantitatively more relevant guidance on the welfare-maximizing shape of the tax function. Second, we show that constrained efficient allocations can be approximately implemented via very simple parametric tax functions, and thus that in practice the Mirrlees and Ramsey approaches to tax design can yield very similar prescriptions.

The standard Mirrlees approach is to set up a planner's problem where the planner maximizes a social welfare function subject to resource feasibility, and subject to incentive constraints such that individuals have incentives to truthfully reveal their productivity types. Given a solution to this problem, a tax schedule can then be inferred, such that the same allocations are decentralized as a competitive equilibrium given those taxes. Note that according to this decentralization, the government is providing all the insurance in the economy.

The first way in which we extend the standard Mirrlees environment is to assume that agents are able to privately insure a share of idiosyncratic labor productivity risk. We assume that the planner cannot directly observe these insurable shocks, or the associated insurance transfers. Given this hidden private insurance, the planner cannot make individual income or consumption a function of insurable shocks. Thus our environment contains distinct roles for both public and private insurance. 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.

Our second extension is to assume that individual labor productivity has a component that is observable by the planner, in addition to a component that the planner cannot observe directly. The logic is that wages vary systematically by observables such as age, gender, race, and education level. To the extent that the planner has some information about workers' productivities, a constrained efficient tax system should explicitly index taxes to these observables (see, e.g., Weinzierl (2011)).

The shape of the optimal tax schedule in any social insurance problem is inevitably sensitive to the assumed social welfare function. 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. Our third innovation relative to previous work in the Mirrlees tradition is to construct a social welfare function motivated by the tax system observed in US data. The logic is that the degree of progressivity built into the US tax and transfer system is informative about the government's taste for redistribution. We show that a summary statistic for the amount of progressivity embedded in the US system borrowed from Benabou (2000) and Heathcote, Storesletten, and Violante (2013) can be mapped parametrically into a summary statistic for the preference for redistribution in the planner's social welfare function. This *empirically-motivated social welfare function* will serve as our baseline objective function.

The environment in which we compare alternative tax systems is a simple model in which agents are heterogeneous only with respect to labor productivity. There are three orthogonal components to idiosyncratic labor productivity: $\log(w) = \alpha + \kappa + \varepsilon$. When individuals first enter the economy they draw two fixed effects. The first (α) is private information to the agent and cannot be privately insured – the standard Mirrlees assumptions. The second (κ) is observed by the planner, and captures differences in wages related to observables like age and education. Again, because this component is drawn before the agent can trade in financial markets, it cannot be insured privately. Agents then purchase explicit private insurance indexed to a third component of the wage (ε). The goal of the planner is to design a tax system that specifies the net taxes to be paid each agent, recognizing that agents will be privately insuring ε shocks in the background. The planner only observes κ and total individual income, which comprises labor earnings plus private insurance income.

Because the goal of the paper is to deliver quantitatively realistic prescriptions, we are carefully to replicate observed dispersion in US wages, and to decompose the overall variance of wages into the three model components described above. First, we set the variance of the observable fixed effect to reflect the amount of wage dispersion that can be accounted for by standard observables in a Mincer regression. Second, we set the variance of privately insurable shocks to the variance of wage shocks estimated to be privately insurable by Heathcote, Storesletten, and Violante (forthcoming).

After laying out the environment, and describing the program that defines constrained efficient allocations, we describe the general form of Ramsey problems in which the planner is forced to choose tax systems that belong to a particular parametric class – for example, affine tax schedules. We then describe how to use a particular ad hoc tax schedule – the power form considered by Heathcote, Storesletten, and Violante (2013) – to construct a mapping from observed progressivity to an implied social welfare function. Given this empirically-motivated welfare function we then tackle the Mirrlees optimal tax problem, and measure the potential welfare gains that could be attained by moving to a constrained efficient tax system. Our key findings are as follows.

First, in the conventional Mirrlees model, where α is the only component of wages and the planner maximizes a utilitarian social welfare function, there are very large welfare gains from tax reform: the gain from an efficient tax reform is equivalent to a 5.75 percent increase in all agents' consumption. At the same time, the tax system that decentralizes efficient allocations reduces hours worked and output by over 10 percent relative to the current tax schedule. The efficient tax schedule features large lump-sum transfers equal to around one quarter of per capita output.

The nature of optimal taxation changes dramatically when we extend the model to incorporate our alternative empirically-motivated social welfare function or to incorporate privately insurable shocks (ε). With the same simple model for wages – $\log(w) = \alpha$ – but the empirically-motivated social welfare function, the welfare gain from tax reform shrinks from 5.75 to 0.09 percent of consumption. Thus, the current tax system does not appear to be particularly inefficient. Introducing insurable shocks to wages reduces quite dramatically efficient marginal tax rates: from 54% to 43%, assuming a utilitarian welfare objective.

When we explore how nearly we can decentralize efficient allocations with simple parametric tax schedules, we compare two alternative ways to redistribute: an affine tax system in which marginal tax rates are constant, and all agents receive a lump-sum transfer, and a power system (a la Benabou (2000)) in which there are no lump-sum transfers, but where marginal tax rates increase with income.[1] The best system in the affine class system does not do particularly well. For example, in the model with a utilitarian planner and uninsurable and insurable shocks, the optimal tax system in the affine class delivers less than one third of the welfare gains from moving to the fully non-parametric income tax schedule that delivers efficient allocations. Moreover, our comparison of alternative tax functions indicates that it is more important to have marginal rates increase with income than to provide

[1] These are both two parameter functions. Taxes paid as a function of income according to the two systems are, respectively, $T(y) = \tau_0 + \tau_1 y$ and $T(y) = y - \lambda y^{1-\tau}$.

universal lump-sum transfers. These findings contrast with much of the existing literature which has argued that affine tax schedules can approximately decentralize efficient allocations.

In the next part of the paper we introduce the observable component of wages (κ). 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. Under an affine system with two observable types (think college versus high school), the type with higher observable productivity (college graduates) should face higher marginal tax rates, and receive smaller lump-sum transfers. Higher marginal rates on the more productive type allow the planner to redistribute across types, while smaller lump-sum transfers offset the disincentive effects of higher marginal tax rates on labor supply.

2 Environment

A unit mass of agents have identical preferences over consumption c, and work effort h. The utility function is separable between consumption and work effort and takes the form

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

Given this functional form, the Frisch elasticity of labor supply is $1/\sigma$.

Agents differ only with respect to labor productivity θ. Individual labor productivity has three orthogonal components

$$\log\theta = \alpha + \kappa + \varepsilon \tag{2}$$

These three idiosyncratic components differ with respect to whether or not they can be insured privately, and whether or not they are publicly observable. We assume that $\alpha \in \mathcal{A} \subset \mathbb{R}^+$ and $\kappa \in \mathcal{K} \subset \mathbb{R}^+$ represent shocks that cannot be insured privately, while perfect private insurance exists against shocks to $\varepsilon \in \mathcal{E} \subset \mathbb{R}^+$. We assume that κ is publicly observable while α and ε are not observed by the tax authority. We let the vector $(\alpha, \kappa, \varepsilon)$ denote an individual's type, and $F(\alpha)$, $F(\kappa)$ and $F(\varepsilon)$ denote the distributions for the three components.

The timing of events is as follows. Agents first draw α and κ. They then trade in a market in which they can purchase private insurance at actuarially fair prices against ε. Then each individual draws an ε, insurance pays out, and individuals choose how much to work. Finally, a government assesses taxes, and income less net taxes is consumed.

The price of insurance claims that will pay one unit of consumption if and only if $\varepsilon \in E \subset \mathcal{E}$ is $Q(E) = \int_E dF(\varepsilon)$. In the first stage of the period, prior to drawing ε, the budget constraint for an agent with uninsurable components α and κ is

$$\int B(\varepsilon; \alpha, \kappa) Q(\varepsilon) d\varepsilon = 0 \tag{3}$$

where $B(x; \alpha, \kappa)$ denotes the quantity (positive of negative) of insurance claims purchased that pay a unit of consumption if and only if the second stage draw for the insurable shock is x.

In the second stage of the period, income before taxes $y(\alpha, \kappa, \varepsilon)$ is the sum of labor earnings plus insurance payouts

$$y(\alpha, \kappa, \varepsilon) = \exp(\alpha + \kappa + \varepsilon) h(\alpha, \kappa, \varepsilon) + B(\varepsilon; \alpha, \kappa) \tag{4}$$

The tax authority observes only two individual level variables: the observable component of productivity κ, and total end of period income $y(\alpha, \kappa, \varepsilon)$. The tax authority does not directly observe α or ε, does not observe hours worked, and does not observe any of the trades $B(\cdot; \alpha, \kappa)$ associated with private insurance against ε. Taxes must be functions of observables. We let $T(y; \kappa)$ denote the income tax schedule, which may be indexed to κ. Given that it observes income taxes collected the planner also effectively observes consumption, since

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

One interpretation for the differential insurance assumption for α and κ versus ε is that α and κ represent fixed effects that are drawn before agents can participate in insurance markets. An alternative interpretation is that ε represents shocks that can be pooled within a family or other risk-sharing group, while α and κ are common across all members of the group but differ across groups.

One interpretation for the differential observability assumption for α and ε versus κ is that α and ε reflect components of productivity that are private information, while κ reflects the impact of observable characteristics on productivity, such as age, education, and gender.

While the model we describe is static it would be straightforward to develop a dynamic extension in which agents draw new values for the insurable shock ε in each period and in which the observable component κ evolves over time. A much more challenging extension would be to allow for persistent shocks to the unobservable non-insurable component of productivity α.

Aggregate output in the economy is simply aggregate effective labor supply

$$Y = \int\int\int \exp(\alpha + \kappa + \varepsilon) h(\alpha, \kappa, \varepsilon) dF(\alpha) dF(\kappa) dF(\varepsilon) \tag{6}$$

where $h(\alpha, \kappa, \varepsilon)$ denotes hours worked by an individual of type $(\alpha, \kappa, \varepsilon)$.

Aggregate output is divided between private consumption and a publicly-provided good that is allocated equally across all agents.

$$Y = \int\int\int c(\alpha, \kappa, \varepsilon) dF(\alpha) dF(\kappa) dF(\varepsilon) + G \tag{7}$$

The government must run a balanced budget, and thus the budget constraint is

$$\int\int\int T(y(\alpha,\kappa,\varepsilon);\kappa)dF(\alpha)dF(\kappa)dF(\varepsilon) = G \tag{8}$$

3 Constrained Efficient Allocations

In the Mirrlees formulation of the program that determines constrained efficient allocations, rather than thinking of the planner choosing taxes, we will instead think of the planner as choosing both consumption $c(\alpha,\kappa,\varepsilon)$ and income $y(\alpha,\kappa,\varepsilon)$ as functions of the individual types $(\alpha,\kappa,\varepsilon)$. 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 $(\alpha,\kappa,\varepsilon)$ an agent of that type weakly prefer to deliver to the planner the value for income $y(\alpha,\kappa,\varepsilon)$ the planner intends for that type, thereby receiving (after-taxes) the value for consumption $c(\alpha,\kappa,\varepsilon)$, rather than delivering any alternative level of income. It is easiest to formulate the Mirrlees problem with the planner inviting agents to report their unobservable characteristics α and ε, and then assigning the agent an allocation for income $y(\tilde{\alpha},\kappa,\tilde{\varepsilon})$ and consumption $c(\tilde{\alpha},\kappa,\tilde{\varepsilon})$ on the basis of their reports $\tilde{\alpha}$ and $\tilde{\varepsilon}$ (recall that the planner observes κ directly). 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 income to consumption. Such a schedule can be decentralized via a tax schedule on income y of the form $T(y;\kappa)$ that defines how rapidly consumption grows with income.[2]

[2] Note that some values for income might not feature in the menu offered by the Mirrlees planner. Those values will not be chosen in decentralization with income taxes as long as income at those values is taxed sufficiently heavily. For example, suppose the lowest value for income in the Mirrlees solution is $\underline{y}$, with corresponding consumption value $\underline{c}$. Lower

The planner maximizes a social welfare function, $W(\alpha, \kappa, \varepsilon)$, with weights that potentially vary with α, κ and ε.[3] The timing with the period is as follows. In the first stage, agents draw (α, κ). They make a report $\hat{\alpha}(\alpha, \kappa)$ to the planner where the reporting function $\hat{\alpha} : \mathcal{A} \times \mathcal{K} \to \mathcal{A}$. At the same time, agents buy private insurance against ε, $B(\cdot; \alpha, \kappa, \tilde{\alpha})$. In the second stage, agents draw ε, and insurance pays out. Agents make a report $\hat{\varepsilon}(\alpha, \kappa, \varepsilon)$ to the planner where the second stage reporting function $\hat{\varepsilon} : \mathcal{A} \times \mathcal{K} \times \mathcal{E} \to \mathcal{E}$. Finally, agents work sufficient hours to deliver $y(\hat{\alpha}(\alpha, \kappa), \kappa, \hat{\varepsilon}(\alpha, \kappa, \varepsilon))$, and receive consumption $c(\hat{\alpha}(\alpha, \kappa), \kappa, \hat{\varepsilon}(\alpha, \kappa, \varepsilon))$.

3.1 Second stage agent's problem

As a first step towards characterizing efficient allocations, we start with the agent's problem in the final stage, given generic reporting strategies $\hat{\alpha}$ and $\hat{\varepsilon}$. Let $\tilde{\alpha} = \hat{\alpha}(\alpha, \kappa)$. Agents choose insurance purchases and labor supply to solve the following problem

$$\max_{h(\alpha,\kappa,\cdot,\tilde{\alpha};\hat{\varepsilon}),B(\cdot;\alpha,\kappa,\tilde{\alpha};\hat{\varepsilon})} \int u(\alpha, \kappa, \varepsilon, \tilde{\alpha}; \hat{\varepsilon}) dF(\varepsilon) \tag{9}$$

where

$$u(\alpha, \kappa, \varepsilon, \tilde{\alpha}; \hat{\varepsilon}) = \frac{c(\tilde{\alpha}, \kappa, \hat{\varepsilon}(\alpha, \kappa, \varepsilon))^{1-\gamma}}{1-\gamma} - \frac{h(\alpha, \kappa, \varepsilon, \tilde{\alpha}, \hat{\varepsilon}(\alpha, \kappa, \varepsilon))^{1+\sigma}}{1+\sigma}$$

income values can be ruled in the competitive equilibrium by assuming that marginal tax rates are 100% for income below $\underline{y}$ and that conditional on delivering at least $\underline{y}$ agents receive a lump-sum transfer equal to $\underline{c}$.

[3] We will show later that the planner will be unable to offer contracts in which income or consumption vary with ε, and given that result we will focus on a social welfare function in which weights do not vary with ε.

subject to

$$\begin{aligned}\int B(\varepsilon;\alpha,\kappa,\tilde{\alpha};\hat{\varepsilon})Q(\varepsilon)d\varepsilon &= 0 \qquad (10)\\ \exp(\alpha+\kappa+\varepsilon)h(\alpha,\kappa,\varepsilon,\tilde{\alpha},\hat{\varepsilon}(\alpha,\kappa,\varepsilon))+B(\varepsilon;\alpha,\kappa,\tilde{\alpha};\hat{\varepsilon}) &= y(\tilde{\alpha},\kappa,\hat{\varepsilon}(\alpha,\kappa,\varepsilon))\end{aligned}$$

Let $v(\alpha,\kappa,\varepsilon,\tilde{\alpha},\hat{\varepsilon}(\alpha,\kappa,\varepsilon))$ denote maximum utility after drawing ε given reports $\tilde{\alpha}$ and $\hat{\varepsilon}(\alpha,\kappa,\varepsilon)$.

3.2 First stage planner's problem

The planner maximizes $W(\alpha,\kappa,\varepsilon)$ subject to the resource constraint, and incentive constraints that ensure that utility from reporting α and ε truthfully and receiving the associated allocation is weakly larger than expected welfare from any alternative report and associated allocation. Formally. the planner solves

$$\max_{\{c(\alpha,\kappa,\varepsilon),y(\alpha,\kappa,\varepsilon)\}}\int\int\int W(\alpha,\kappa,\varepsilon)v(\alpha,\kappa,\varepsilon,\alpha,\varepsilon)dF(\alpha)dF(\kappa)dF(\varepsilon) \qquad (11)$$

subject to

$$\int\int\int c(\alpha,\kappa,\varepsilon)dF(\alpha)dF(\kappa)dF(\varepsilon)+G=\int\int\int y(\alpha,\kappa,\varepsilon)dF(\alpha)dF(\kappa)dF(\varepsilon) \qquad (12)$$

and subject to allocations being incentive compatible.

$$\int v(\alpha,\kappa,\varepsilon,\alpha,\varepsilon)dF(\varepsilon)\geq\int v(\alpha,\kappa,\varepsilon,\tilde{\alpha},\varepsilon)dF(\varepsilon) \qquad \forall\alpha,\kappa \text{ and } \forall\tilde{\alpha} \qquad (13)$$

$$v(\alpha,\kappa,\varepsilon,\tilde{\alpha},\varepsilon)\geq v(\alpha,\kappa,\varepsilon,\tilde{\alpha},\tilde{\varepsilon}) \qquad \forall\alpha,\kappa,\varepsilon,\tilde{\alpha} \text{ and } \forall\tilde{\varepsilon} \qquad (14)$$

where $v(\alpha,\kappa,\varepsilon,\tilde{\alpha},\tilde{\varepsilon})$ is the utility achieved by an agent who follows reporting strategies $\hat{\alpha}$ and $\hat{\varepsilon}$, who draws $(\alpha,\kappa,\varepsilon)$, and who makes reports $\tilde{\alpha}=\hat{\alpha}(\alpha,\kappa)$ and $\tilde{\varepsilon}=\hat{\varepsilon}(\alpha,\kappa,\varepsilon)$.

The second set of incentive constraints (14) imposes that agents weakly prefer to report ε truthfully for any report $\tilde{\alpha}$ in the first stage. This ensures that truth-telling $\tilde{\varepsilon}(\alpha, \kappa, \varepsilon) = \varepsilon$ is incentive compatible for any first stage strategy $\hat{\alpha}$. The first set of incentive constraints (13) impose that agents weakly prefer to report α truthfully, assuming truthful reporting in the second stage. Thus truth-telling $\hat{\alpha}(\alpha, \kappa) = \alpha$ is incentive compatible at the first stage.

3.3 Preliminary result: allocations cannot be conditioned on ε

We start by assuming agents follow arbitrary reporting strategies and compute the implied equilibrium allocations and utility. We show that for any reporting strategies, hours worked are not a function of the report $\tilde{\varepsilon}$. It follows that a truth-telling reporting strategy is incentive compatible if and only if consumption is also independent of the report $\tilde{\varepsilon}$. It follows further that there is nothing to be gained from the planner asking agents to report ε – since neither component of individual welfare (and hence social welfare) can be made contingent on those reports.

Let $\lambda(\alpha, \kappa, \tilde{\alpha}; \hat{\varepsilon})$ denote the multiplier on the first constraint in the agent's problem (maximize 9 subject to 10), where $\tilde{\alpha} = \hat{\alpha}(\alpha, \kappa)$. This multiplier cannot be a function of ε because the constraint applies before ε is drawn. Let $\rho(\alpha, \kappa, \varepsilon, \tilde{\alpha}, \tilde{\varepsilon})$ denote the multiplier on the second constraint given a draw ε, where $\tilde{\varepsilon} = \hat{\varepsilon}(\alpha, \kappa, \varepsilon)$. The first-order conditions imply

$$h(\alpha, \kappa, \varepsilon, \tilde{\alpha}) = [\exp(\alpha + \kappa + \varepsilon)\lambda(\alpha, \kappa, \tilde{\alpha}; \hat{\varepsilon})]^{\frac{1}{\sigma}} \tag{15}$$

The crucial thing to note here is that hours are independent of the report $\tilde{\varepsilon} = \hat{\varepsilon}(\alpha, \kappa, \varepsilon)$. The logic for this result is that agents exploit insurance markets to deliver target income $y(\tilde{\alpha}, \kappa, \tilde{\varepsilon})$ at minimum cost in terms of labor effort. That implies that they choose hours as a function of the draw for ε to make marginal disutility of hours $h(\alpha, \kappa, \varepsilon, \tilde{\alpha})^{\sigma}$ proportional to the wage $\exp(\alpha+\kappa+$

ε), – a familiar complete markets result. The wage is obviously independent of $\tilde{\varepsilon}$ and the constant of proportionality is too – it is the multiplier on the insurance purchases constraint, which applies prior to ε and thus $\tilde{\varepsilon}$ being realized.

Consider an agent of type $(\alpha, \kappa, \varepsilon)$ who is following an arbitrary reporting strategy in the first stage, so that $\tilde{\alpha} = \hat{\alpha}(\alpha, \kappa)$. In the second stage, after drawing ε, this agent will want to report the value $\tilde{\varepsilon}$ whose associated contract $\{c(\tilde{\alpha}, \kappa, \tilde{\varepsilon}), y(\tilde{\alpha}, \kappa, \tilde{\varepsilon})\}$ maximizes utility. Because hours worked, and thus disutility from hours, are independent of $\tilde{\varepsilon}$, it follows that the agent will choose whichever contract offers the highest value for $c(\tilde{\alpha}, \kappa, \tilde{\varepsilon})$. Thus truth telling is incentive compatible if and only if consumption is independent of the individual report $\tilde{\varepsilon}$. Thus neither consumption nor hours can be made contingent on the report of ε in any truth-telling allocation. Thus contracts must be of the form $\{c(\tilde{\alpha}, \kappa), y(\tilde{a}, \kappa)\}$. Since these contracts are independent of $\tilde{\varepsilon}$ we can envision these contracts being offered prior to ε being revealed.

Note that because neither consumption nor income depend on ε in the constrained efficient allocation, the income or consumption tax system that decentralizes efficient allocations is also independent of ε. We have derived this result endogenously, given the assumption that the planner cannot observe ε. An alternative approach that gives the same outcome

3.4 Restatement of Mirrlees problem

We now apply this result to produce a much simpler representation of the program that defines constrained efficient allocations. We will drop the dependence of consumption and income on the insurable shock. In addition, because the planner weights do not vary with ε, and because we do not need to worry about the second set of incentive constraints (14), we can also dispense with keeping track of private insurance markets altogether, and work with expected utility (conditional on α and κ) prior to drawing ε.

From the agent's budget constraint we can solve for the multiplier on the insurance purchases constraint, $\lambda(\alpha,\kappa,\tilde{\alpha})$, and thus simplify the expression for hours $h(\alpha,\kappa,\varepsilon,\tilde{\alpha})$. Substituting (4) into (3) and then using (15) to substitute for hours gives

$$\lambda(\alpha,\kappa,\tilde{\alpha}) = \left(\frac{y(\tilde{\alpha},\kappa)}{\int \exp(\alpha+\kappa+\varepsilon)^{\frac{1+\sigma}{\sigma}} dF(\varepsilon)} \right)^{\sigma}$$

Substituting this into (15) gives

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

Thus utility, given type $(\alpha,\kappa,\varepsilon)$ and report $\tilde{\alpha}$, is

$$v(\alpha,\kappa,\varepsilon,\tilde{\alpha}) = \frac{c(\alpha,\kappa,\tilde{\alpha})^{1-\gamma}}{1-\gamma} - \frac{1}{1+\sigma} \left(\frac{\exp(\alpha+\kappa+\varepsilon)^{\frac{1}{\sigma}} y(\tilde{\alpha},\kappa)}{\int \exp(\alpha+\kappa+\varepsilon)^{\frac{1+\sigma}{\sigma}} dF(\varepsilon)} \right)^{1+\sigma}$$

Expected utility, prior to drawing ε, is

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

Let $U(\alpha,\kappa,\tilde{\alpha}) = \int v(\alpha,\kappa,\varepsilon,\tilde{\alpha}) dF(\varepsilon)$. We can now restate the planner's problem that defines constrained efficient allocations.

$$\max_{\{c(\alpha,\kappa),y(\alpha,\kappa)\}} \int \int W(\alpha,\kappa) U(\alpha,\kappa,\alpha) dF(\alpha) dF(\kappa) \tag{16}$$

subject to

$$\int \int c(\alpha,\kappa) dF(\alpha) dF(\kappa) + G = \int \int y(\alpha,\kappa) dF(\alpha) dF(\kappa) \tag{17}$$

and

$$U(\alpha, \kappa, \alpha) \geq U(\alpha, \kappa, \tilde{\alpha}) \qquad \forall \alpha, \kappa \text{ and } \forall \tilde{\alpha} \tag{18}$$

Note that ε no longer appears anywhere in this problem, and nor does the second stage agent's problem. The problem is identical to a standard static Mirrlees type problem, where the planner faces a distribution of agents with heterogeneous unobserved productivity α. We will solve this problem numerically. Note, however, that the period utility function for each agent in this problem is not identical to the utility function we started with in the underlying problem, since the weight on hours worked is now $\left(\int \exp(\varepsilon)^{\frac{1+\sigma}{\sigma}} dF(\varepsilon)\right)^{-\sigma}$.

3.5 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.

4 Ramsey Policies

We will compare outcomes and welfare under the solution to the Mirrlees described above to those under alternative ad hoc tax systems. One comparison of particular interest will be the US tax system. We will consider taxes based on individual earnings, and taxes based on income.

Consider a decentralized economy with income taxes $T(y; \kappa)$ and private insurance markets against ε shocks. Knowing α and κ, but before drawing ε, agents solve

$$\max_{c(\alpha,\kappa,\varepsilon), h(\alpha,\kappa,\varepsilon), B(\alpha,\kappa,\varepsilon)} \int \left(\frac{c(\alpha, \kappa, \varepsilon)^{1-\gamma}}{1-\gamma} - \frac{h(\alpha, \kappa, \varepsilon)^{1+\sigma}}{1+\sigma} \right) dF(\varepsilon) \tag{19}$$

subject to equations (3), (5) and (4).

The first-order conditions are

$$\begin{aligned} c(\alpha,\kappa,\varepsilon)^{-\gamma}\left[1-T'\left(y(\alpha,\kappa,\varepsilon);\kappa\right)\right]\exp(\alpha+\kappa+\varepsilon) &= h(\alpha,\kappa,\varepsilon)^{\sigma} \\ c(\alpha,\kappa,\varepsilon)^{-\gamma}\left[1-T'\left(y(\alpha,\kappa,\varepsilon);\kappa\right)\right] &= \lambda(\alpha,\kappa) \end{aligned} \tag{20}$$

where $\lambda(\alpha,\kappa)$ is the multiplier on the insurance purchases constraint.

Proposition. Assume that the tax function satisfies

$$T''(y;\kappa) > -\gamma\frac{(1-T'(y;\kappa))}{(y-T(y;\kappa))}$$

for all feasible y. Then income, taxes and consumption are independent of ε.[4]

Note that this condition is automatically satisfied for any affine tax schedule. Suppose the condition is satisfied. Then the first-order condition for labor supply can be written as

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

where $y(\alpha,\kappa)$ can be solved for from the budget constraint

$$\int B(\alpha,\kappa,\varepsilon)dF(\varepsilon) = \int\left[y(\alpha,\kappa)-\exp(\alpha+\kappa+\varepsilon)h(\alpha,\kappa,\varepsilon)\right]dF(\varepsilon) = 0$$

[4] **Proof**. Subsituting **??** into 20 gives

$$\left[y(\alpha,\kappa,\varepsilon)-T(y(\alpha,\kappa,\varepsilon);\kappa)\right]^{-\gamma}\left[1-T'(y(\alpha,\kappa,\varepsilon);\kappa)\right] = \lambda(\alpha,\kappa)$$

The derivative of the left-hand side of this equation with respect to the insurable shock ε is, by the Chain Rule

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

The first term is strictly negative iff the condition is satisfied, which immediately implies that $\frac{\partial y}{\partial \varepsilon}=0$.

4.1 Earnings taxes

Let $x(\alpha,\kappa,\varepsilon) = \exp(\alpha+\kappa+\varepsilon)h(\alpha,\kappa,\varepsilon)$ denote labor earnings. Consider a decentralized economy with earnings taxes $T(x(\alpha,\kappa,\varepsilon);\kappa)$ and private insurance markets against ε shocks. Knowing α and κ, but before drawing ε, agents solve a very similar problem to the one described above with income taxes.

The first order conditions are

$$\begin{aligned} c(\alpha,\kappa,\varepsilon)^{-\gamma}\left[1-T'(x(\alpha,\kappa,\varepsilon);\kappa\right]\exp(\alpha+\kappa+\varepsilon) &= h(\alpha,\kappa,\varepsilon)^{\sigma} \qquad (21) \\ c(\alpha,\kappa,\varepsilon)^{-\gamma} &= \mu(\alpha,\kappa) \end{aligned}$$

In this case it is immediate from the second first-order condition that consumption is independent of ε. Hours worked are implicitly given by

$$h(\alpha,\kappa,\varepsilon) = c(\alpha,\kappa)^{-\gamma}\exp(\alpha+\kappa+\varepsilon)\left[1-T'\left(\exp(\alpha+\kappa+\varepsilon)h(\alpha,\kappa,\varepsilon);\kappa\right)\right]^{\frac{1}{\sigma}}$$

while the budget constraints can be combined to give

$$c(\alpha,\kappa) = \int\left[\exp(\alpha+\kappa+\varepsilon)h(\alpha,\kappa,\varepsilon) - T(\exp(\alpha+\kappa+\varepsilon)h(\alpha,\kappa,\varepsilon);\kappa)\right]dF(\varepsilon) = 0$$

4.1.1 Affine taxes

Given an affine tax schedule, it is possible to characterize equilibrium allocations more sharply. Suppose earnings taxes are affine, with (potentially) $\kappa-$contingent intercept and slope, so that

$$T(x;\kappa) = \tau_0(\kappa) + \tau_1(\kappa)x$$

Then we have an explicit solution for hours worked, as a function of the wage and consumption.

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

Note that under an affine tax structure, equilibrium allocations for consumption and hours worked are identical for income versus earnings taxes, assuming the coefficients $\tau_0(\kappa)$ and $\tau_1(\kappa)$ are the same in the two systems. The logic is that the only difference between the corresponding intra-temporal FOCs is that in one case the relevant marginal tax rate is the rate on income while in the other it is the rate on earnings.

4.2 Decentralization of constrained efficient allocations

Let $c^*(\alpha, \kappa)$ and $y^*(\alpha, \kappa)$ denote consumption and income in the constrained efficient allocation and let $h^*(\alpha, \kappa, \varepsilon)$ denote the constrained efficient allocation rule for hours (all given a particular social welfare function $W(\alpha, \kappa)$):

$$h^*(\alpha, \kappa, \varepsilon) = \frac{y^*(\alpha, \kappa)}{\exp(\alpha + \kappa) \int \exp(\varepsilon)^{\frac{1+\sigma}{\sigma}} dF(\varepsilon)} \exp(\varepsilon)^{\frac{1}{\sigma}}$$

Substituting constrained efficient allocations into the competitive equilibrium first-order condition for hours worked in the economy with income taxes (20) implicitly defines the marginal tax rates that decentralize constrained efficient allocations:

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

Note that marginal tax rates do not vary with ε because income (including insurance payouts) does not vary with ε. Note also that while the decentralization above is based on income taxes, it is clear that the Mirrlees solution could equivalently be decentralized using consumption taxes. In that case we would get

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

Note, finally, that we cannot decentralize the Mirrlees solution using taxes on individual earnings, because in the competitive equilibrium individual earnings – and thus taxes – vary with ε, while in the solution to the Mirrlees planner taxes paid (which are equal to income minus consumption) are independent of ε and vary only with the components α and κ.

5 Social Welfare Function

We now describe our methodology for using the degree of progressivity built into the actual tax system to infer social preferences. We will assume the social welfare function takes the form

$$W(\alpha,\kappa) = \frac{\exp(-\omega\,(\alpha+\kappa))}{\int\int \exp(-\omega\,(\alpha+\kappa))dF(\alpha)dF(\kappa)} \tag{22}$$

where the parameter ω controls the extent to which the planner puts relatively more or less weight or relatively high productivity workers. We implicitly assume that the planner puts equal weight on agents with different realizations for ε. The logic for this is that, as explained above, constrained efficient allocations cannot be made to vary with ε. One motivation for choosing a social welfare function that treats the components α and κ symmetrically is that for any non-κ contingent tax system based on individual earnings or income, only the total uninsurable component of productivity (i.e. $\alpha+\kappa$) is relevant for individual choices and individual welfare.

Heathcote, Storesletten, and Violante (2013) argue that the following earnings tax function offers a reasonable approximation to the US tax and transfer system

$$T\,(y) = y - \lambda y^{1-\tau} \tag{23}$$

Thus the marginal tax rate on individual income is given by

$$T'(y) = 1 - \lambda(1-\tau)(y)^{-\tau}$$

Suppose the US government is solving the following Ramsey problem

$$\max_{T(y(\alpha,\kappa,\varepsilon);\kappa)} \int\int W(\alpha,\kappa)\left(\int u\left(c(\alpha,\kappa,\varepsilon),h(\alpha,\kappa,\varepsilon)\right)dF(\varepsilon)\right)dF(\alpha)dF(\kappa)$$

subject to

1. $c(\alpha,\kappa,\varepsilon)$ and $h(\alpha,\kappa,\varepsilon)$ are solutions to the household's problem **??** given the tax function $T(y(\alpha,\kappa,\varepsilon);\kappa)$.

2. The government budget constraint is satisfied given an exogenous value for public expenditure G

$$\int\int\int T(y(\alpha,\kappa,\varepsilon);\kappa)dF(\varepsilon)dF(\alpha)dF(\kappa) = G.$$

3. The tax function is in the two-parameter class described by 23.

4. The social welfare function takes the form described by 22.

Note that while in principle the government chooses two tax parameters, λ and τ, because it has to respect the government budget constraint, the government effectively has a single choice variable, τ. Let $\tau^*(\omega)$ denote the welfare maximizing choice for τ given a social welfare function indexed by ω.

The logic underlying our empirically-motivated social welfare function is that while we do not directly observe ω, we do observe the degree of progressivity chosen by the US political system. Let τ^{US} denote that value for τ. We then infer that US social preferences ω^{US} must satisfy

$$\tau^*(\omega^{US}) = \tau^{US} \tag{24}$$

and use this equation to reverse engineer ω^{US} and thus a social welfare function.

We now describe how we operationalize this.

Substituting the tax function 23 into the first-order conditions for an economy with earnings taxes (20) gives

$$h(\alpha,\kappa,\varepsilon) = \left[c(\alpha,\kappa)^{-\gamma-\frac{\tau}{1-\tau}} \exp(\alpha+\kappa+\varepsilon)(1-\tau)\lambda^{\frac{1}{1-\tau}} \right]^{\frac{1}{\sigma}} \tag{25}$$

The budget constraint is

$$c(\alpha,\kappa) = \lambda y(\alpha,\kappa,\varepsilon)^{1-\tau}$$

Since this applies for every ε,

$$c(\alpha,\kappa)^{\frac{1}{1-\tau}} = \int \lambda^{\frac{1}{1-\tau}} y(\alpha,\kappa,\varepsilon) dF(\varepsilon)$$

which, using the definition for income $y(\alpha,\kappa,\varepsilon)$ and substituting in the expression for hours, can be expressed as

$$c(\alpha,\kappa) = \lambda^{\frac{1+\sigma}{\sigma+\tau+\gamma(1-\tau)}} (1-\tau)^{\frac{1-\tau}{\sigma+\tau+\gamma(1-\tau)}} \exp\left(\frac{(1+\sigma)(1-\tau)}{\sigma+\tau+\gamma(1-\tau)} (\alpha+\kappa) \right) \left[\int \exp\left(\frac{1+\sigma}{\sigma} \varepsilon \right) dF(\varepsilon) \right]^{\frac{\sigma(1\cdots}{\sigma+\tau+\gamma\cdots}} \tag{26}$$

Given expressions 25 and 26, and using the government budget constraint to solve for λ, we can compute social welfare for any values for τ and ω. When we calibrate the model we will search numerically for a value for ω that satisfies equation 24.

It is instructive to consider a special case, in which the function $\tau^*(\omega)$ can be characterized in closed form. Consider the special case in which utility is logarithmic in consumption, so $\gamma = 1$. Assume, in addition, that all shocks are normally distributed: $\alpha\tilde{}N(-v_\alpha/2, v_\alpha)$, $\kappa\tilde{}N(-v_\alpha/2, v_\alpha)$ and $\varepsilon\tilde{}N(-v_\alpha/2, v_\alpha)$ is normally distributed with variance v_ε and mean $-v_\varepsilon/2$.

Then

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

$$h(\alpha,\kappa,\varepsilon;\tau) = (1-\tau)^{\frac{1}{1+\sigma}} \exp\left(-\frac{(1-\tau)}{\sigma+\tau}\left(\frac{1-2\tau-\sigma\tau}{\sigma+\tau}\right)\frac{v_\varepsilon}{2}\right) \exp\left(\frac{1-\tau}{\sigma+\tau}\varepsilon\right)$$

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

We can substitute these expressions into the planner's objective function, to give an unconstrained optimization problem with one choice variable, τ. Given the social welfare function 22 the planner's problem is

$$\max_{\tau}\left\{\int\int \frac{\exp(-\omega(\alpha+\kappa))}{\int\int \exp(-\omega(\alpha+\kappa))dF(\alpha)dF(\kappa)}\left(\log\left(c(\alpha,\kappa;\tau)\right) - \int \frac{h(\alpha,\kappa,\varepsilon;\tau)^{1+\sigma}}{1+\sigma}dF(\varepsilon)\right)dF(\alpha)dF\right.$$

Substituting in the expressions above for $c(\alpha,\kappa;\lambda,\tau)$, $h(\alpha,\kappa,\varepsilon;\tau)$ and $\lambda(\tau)$ and differentiating with respect to τ gives a first-order condition that maps ω and G into the optimal choice for τ. In the appendix we show that the first order condition is

$$\frac{1}{1-g(G,\tau)}\frac{-1}{(1+\sigma)(1-\tau)} + (v_\alpha+v_\kappa)(1+\omega-\tau) + \frac{1}{1+\sigma} = 0$$

where $g(G,\tau) = G \times Y(\tau)$ and $Y(\tau)$ denotes aggregate output.

Let $*$'s denote observed variables. If the US political system is solving this problem, we can use this first-order condition in conjunction with the observed choices for G and τ to infer ω. In particular, since $g(G^*,\tau^*) = G^* \times Y(\tau^*) = g^*$ we infer that

$$\omega^* = -(1-\tau^*) + \frac{1}{v_\alpha+v_\kappa}\frac{1}{1+\sigma}\left[\frac{1}{(1-g^*)(1-\tau^*)} - 1\right]$$

Note that ω^* is increasing in τ^*, as expected. Thus if we observe more progressive taxation, we can infer that the social planner puts less weight on higher wage individuals. Holding fixed observed τ^*, $\frac{\partial \omega^*}{\partial v_\alpha} < 0$ and $\frac{\partial \omega^*}{\partial v_\kappa} < 0$. Thus more uninsurable risk but the same tax progressivity means we can infer the planner has less desire to redistribute. Similarly, holding fixed observed τ^*, $\frac{\partial \omega^*}{\partial \sigma} < 0$, meaning that for the same observed τ, the less elastic we think labor supply is (and thus the smaller the distortions associated with progressive taxation) the less desire to redistribute we should attribute to the planner. Finally $\frac{\partial \omega^*}{\partial g^*} > 0$, meaning that, holding fixed progressivity, the larger the share of output devoted to public goods, the more the planner wants to redistribute. The logic here is that tax progressivity tends to reduce labor supply, making it more difficult to finance public goods, so governments that need to finance large expenditure will tend to choose less progressivity – unless they have a strong desire to redistribute.

6 Calibration

This section illustrates the calibration of this economy.

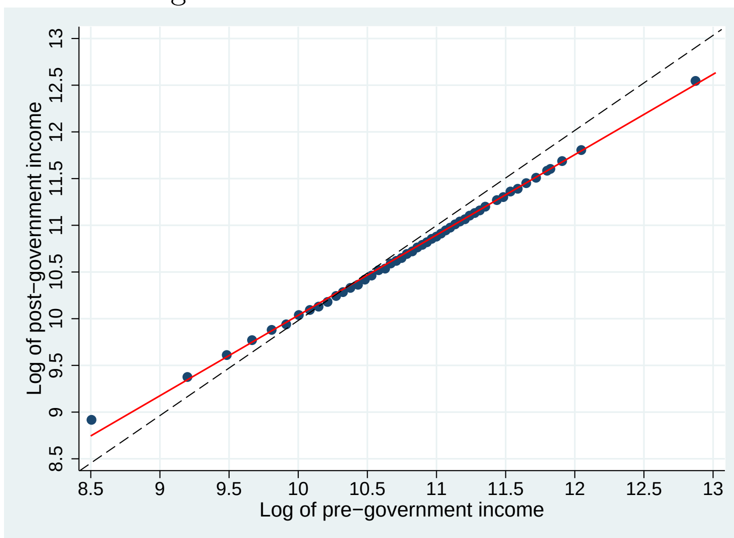
6.1 Preference

The agents' utility is given by

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

As a baseline parametrization, we assume the relative risk aversion being unity $\gamma = 1$, i.e., log preference for consumption. We choose $\sigma = 2$ so that the Frisch elasticity is 0.5.

Figure 1: Fit of HSV tax function



6.2 Tax System and Government Expenditure

We regard the U.S. income tax system is well-represented by a particular tax function

$$T(y; \lambda, \tau) = y - \lambda y^{1-\tau}$$

which is first introduced by Benabou (2000). Heathcote, Storesletten, and Violante (2013) estimate the progressivity parameter in this baseline tax function as $\tau = 0.151$. Figure 1 shows log of the pre- and post-government income of the U.S. households and one can see it fits the actual tax system fairly well. The remaining parameter λ is pinned down by the resource feasibility. We call this tax function as the HSV tax function.

The government consumption and investment constitute 18.8% of GDP in 2005 in the U.S., and hence we set government spending G such that $G/Y = 0.188$.

6.3 Wage distribution

The baseline model has three independent shocks, α, ε and κ. We need to estimate the variance of these shocks.

The variance of insurable shocks v_ε is given by 0.193, the estimate from Heathcote, Storesletten, and Violante (forthcoming). Carefully taking into account the measurement errors, they also estimate the total variance of wages to be 0.466.

For the variance of observable shocks v_κ, we use the estimate in Heathcote, Perri, and Violante (2010) who estimate the variance of cross sectional wage dispersion attributable to observables to be 0.108. We can therefore estimate the variance of uninsurable shocks which are not observable as the residual and is then given by $v_\alpha = 0.165$.

The bounds for α are given by

$$\exp(\alpha) \in \left[\frac{5.15/2}{19.60}, \frac{2,138.16}{19.60} \right].$$

Specifically, we assume the minimum productivity level is the half of the Federal minimum wage \$5.15 in 2005. Taking into account the fat right tail of the wage distribution (Saez (2001)), we assume the maximum level of the productivity is given by \$2,138.16, the earnings per hour at 99.99th percentile of 2005 earnings distribution from Piketty and Saez (2003). We assume annual hours worked are 2000 to calculate this number. Both bounds are normalized by \$19.60, the average hourly earnings in 2005 taken from BLS data.

Saez (2001) argues that the wage distribution has a thick right tail and is well-approximated by a Pareto distribution. We address this issue by assuming that the distribution being log-normal for $exp(\alpha) \leq x$ and being Pareto for $exp(\alpha) > x$ where $x = 1.77$ so that 95% of the population lie in the log-normal range and 5% lie in the Pareto range. This value for x corresponds to about \$34.6. For the Pareto parameter, we use 2.0, estimated from Piketty and Saez (2003). The shape of the wage distribution is found in Figure 2.

We assume two-point equal-weight distribution for κ. This gives $\exp(\kappa_{high})/\exp(\kappa_{low}) = 1.93$. We also assume that the uninsurable shocks ε are drawn from a continuous normal distribution $N(-\frac{v_\varepsilon}{2}, v_\varepsilon)$.

We use 10,000 evenly spaced grid points in the baseline.

6.4 Social Welfare

We estimate the empirically motivated social welfare function. Specifically, we estimate the welfare parameter ω^* described in Section 5 using the U.S. tax data. The resulting weight is illustrated in Figure 3. It shows that the relative weights are increasing in wage, which might be intuitive given the low marginal weight in the U.S. economy.

7 Quantitative Analysis

In this section, we compare outcomes and welfare under the solution to the Mirrlees described above to those under alternative ad hoc tax systems. One comparison of particular interest is with the US tax system. Welfare measure we use here is the consumption equivalent variation.

We are interested in computing the welfare gains in moving to tax policies that are optimal when we restrict the tax code to take a simple functional form. In particular, we consider the HSV income tax as our baseline economy. We also consider polynomial tax schedules, starting with affine tax schedules, and then moving to add quadratic and cubic terms. For each functional form we consider, we search for welfare-maximizing values for the parameters that define the tax schedule[5]. We consider both tax schedules that explicitly

[5]For higher order polynomial tax functions, we impose a restriction that marginal rates become constant above ten times of average income. This cutoff value corresponds to 99.5th percentile of income distribution. This is for computational purpose and does

Figure 2: Wage Distribution

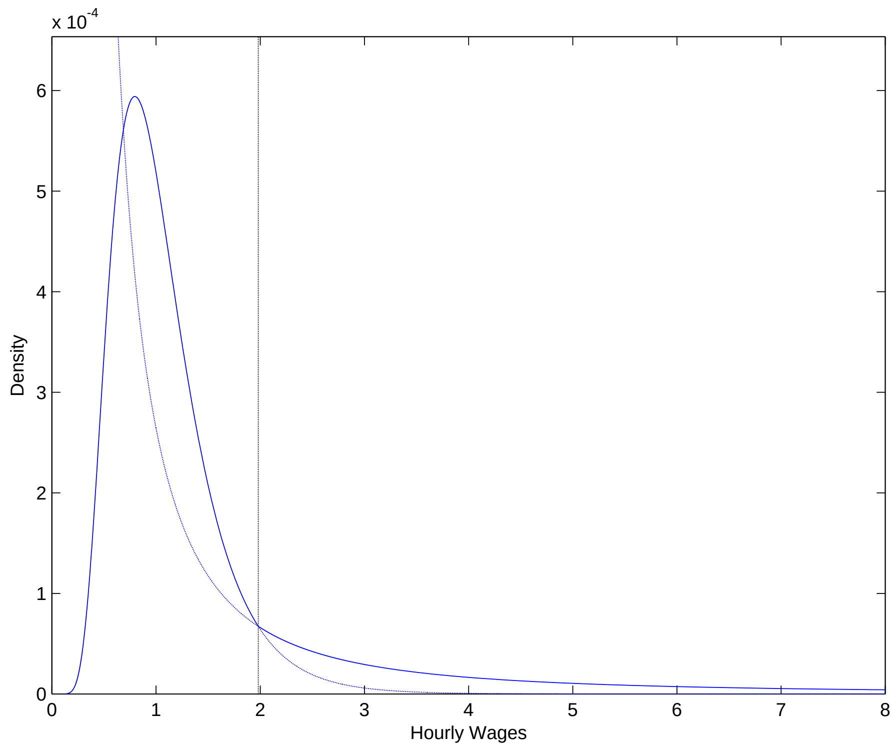
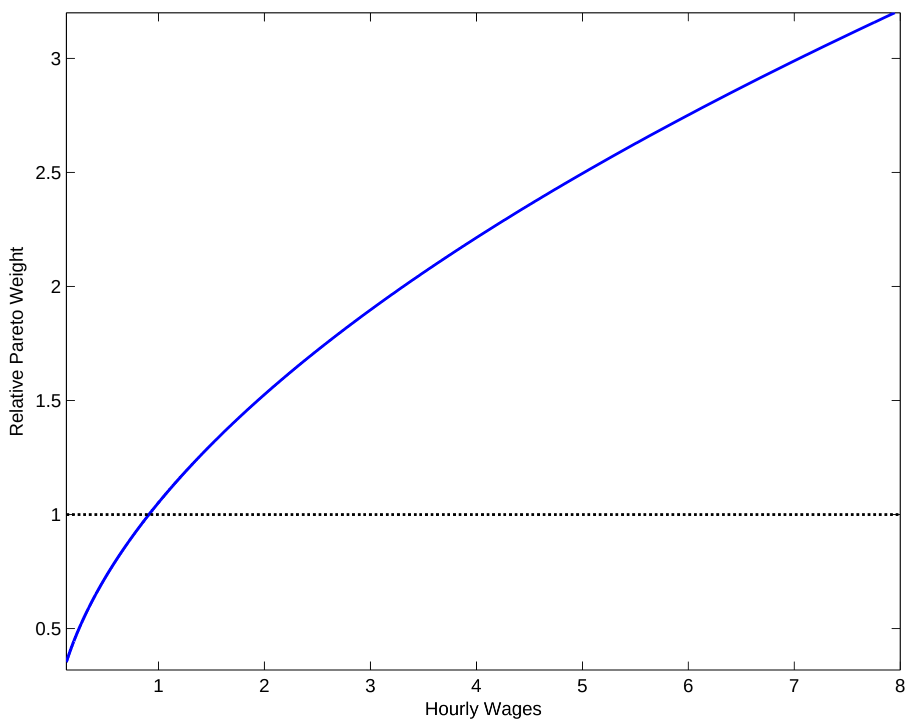


Figure 3: Social Welfare



condition on the observable characteristic κ, as well as tax schedules that do not.

7.1 Impact of Insurable Risks and Social Welfare Function

We consider the impact of insurable shocks and welfare function. In all the exercises, we keep the total variance of the wage shocks constant.

We start with the "standard" Mirrlees model in which Utilitarian planner provides all the insurance in the economy. In this case, all the wage variance comes from the variation in uninsurable and unobservable shocks, i.e., $\log w = \alpha$.

The top panel of Table 1 compares optimal tax systems under this economy. There are large welfare gains from moving to the second best allocation (5.74% of the consumption) and significant output losses (more than 10% of GDP). The marginal tax rates are very high and more than 50% on average. The gains from the second best allocation are approximated well with the affine tax system that displays sizable lump-sum transfers, which appears in the column TR/Y, whereas the HSV tax system is worse than the affine tax. The intuition of this well-known result in the literature (e.g., Mirrlees (1971)) is the following. In this economy, the variance of uninsurable shocks is large and the planner provides all the insurance. Because she has a strong preference for redistribution, she needs a large amount of transfers from productive agents to unproductive ones. The HSV tax has no lump-sum feature by its nature and cannot come close to the second best.

Next, we introduce insurable shocks in the "standard" Mirrlees model. The planner still has a taste for redistribution, but the variance of uninsurable shocks is smaller because agents have access to private insurance markets. The marginal tax rates are then reduced to 43% and the size of the welfare

not change the results quantitatively.

gains is much smaller. This is intuitive given that the role of the planner is relatively small. Perhaps surprisingly, the HSV tax dominates the affine tax unlike the previous case. The reason is because there is a relatively small number of people who receive adverse uninsurable shocks and hence the redistribution effects from the lump-sum transfer is much weaker. The planner wants to redistribute resources to poor agents, but because the target population is small, she does so by having increasing marginal tax rates instead.

The same result appears when we switch to the empirical motivated Social Welfare Function. Note that now the Ramsey HSV tax is our benchmark and the current U.S. tax code is the optimal HSV tax system by construction. The variance of underlying shocks is as large as the "standard" Mirrlees model, but the planner has less incentive to redistribute resources to unproductive agents. The last column shows that the transfers are now very small and hence the marginal tax rates decline to even around 30%. As in the previous case, switching from the HSV tax to the optimal affine tax exhibits welfare losses.

The last panel of Table 1 is the economy with insurable shocks and empirical motivated SWF, which is our benchmark. It shows that the marginal tax is further reduced and is around 30%. The second best allocation displays both welfare gains and output gains. This is because the planner raises the tax revenue more efficiently in this economy.

7.2 Richer Tax Functions

In this section, we consider richer tax schemes in our benchmark economy. Specifically, we compare the baseline HSV tax to polynomial tax systems.

As we saw in the previous section, Table 2 shows that the affine tax system is worse than the baseline HSV tax. However, the quadratic term delivers gains relative to the affine tax and the overall welfare is comparable with the baseline HSV tax. The size of the transfers is smaller than the affine tax, but

Table 1: Impact of Insurable Risks and Social Welfare Function

Wages	SWF	Tax Parameters		Outcomes			
				welfare	Y	mar. tax	TR/Y
$\log w = \alpha$	Utilitarian						
	Affine Tax	$\tau_0 : -0.280$	$\tau_1 : 0.541$	5.04	-10.57	0.541	0.293
	HSV Tax	$\lambda : 0.805$	$\tau : 0.377$	4.32	-9.79	0.507	0.110
	Mirrlees Tax			5.74	-10.67	0.542	0.248
$\log w = \alpha + \varepsilon$	Utilitarian						
	Affine Tax	$\tau_0 : -0.228$	$\tau_1 : 0.445$	0.47	-6.11	0.445	0.210
	HSV Tax	$\lambda : 0.821$	$\tau : 0.289$	1.25	-5.75	0.431	0.066
	Mirrlees Tax			1.60	-5.86	0.434	0.167
$\log w = \alpha$	Empirical						
	Affine Tax	$\tau_0 : -0.120$	$\tau_1 : 0.315$	-0.44	-0.04	0.315	0.096
	HSV Tax (baseline)	$\lambda : 0.842$	$\tau : 0.151$	–	–	0.311	0.020
	Mirrlees Tax			0.09	-0.04	0.312	0.047
$\log w = \alpha + \varepsilon$	Empirical						
	Affine Tax	$\tau_0 : -0.113$	$\tau_1 : 0.300$	-0.60	0.57	0.300	0.083
	HSV Tax (baseline)	$\lambda : 0.835$	$\tau : 0.151$	–	–	0.311	0.018
	Mirrlees Tax			0.20	0.91	0.285	0.033

Table 2: No Type-Contingent Taxes

Tax System				Outcomes				
HSV		λ	τ	welfare	Y	mar. tax	G/Y	TR/Y
		0.835	0.151	-	-	0.311	0.188	0.018
τ_0	τ_1	τ_2	τ_3					
-	0.178	-	-	-1.32	5.61	0.178	0.178	-0.023
-0.113	0.300	-	-	-0.60	0.57	0.300	0.187	0.083
-0.069	0.213	0.022	-	0.01	0.72	0.292	0.187	0.045
-0.032	0.126	0.063	-0.003	0.14	0.80	0.288	0.187	0.015
Second Best (Mirrlees)				0.20	0.91	0.285	0.186	0.033
First Best				9.14	16.03	0	0.162	0.221

the marginal tax rates are also lower on average. This means the higher order term is important to raise the revenue efficiently and it effectively reduces the tax distortion.

The third order term adds more gains and delivers almost 75% of the gains of the second best allocation. This tells us that the increasing marginal tax rates are more important than the universal lump-sum transfers.

One interesting observation of this benchmark case is that additional gains from switching to Mirrlees tax are tiny (0.2% of consumption). Under the empirically motivated social welfare function, the current tax system performs very well and hence there are not much room for welfare improving tax reforms. In the next section, we see that this is not the case once we introduce the observable types and consider type contingent tax systems.

We plot decision rules and the optimal tax schedule for each tax system in Figure 4 to 6. For each figure, we take log hourly wages in the x-axis.

Figure 4 compares the baseline HSV tax with the Mirrlees second best allocation and taxes. The optimal marginal tax, the dashed line in the bottom left figure, displays the usual zero tax result at the bottom and top of the distribution. However, it also shows high non linearity in the middle.

The tax schedule starts at a very low rate around 20% in the area where most people are located, and then sharply increases in the right half of the wage distribution, especially the range where the uninsurable shock follows the Pareto distribution. The planner has a strong incentive to keep the rates low where most people are affected, but wants to increase it to raise enough revenue once the number of people affected decreases. The baseline HSV Ramsey tax scheme captures this increase, while the optimal tax exhibits higher non-linearity.

Figure 5 compares the affine tax with the Mirrlees tax. This reveals the reason why the Ramsey affine tax is worse than the baseline HSV tax. The marginal tax rate is flat in this case, because it is simply given by τ_1. Thus it hardly captures the increasing pattern of the Mirrlees tax. This creates a large gap in the average tax, which is found in the bottom right figure.

Figure 6 compares the cubic tax with the Mirrlees tax. The marginal tax of the Ramsey policy in this case captures not only the increasing pattern but the convex shape of the optimal tax rates. Thus it approximates the optimal average tax and hence delivers most of the welfare gains.

7.3 Type-Contingent Taxes

In this section, we consider $\kappa-$type contingent taxation, the full model considered in the model section.

Table 3 shows the results. We find significant welfare gains relative to non-contingent taxes. For example, the Mirrlees second best tax brings the gains of 1.74% of consumption. By conditioning on observables, the planner can raise the tax revenue more efficiently and the marginal tax rates decreases to the level less than 25%. In contrast to the literature, the optimal tax system also creates significant output gains.

The result that the HSV tax performs better than the affine tax survives in this class of taxes. Even though the affine tax schedule effectively delivers

Figure 4: Baseline HSV Tax System

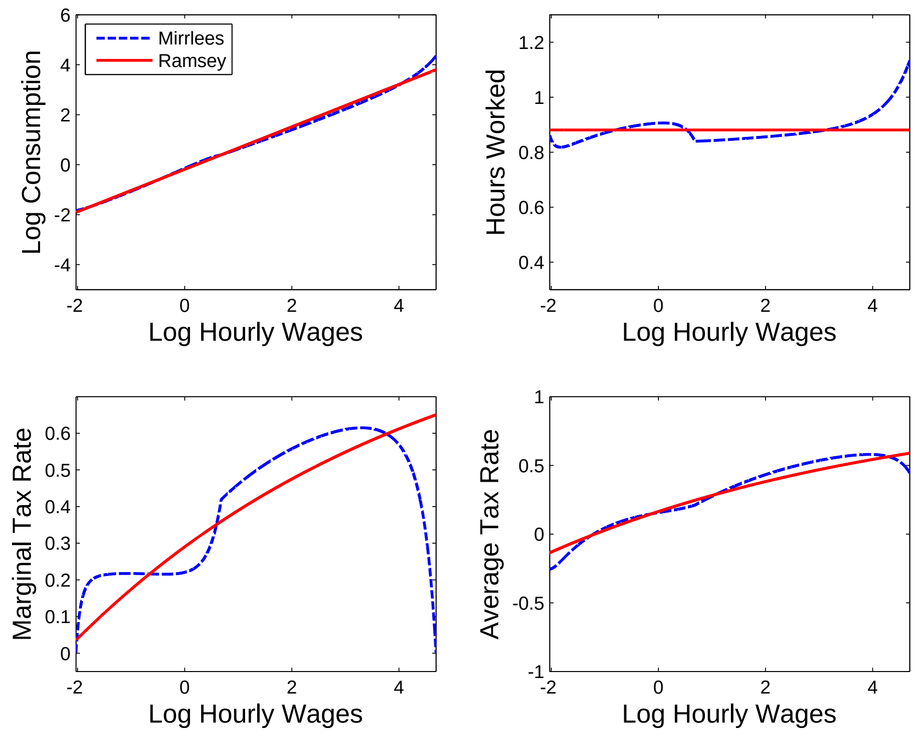


Figure 5: Affine Tax System

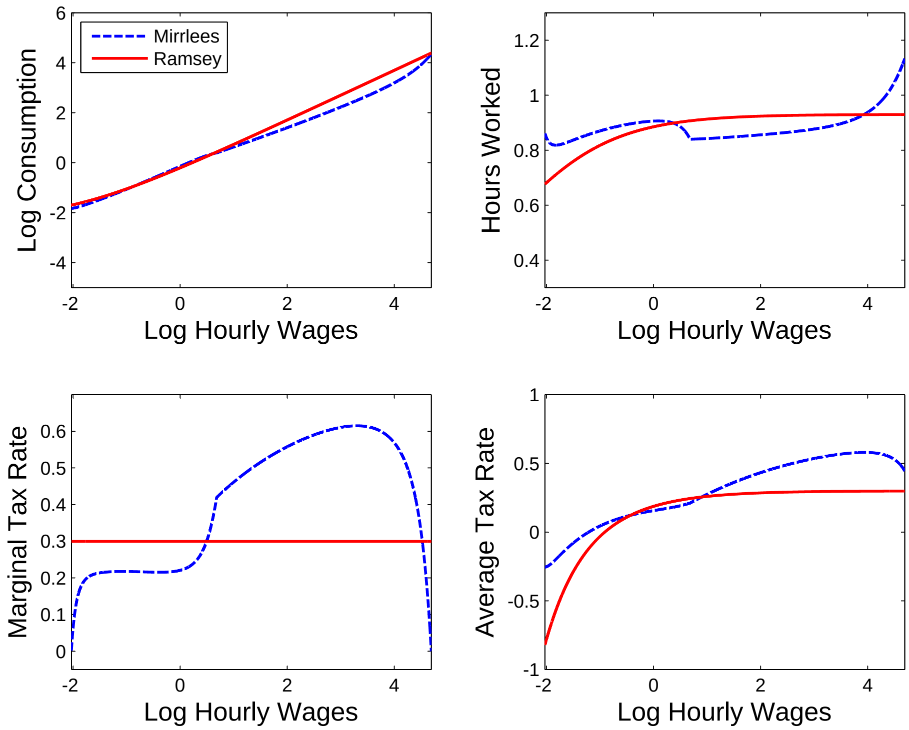
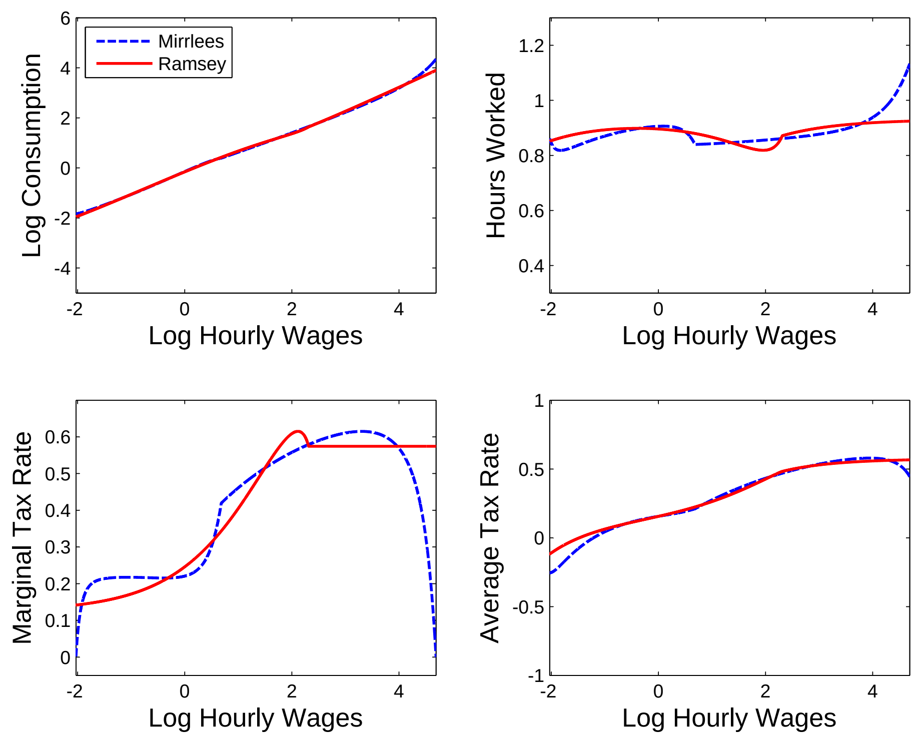


Figure 6: Cubic Tax System



the transfers to low κ-type and sets the marginal tax rates differently for each κ type, it delivers less than 50% of the welfare gains of the second best. Thus the linear tax is again not a good approximation of the optimal tax. On the other hand, the type-specific HSV schedule delivers much larger welfare gains. This fact tells us that the lump-sum transfers are not so important to mimic the optimal allocation and, rather, the increasing marginal tax rates are crucial.

However, this does not mean that we need a complex highly non-linear tax system. It turns out that the second best optimal allocation can be approximately implemented with the type-specific cubic tax and this simple tax scheme delivers 94% of the welfare gain of the second best.

8 Conclusions

There has been a long debate on the structure of labor income taxation in and out of the academia. This paper addresses its main question: how to balance redistribution versus distortions to labor supply. When doing so, we emphase that it is important to measure the gap in terms of allocations and welfare, not in terms of marginal tax rates. The first finding is that Ramsey and Mirrlees tax schemes not far apart in terms of welfare: we can approximately decentralize the second best constrained efficient allocation with a simple tax scheme. We find that increasing marginal tax systems are crutial, but universal lump-sum transfers might not be so important. In the same vein, we find that the current tax scheme might actually be close to Mirrlees. This means that there are not much room for welfare improving tax reforms. However, there will be potential large welfare gains from type-contingent taxes, and to acheive them, we need to condition both transfers and tax rates on observables.

Table 3: Type-Contingent Taxes

HSV		λ	τ	welfare	Y	mar. tax	TR/Y
baseline		0.833	0.151	-	-	0.311	0.018
		0.978 0.715	0.236 0.042	1.06	1.53	0.285	0.063 -0.041
τ_0	τ_1	τ_2	τ_3				
-0.048	0.157	0.054	-0.002	0.12	0.49	0.297	0.035
-	-0.009 0.275	-	-	0.06	5.61	0.178	0.001 -0.048
-0.211 0.075	0.248	-	-	0.75	3.29	0.248	0.193 -0.127
-0.079	0.111 0.338	-	-	0.33	2.26	0.262	0.070 0.028
-0.188 0.052	0.203 0.269	-	-	0.79	3.34	0.248	0.172 -0.105
-0.129 0.234	0.023 0.003	0.097 0.081	-0.005 -0.003	1.64	3.86	0.226	0.123 -0.235
Second Best (Mirrlees)				1.74	4.01	0.221	0.128 -0.144

References

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

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. VIOLANTE (2013): "Redistributive Taxation in a Partial-Insurance Economy," mimeo.

——— (forthcoming): "Consumption and Labor Supply with Partial Insurance: An Analytical Framework," *American Economic Review*.

MIRRLEES, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(114), 175–208.

PIKETTY, T., AND E. SAEZ (2003): "Income Inequality In The United States, 1913-1998," *The Quarterly Journal of Economics*, 118(1), 1–39.

SAEZ, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68(1), 205–29.

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