Practical Optimal Income Taxation*

Jonathan Heathcote†
Federal Reserve Bank of Minneapolis
and CEPR

Hitoshi Tsujiyama‡
University of Surrey

January 2025

Abstract

We review approaches to formulating and solving optimal tax problems in heterogeneous-agent economies. We first show that whether worker heterogeneity is represented through a small or a large number of different productivity types controls the tightness of incentive constraints facing the Mirrlessian planner and therefore has an important impact on policy prescriptions. A popular computational approach that iterates on the Diamond-Saez implicit optimal tax formula does not deliver the constrained efficient allocation when a coarse productivity grid is used. For the purpose of providing quantitative policy recommendations, one safe approach is to solve for the Mirrleesian optimum assuming a very fine grid of productivity types. Alternatively, one can formulate the problem assuming that the distribution of types is continuous, and search for a numerical solution to the system of ordinary differential equations that then define the optimal policy. If these options are infeasible, then optimizing within a flexible parametric class for taxes is preferable to a coarse grid Mirrleesian approach.

Keywords: Optimal income taxation; Mirrlees taxation; Diamond Saez formula; Ramsey taxation

*We thank Job Boerma, Antoine Ferey, Jonas Loebbing, Paweł Doligalski, Dominik Sachs, Nicolas Werquin and Yuki Yao for helpful comments. Tsujiyama acknowledges financial support from Zengin Foundation for Studies on Economics and Finance. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

†Email: heathcote@minneapolisfed.org

‡Email: h.tsujiyama@surrey.ac.uk

1 Introduction

What is the optimal income tax and transfer system? A large literature in the tradition of Mirrlees (1971) has tackled this question. In the classic Mirrleesian setup, individuals differ by labor productivity, leading to income and consumption inequality and a motive for redistributive taxation. Labor productivity is assumed to be private information, so taxes can be conditioned only on labor earnings. Most of the papers on Mirrleesian optimal taxation have a theoretical focus and devote relatively little space to discussing how to produce an accurate numerical solution to the Mirrlees problem.[1] However, such a solution is essential if the goal is to deliver practical policy recommendations.

We show that one aspect of computation that is very important for accurately characterizing optimal policy is the representation for individual productivity. In reality, individual productivity can take a continuum of possible values. However, to compute optimal policy researchers have typically assumed that individual productivity must lie on a discrete grid of possible values. We show that in the class of models with incentive constraints, discretization is a key part of the model environment that controls the strength of information frictions. Policy prescriptions based on analyses with a coarse grid are of little practical value, at best. At worst, depending on the details of how those prescriptions are derived, they are wrong.

We first review the “Mirrlees approach” (sometimes called the “mechanism design approach”) to solving for optimal taxes, which follows Mirrlees (1971). Mirrlees’ insight was that a fruitful way to tackle the tax design problem is to conceptualize the planner as choosing type-specific allocations directly, subject to a resource constraint and incentive constraints that ensure that each productivity type weakly prefers the allocation intended for its type. Given the solution to this problem, in a second step one can search for a tax schedule that will decentralize the constrained efficient allocation as part of a competitive equilibrium. Note that the details of how productivity is discretized are crucial in this problem, because the number of grid points determines the number of incentive constraints the planner faces and thus how much bite the private information friction has.

We show that the Mirrlees approach works in that it delivers the correct constrained efficient allocation. However, for a conventional model calibration, the solution to the Mirrlees problem varies quite dramatically with the coarseness of the grid on productivity. The Mirrleesian planner achieves notably higher welfare when the grid is coarse. There is a straightforward economic intuition for these results. In particular, as is well-known, at

[1] Seminal papers include Mirrlees (1971), Diamond (1998), and Saez (2001). More recent work in dynamic environments builds on Farhi and Werning (2013) and Golosov et al. (2016).

the Mirrleesian optimum only local incentive constraints are typically binding. Suppose one solves for the constrained efficient allocation given a fine grid, then removes every other grid point. All the previously binding local incentive constraints have been eliminated! Thus, the previous allocation, while still feasible, is no longer optimal; the planner can reduce relative consumption of more productive types, or increase their relative earnings.

We find that when the productivity grid is made coarser, the marginal tax rates that decentralize the Mirrleesian optimum are generally lower. We also show that this numerical finding has a theoretical counterpart. In particular, we derive a discrete-grid version of the well-known Diamond-Saez formula, in which the coarseness of the grid appears as a new factor determining (implicit) optimal marginal rates, with greater coarseness translating into lower marginal rates.

However, we also find that a coarser grid implies *average* tax rates that rise more swiftly with income. How can average rates rise more rapidly with income, if marginal rates are lower? To reconcile these findings, it is important to remember that the Mirrlees approach only directly pins down marginal tax rates at each of the discrete set of income values that the different productivity types earn in equilibrium. Fully characterizing the optimal tax schedule requires "filling in" marginal rates in between these income values. This filling in must be done in a way that preserves incentive compatibility and budget feasibility.

When we conduct this filling-in exercise with a coarse grid, we find that while marginal tax rates are low at the income values that are chosen at the optimum, they are very high in between those earnings levels. Thus, the tax planner minimizes distortions by keeping marginal rates low at earnings levels that are actually chosen in equilibrium, while raising revenue via high marginal rates in between those earnings levels. However, note that the highly non-linear optimal tax and transfer function that emerges is purely an artifact of assuming an unrealistically coarse support for productivity; with a very fine grid, the optimal tax schedule is smooth.

Next, we turn to the alternative "tax formula approach" to optimal taxation, which solves simultaneously for the constrained efficient allocation and the corresponding tax schedule by iterating on the Diamond-Saez implicit formula for optimal marginal tax rates.[2] When the formula is used for computation, the productivity distribution is again discretized. A common approach (e.g., Mankiw et al. 2009) is to assume that the tax schedule is piecewise linear, with each discrete productivity type facing a type-specific marginal tax rate. Solving for the optimum involves computing equilibrium allocations given a candidate tax schedule,

[2]This formula can be derived from the Mirrlees problem first order conditions. An alternative derivation starts by noting that at an optimum, the marginal social welfare gains from perturbing marginal tax rates at any earnings level must be zero, and it uses an expression for those marginal gains to derive the same condition. Thus, this is sometimes called the "perturbation approach" to optimal taxation.

then plugging them into the Diamond-Saez formula to update the vector of guesses for optimal marginal tax rates.

We find that when the productivity grid is very fine, the method works: we uncover the same allocation and the same smooth optimal tax schedule that we find under the Mirrlees approach. However, the method does not work when the grid is coarse, in that allocations and tax rates converge neither to the coarse grid optimum nor to the fine grid optimum. The allocation the method delivers typically has the property that downward incentive constraints are slack, indicating that the allocation is not constrained efficient. The approach is doomed to fail because any tax schedule that decentralizes the true Mirrleesian solution is far from piecewise linear when the grid on productivity is coarse. An additional problem we identify with the tax formula approach is that the Diamond-Saez tax formula is typically derived assuming a continuous distribution for productivity. Deriving the formula correctly, given a discrete distribution, leads to a different expression involving the coarseness of the grid.

We conclude that to accurately characterize the optimal tax and transfer schedule in a setting with a discrete type distribution, a very fine grid on productivity is required, especially if the tax formula approach is being used. A useful check that we suspect most existing papers would fail is to recompute optimal policy with a finer grid and verify that this does not lead to noticeably different policy prescriptions. Future research should be explicit about numerical solution methods used and especially about the number of grid points.[3]

While using a very fine grid is the simplest solution, it is not always practical. An alternative approach, which we consider next, is to formulate the Mirrlees problem assuming a continuous type space, in which case the optimal policy can be represented as the solution to a system of ordinary differential equations. Since this "differential equations approach" embeds a continuous set of incentive constraints, it is free from the fundamental problem of coarse discretization. Solving the differential equations numerically does require choosing a step size for approximation, but this step size is a detail of the numerical approximation method, rather than a fundamental feature of the original physical environment. We show that the differential equations approach gives a very accurate solution as long as the step size is not too large. We also show that the differential equations to be solved can be described in a way that builds naturally and intuitively on the Diamond-Saez optimal tax formula.

In sum, to accurately characterize optimal taxes, it is necessary to either work with a very fine productivity grid or to use the differential equations approach. Suppose, however, that neither of these options is feasible, as they might not be in richer environments than the

[3]Two widely cited papers that are laudably explicit about computational details are Brewer et al. (2010) and Mankiw et al. (2009). Other papers that report the number of grid points used in computation include Chang and Park (2020) and Boar and Midrigan (2021) (who use a very fine grid).

simple static model we focus on. In such contexts, we propose that a sensible way to proceed is to solve for the tax schedule that is optimal within a simple but flexible parametric class. That is, one should search for taxes that are optimal in the Ramsey tradition—given a parametric functional form—rather than optimal in the non-parametric Mirrleesian sense.[4] We propose a particular parametric form, according to which marginal tax rates are a polylogarithmic function of income. We show that in our calibrated model, a cubic function in this class delivers allocations and welfare very nearly identical to those in the Mirrleesian optimum. And, just as importantly, the optimal policy and associated allocations are largely insensitive to the number of grid points assumed for productivity. Thus, the Ramsey policy that is optimal when the policy problem is solved on a coarse grid remains close to optimal given a nearly continuous distribution, while the Mirrlees policy is not similarly robust.

2 The Mirrlees Model

This section describes the standard static Mirrlees model, in which individuals differ only by productivity, and in which labor supply is the only margin distorted by taxation.

2.1 Environment

Productivity. There is a unit mass of individuals. They differ with respect to labor productivity $\theta \in \Theta \subset \mathbb{R}_+$, where θ is continuously distributed with CDF $F(\theta)$.

Preferences. Agents have identical preferences with the separable utility function

$$u(c, y; \theta) = \log(c) - \frac{1}{1+\sigma}\left(\frac{y}{\theta}\right)^{1+\sigma}, \tag{1}$$

where c is consumption, y is earnings, and $\sigma > 0$. We use $c(\theta)$ and $y(\theta)$ to denote consumption and earnings for an individual of type θ. The planner observes y, but not θ.

Technology. Aggregate output is equal to aggregate earnings and is divided between private consumption and a non-valued publicly provided good G. The resource constraint of the economy is thus

$$\int c(\theta) dF(\theta) + G = \int y(\theta) dF(\theta). \tag{2}$$

2.2 Mirrlees Problem

The Mirrlees problem determines constrained efficient allocations. We envision each individual drawing $\theta \in \Theta$ and making a report $\tilde{\theta} \in \Theta$ to the planner. The Mirrlees planner

[4] Papers that integrate Mirrleesian and Ramsey approaches to optimal taxation include Huggett and Parra (2010), Blundell and Shephard (2012), and Heathcote and Tsujiyama (2021).

maximizes social welfare by choosing both consumption $c(\tilde{\theta})$ and income $y(\tilde{\theta})$ as functions of reported type $\tilde{\theta}$. Given true type θ and reported type $\tilde{\theta}$, individual welfare is given by

$$U(\theta,\tilde{\theta}) = \log\left(c(\tilde{\theta})\right) - \frac{1}{1+\sigma}\left(\frac{y(\tilde{\theta})}{\theta}\right)^{1+\sigma}. \tag{3}$$

Planner's Problem. We focus on a utilitarian social welfare objective. The planner thus maximizes average welfare subject to the resource constraint and to incentive constraints that ensure truth-telling:

$$\max_{\{c(\theta),y(\theta)\}_{\theta\in\Theta}} \quad \int U(\theta,\theta)dF(\theta), \tag{4}$$

$$\text{subject to} \quad \int c(\theta)dF(\theta) + G = \int y(\theta)dF(\theta), \tag{5}$$

$$U(\theta,\theta) \geq U(\theta,\tilde{\theta}) \qquad \text{for all } \theta \text{ and } \tilde{\theta}. \tag{6}$$

2.3 Decentralization

The allocation that solves the Mirrlees problem can be decentralized as a competitive equilibrium by an appropriate choice of a tax schedule $T(y)$ that defines how rapidly consumption grows with income.

In a competitive equilibrium, individuals maximize utility (1) subject to the budget constraint

$$c(\theta) = y(\theta) - T(y(\theta)). \tag{7}$$

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

(i) the decision rules $\{c,y\}$ solve the individual maximization problem;

(ii) the resource feasibility constraint (2) is satisfied; and

(iii) the government budget constraint is satisfied: $\int T(y(\theta))\,dF(\theta) = G$.

Taxation. Let $c^*(\theta)$ and $y^*(\theta)$ denote the values for consumption and income that solve the Mirrlees problem (4) for workers with productivity θ.

The tax schedule that decentralizes this allocation must satisfy

$$T(y^*(\theta)) = y^*(\theta) - c^*(\theta).$$

With the caveat that the tax function that decentralizes the Mirrlees solution might not be differentiable, we will define the marginal tax rate at the Mirrlees optimum as the rate at which the household is indifferent about working marginally more. Given the assumed utility function, this marginal tax rate satisfies

$$T'\left(y^*(\theta)\right) = 1 - \frac{c^*(\theta)}{\theta}\left(\frac{y^*(\theta)}{\theta}\right)^{\sigma}. \tag{8}$$

2.4 Diamond-Saez Formulae

A standard way to interpret optimal tax rates is through the famous Diamond-Saez formula (hereafter, DS formula). Given our preference specification (1), this formula is given by

$$\frac{T'\left(y^*(\theta)\right)}{1 - T'\left(y^*(\theta)\right)} = \underbrace{(1+\sigma)}_{A} \times \underbrace{\frac{1}{\theta f(\theta)} \int_{\theta}^{\infty} \left(1 - \frac{\mathbb{E}[c^*(\theta)]}{c^*(s)}\right) \frac{c^*(s)}{c^*(\theta)} dF(s)}_{B}, \tag{9}$$

where the labels A and B on the right hand side correspond to the decomposition in Saez (2001).[5]

Discrete-State Version of the Diamond-Saez Formula. In papers that seek to solve for the Mirrlees optimum following the tax formula approach, equation (9) is typically evaluated on a discrete grid, where the integral in term B is replaced by a summation over the discretized productivity space.

However, if the productivity distribution is really discrete, then the correct DS equation takes a different form, which we now derive. In particular, suppose there is a grid of N evenly spaced values for the log of productivity, $\{\log(\theta_1), \cdots, \log(\theta_N)\}$, with corresponding probabilities $\{\pi_1, \cdots, \pi_N\}$. The discrete-state version of the Mirrlees planner's problem (4) is

$$\max_{\{c_i, y_i\}_{i=1}^N} \quad \sum_i \pi_i U(\theta_i, \theta_i), \tag{10}$$

$$\text{subject to} \quad \sum_i \pi_i c_i + G = \sum_i \pi_i y_i, \tag{11}$$

$$U(\theta_i, \theta_i) \geq U(\theta_i, \theta_j) \qquad \text{for all } i \text{ and } j. \tag{12}$$

Let $\{y_i^*, c_i^*\}_{i=1}^N$ denote the optimal allocation. Suppose values in the grid for productivity are evenly spaced (in logs), and let κ define the coarseness of the grid, where $\kappa = \frac{\theta_{i+1}}{\theta_i} = \left(\frac{\theta_N}{\theta_1}\right)^{\frac{1}{N-1}}$.

[5] See Appendix A.1 for the derivation.

Proposition 1 *The discrete-state version of the DS formula is given by*

$$\frac{T_i^{*\prime}}{1-T_i^{*\prime}} = \left[1-\kappa^{-(1+\sigma)}\right]\frac{1}{\pi_i}\sum_{s=i+1}^{N}\pi_s\left(1-\frac{\mathbb{E}\left[c^*\right]}{c_s^*}\right)\frac{c_s^*}{c_i^*}. \tag{13}$$

Proof. See Appendix A.2. ■

What is novel about formula (13) is that the coarseness of the grid κ appears as a new parameter determining optimal tax rates.

To make eq. (13) comparable to eq. (9), we approximate the discrete probability as $\pi_i \approx f(\theta_i)(\theta_{i+1}-\theta_i) = f(\theta_i)\theta_i(\kappa-1)$ and obtain the following.

Corollary 2 *The discrete-state version of the DS formula (13) can be approximated by*

$$\frac{T_i^{*\prime}}{1-T_i^{*\prime}} \approx \underbrace{\frac{1-\kappa^{-(1+\sigma)}}{\kappa-1}}_{\tilde{A}} \times \underbrace{\frac{1}{\theta_i f(\theta_i)}\sum_{s=i+1}^{N}\pi_s\left(1-\frac{\mathbb{E}\left[c^*\right]}{c_s^*}\right)\frac{c_s^*}{c_i^*}}_{\tilde{B}}. \tag{14}$$

Eq. (14) resembles eq. (9), so we use analogous labels for each component on the right hand side. As $\kappa \to 1$ and the productivity grid becomes arbitrarily fine, term $\tilde{A}$ converges (using l'Hôpital's rule) to $1+\sigma$, and thus eq. (14) converges to eq. (9). Term A in eq. (9) depends only on σ, the labor supply elasticity parameter.[6] In contrast, when the distribution is discrete, term $\tilde{A}$ contains a new driver of optimal marginal tax rates: the coarseness of the grid κ. In particular, the higher is κ (the coarser the grid), the lower are marginal rates, all else equal. Recall that the larger is κ, the fewer incentive constraints the Mirrlees planner faces. Equation (13) indicates that this translates into a less distortionary tax system. We will return to this point in the quantitative analysis in Section 4.

2.5 System of Ordinary Differential Equations

The planner's solution is characterized by the incentive constraint (6) and by a first-order condition. The latter can be restated as the DS formula (9). Both of these conditions must be satisfied for all $\theta \in \Theta$. These conditions can be formulated as a system of ordinary differential equations (ODE), with utility $U(\theta)$ as the state and labor supply $h(\theta) = y(\theta)/\theta$ as the control.

[6] More generally, this term is one plus the uncompensated labor supply elasticity divided by the compensated labor supply elasticity.

The incentive constraint (6) states

$$U(\theta) = \max_{\tilde{\theta}} \left\{ \log c(\tilde{\theta}) - \frac{1}{1+\sigma} \left(\frac{y(\tilde{\theta})}{\theta} \right)^{1+\sigma} \right\}.$$

Using the envelope condition, we have

$$U'(\theta) = \frac{1}{\theta} h(\theta)^{1+\sigma}. \tag{15}$$

Next, we construct a differential equation that summarizes the planner's optimality condition. There are several ways one could do this in our economy. Here we derive a differential equation directly from the DS formula (9), and use the marginal tax ratio on the left-hand side of that formula as the variable whose dynamics we will compute. Thus, the variables appearing in our ODE system have direct economic interpretations, in contrast to the formulations of Mirrlees (1971) and Saez (2001).[7]

In particular, define the optimal marginal tax ratio $Q(\theta) = \frac{T'(\theta)}{1-T'(\theta)}$. From the household first-order condition (8), we can write

$$Q(\theta) = \frac{\theta - c(\theta)h(\theta)^{\sigma}}{c(\theta)h(\theta)^{\sigma}}. \tag{16}$$

Now taking the derivative of the DS formula (9), we obtain a differential equation for the tax ratio:

$$Q'(\theta) = -Q(\theta) \left(\frac{1}{\theta} + \frac{f'(\theta)}{f(\theta)} + \frac{c'(\theta)}{c(\theta)} \right) + \frac{1+\sigma}{\theta} \left(\frac{1}{\zeta c(\theta)} - 1 \right), \tag{17}$$

where the consumption function and its derivative are implicitly given by

$$\begin{aligned} c(\theta) &= \exp\left\{ U(\theta) + \frac{1}{1+\sigma} h(\theta)^{1+\sigma} \right\}, \quad \text{and} \\ c'(\theta) &= c(\theta) \left[U'(\theta) + h(\theta)^{\sigma} h'(\theta) \right]. \end{aligned}$$

Equations (15) and (17) are a system of ODEs, which describe how individual welfare and the marginal tax schedule change along the productivity distribution. By solving them—with appropriate terminal conditions—we can characterize the optimal tax system and associated labor supply.

[7] See Appendix B for more on the derivation.

3 Calibration

The calibration closely follows Heathcote and Tsujiyama (2021).

Preferences. We assume $\sigma = 2$ so that the Frisch elasticity $(1/\sigma)$ is 0.5.

Current Taxes and Spending. We approximate the current tax and transfer system using the function in Benabou (2000) and Heathcote et al. (2017) (henceforth labeled "HSV"), according to which taxes net of transfers are given by $T(y) = y - \lambda y^{1-\tau}$. The parameter τ indexes the progressivity of the system, and Heathcote et al. (2017) estimate $\tau = 0.181$. The parameter λ is set so that government purchases G are 18.8 percent of model GDP. When solving for the Mirrleesian tax policies, we hold G fixed at its baseline value.

Wage Distribution. Log wages are drawn from an exponentially modified Gaussian distribution (EMG): $\log(\theta) \sim EMG(\mu_\theta, \sigma_\theta^2, \lambda_\theta)$. We use the estimates from Heathcote and Tsujiyama (2021) for the distributional parameters λ_θ and σ_θ^2, focusing on their estimates for the privately uninsurable component of wages.[8] Thus, $\lambda_\theta = 2.2$, $\sigma_\theta^2 = 0.142$, and the total variance of log wages is 0.348.[9]

Discretization. We construct a grid of N evenly spaced values for the log wage, $\{\log(\theta_1), \cdots, \log(\theta_N)\}$, with corresponding probabilities $\{\pi_1, \cdots, \pi_N\}$. We set θ_1 so that $\theta_1 / \sum_i \pi_i \theta_i = 0.05$ and set θ_N so that $\theta_N / \sum_i \pi_i \theta_i = 74$, which corresponds to household labor income at the 99.99th percentile of the labor income distribution in the Survey of Consumer Finances (\$6.17 million). The coarseness of the grid is given by $\kappa = \left(\frac{74}{0.05}\right)^{\frac{1}{N-1}}$. We read densities directly from the continuous EMG distribution and rescale them to obtain corresponding probabilities π_i such that $\sum_i \pi_i = 1$.[10]

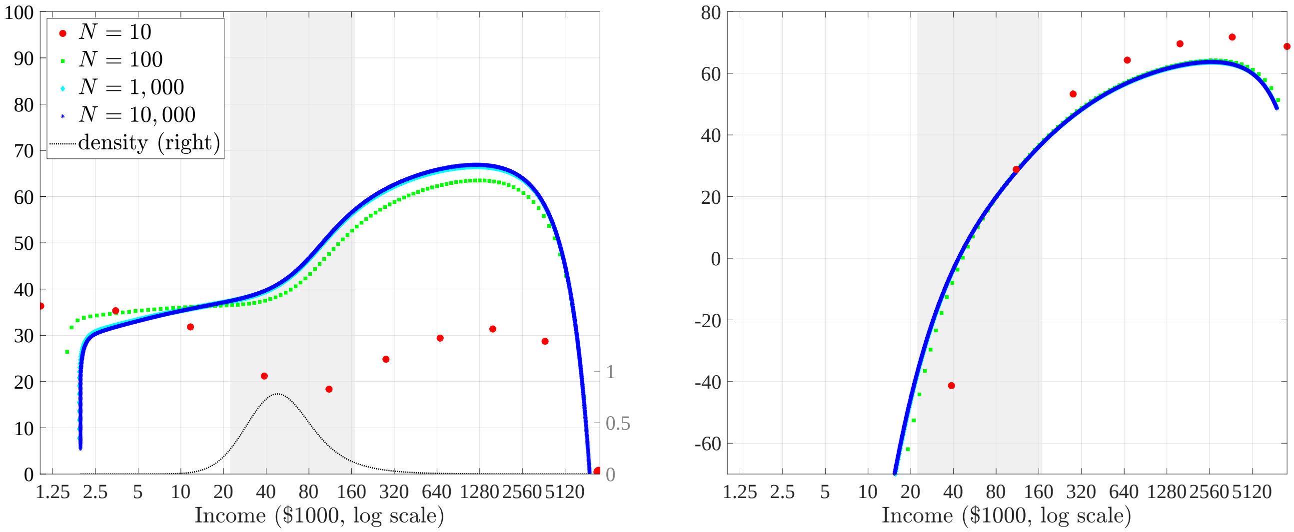
4 Quantitative Analysis

This section describes how the number of grid points N affects the optimal tax and transfer system. We then discuss how to address the problem associated with coarse discretization.

[8] Heathcote and Tsujiyama (2021) show that the assumptions on the productivity distribution deliver extremely close fits to the top of the earnings distribution in the SCF and to the bottom of the latent offered wage distribution estimated in Low and Pistaferri (2015).

[9] Appendix C.1 describes how we scale model units.

[10] We adjust the value for σ_θ^2 in the continuous distribution from which we draw densities to ensure that the variance of the discretely distributed variable $\log(\theta_i)$ remains exactly equal to 0.348 for every different value for N we consider.



(A) Marginal Tax Rate (%)

(B) Average Tax Rate (%)

Figure 1: Optimal Tax Policy. Panels A and B plot the optimal Mirrleesian marginal and average tax schedules with the number of grid points from 10 to 10,000. Panel A also plots the density of income distribution for 10,000 grid points. The area between the 5th and 95th percentiles is shaded gray.

4.1 Mirrlees Approach

The *Mirrlees approach* conceptualizes the planner choosing type-specific allocations directly. Given a solution $\{y_i^*, c_i^*\}_{i=1}^N$ to the discrete-state Mirrlees problem (10), the corresponding optimal average and marginal tax rates at each grid point are given by[11]

$$\begin{aligned} \frac{T_i^*}{y_i^*} &= \frac{y_i^* - c_i^*}{y_i^*}, \\ T_i^{*\prime} &= 1 - \frac{c_i^*}{\theta_i}\left(\frac{y_i^*}{\theta_i}\right)^{\sigma}. \end{aligned} \tag{18}$$

Optimal Tax Policy and Grid Points. Figure 1 plots optimal marginal and average tax rates when N varies from 10 to $10,000$ (κ varies from 2.25 to 1.0007). With a very fine grid—say, $N > 1,000$—these tax schedules become insensitive to further increasing N.[12] We have verified this result by solving a version with $N = 100,000$; the resulting tax schedules are indistinguishable from those with $N = 10,000$. We therefore think of the solution with $N = 10,000$ as an accurate representation of optimal taxation in an economy whose

[11] Appendix C.2 provides the details of the computation.

[12] This result holds only when the wage grid points are evenly spaced in the log space. If they are evenly spaced in *levels*, there will be relatively few grid points in the densely populated range. In this case, even the optimal schedule calculated with $N = 1,000$ is a poor approximation to the true optimal schedule.

distribution of productivity is continuous, as it is in reality. Note that optimal marginal rates are generally increasing in income.[13] Since the wage distribution is bounded, the Mirrleesian marginal tax rate drops to zero at the very top of the income distribution.

With fewer grid points, however, optimal tax rates are quite different, especially in the case with only 10 grid points. Marginal tax rates are higher at the very bottom, but are much lower over the rest of the income distribution. This observation is qualitatively consistent with how κ features in eq. (14).[14] Furthermore, the level differences in plotted marginal tax rates across different grid coarsenesses are quantitatively of similar magnitude to what one might predict based on the corresponding grid-specific values for κ. For example, the average income-weighted value for the marginal tax ratio $T_i'/(1-T_i')$ falls from 1.029 to 0.280 (a ratio of 3.7) as N is reduced from $10,000$ to 10. The corresponding value for term $\tilde{A}$ in eq. (14) falls from 2.996 to 0.730, a ratio of 4.1. The declines are not identical, because changing the coarseness of the grid also changes the constrained efficient consumption allocation and thus term $\tilde{B}$. However, it is clear that a coarser productivity grid implies lower optimal marginal rates and that this result is hard-wired into the Mirrleesian optimality conditions. We will shortly offer some economic intuition for this result.

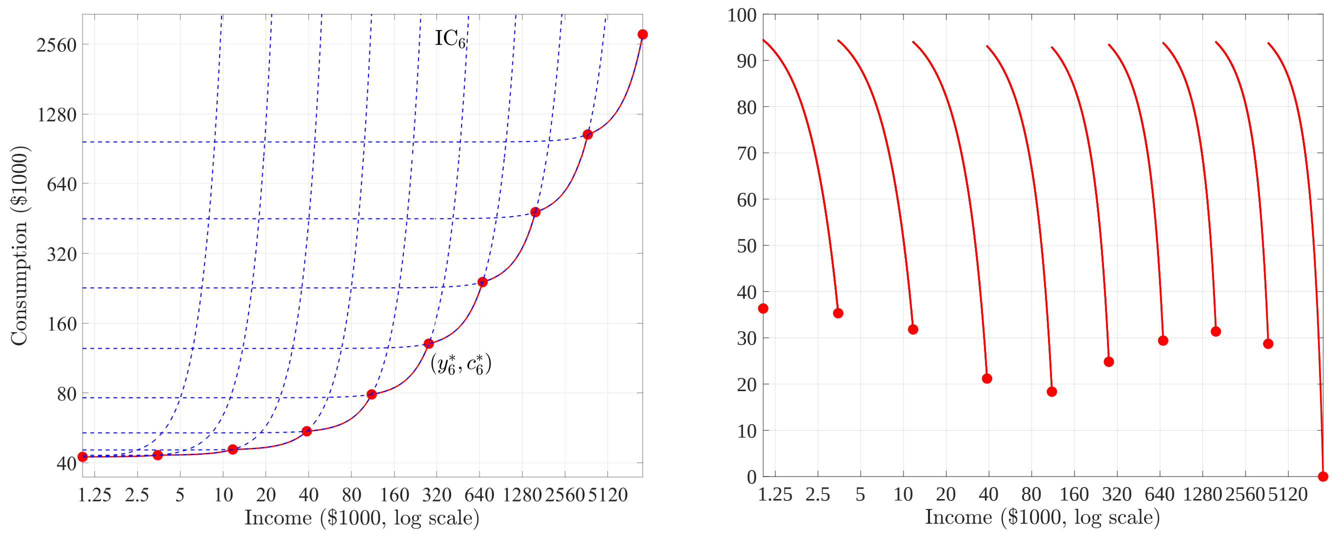
An apparent contradiction emerges when comparing the profiles for marginal tax rates (panel A) with the ones for average tax rates (panel B). With only 10 grid points, marginal rates are mostly lower than under the nearly continuous case, suggesting less redistribution. However, average tax rates increase more rapidly than income, suggesting a more redistributive tax schedule. How can it be that marginal tax rates rise more slowly with income, while average tax rates rise more swiftly? To understand this, it is important to remember that eq. (18) only pins down the optimal tax rates *at each point in the grid on productivity.* We need to fill in how taxes vary in between grid points.

Filling in the Tax Schedule between Grid Points. We have already pinned down taxes and marginal tax rates at the income values that the planner intends for each productivity type; that is, we know $T(y_i^*) = T_i^*$ and $T'(y_i^*) = T_i^{*\prime}$. We now fill in the rest of the tax schedule in a way that ensures taxes are continuous in income and marginal tax rates are everywhere between 0 and 100 percent.

When filling in the tax schedule T in between grid points, we need to ensure that the constrained efficient allocation is a competitive equilibrium given that schedule. More specifically, each type θ_i, taking T as given, must weakly prefer income y_i^* not only to any alternative value in the vector $\{y_i^*\}_{i=1}^N$ but also to any other off-the-grid income value $y \in \mathbb{R}_+$.

[13] Heathcote and Tsujiyama (2021) explore the economic forces behind this increasing pattern of optimal marginal tax rates.

[14] We verify that eq. (13) delivers the same marginal tax rates as those plotted in figure 1A at each grid point and for each grid coarseness.



(A) Income versus Consumption

(B) Marginal Tax Rate (%)

Figure 2: Income-Consumption Menu and Optimal Marginal Taxes with a Coarse Grid. Panel A plots the optimal income-consumption pairs when $N = 10$ (red dot) and the indifference curve for each wage type (blue dashed line). It also plots an optimal income-consumption menu off the grid (red solid line). Panel B plots the corresponding marginal tax schedule that decentralizes the optimal allocation.

To illustrate the set of tax schedules that satisfy this requirement, we plot income against consumption in figure 2A for the case $N = 10$. The blue dashed lines indicate the indifference curves for each type that go through the optimal income-consumption pair for that type, (y_i^*, c_i^*), which are marked by red dots. Thus, $IC_i(y)$ defines the level of consumption c at income level y such that $u\left(c, \frac{y}{\theta_i}\right) = u\left(c_i^*, \frac{y_i^*}{\theta_i}\right)$. Notice that the indifference curve labeled $IC_6(y)$ for type θ_6, passes through not only (y_6^*, c_6^*) but also (y_5^*, c_5^*), which is a graphical representation of the fact that local downward incentive constraints bind at the optimum.

Any continuous tax schedule that decentralizes the constrained efficient allocation must connect the dots. Given a schedule that connects the dots, each type θ_i will choose the corresponding efficient income level y_i^* if and only if $y - T(y)$ lies weakly below $IC_i(y)$. For a tax schedule to decentralize the efficient allocation, this condition must be satisfied for all types; that is, $y - T(y) \leq IC_i(y)$ for all i.

There are many tax schedules that satisfy this requirement. Here, we propose an optimal tax schedule among the set of feasible continuous schedules that is the least extreme in the sense that discontinuous jumps in the marginal rate are the smallest.[15] This tax schedule

[15] Kocherlakota (2010) considers a schedule that features 100 percent marginal tax rates in between grid points. In this case, the income-consumption profile would be a step function, and the average tax rate would drop discontinuously at each grid point.

has the property that the implied consumption profile $c(y) = y - T(y)$ traces the relevant indifference curve in between grid points, which is the red line in figure 2A. In particular, in between the income levels y_i^* and y_{i+1}^*, the residual income function traces the indifference curve for type θ_{i+1}. Thus, type θ_{i+1} is indifferent between delivering y_{i+1}^* versus any income level in the interval $[y_i^*, y_{i+1}^*)$. The associated optimal marginal tax schedule against income is depicted in figure 2B. It shows that the schedule is highly nonlinear and jumps to more than 90 percent each time it crosses an income threshold y_i^* before declining as income rises to y_{i+1}^*.[16] Note that *any* tax schedule that decentralizes the optimum must feature high marginal tax rates in between grid points; these high marginal rates in between grid points are required to reconcile low marginal rates but high average tax rates at the income levels in the vector $\{y_i^*\}$; see figure 1B.[17] As the grid is made finer, the jumps in the optimal marginal tax schedule obtained with this procedure become smaller. Homburg (2001) decentralizes the Mirrlees allocation using a similar tax schedule, and argues that with a fine enough grid the marginal tax schedule becomes continuous.

In the existing literature, the optimal tax schedule is often reported by linearly interpolating between the tax rates at each grid point (graphing software does this automatically!). The optimal marginal schedule with $N = 10$ in figure 2B would appear U-shaped if the rates at each grid point were connected linearly. However, this seemingly innocuous visualization convention is misleading; the true optimal tax schedule in between grid points is highly nonlinear.

Welfare. Table 1 reports how welfare varies with the coarseness of the productivity grid under various alternative tax systems. In each case, welfare gains are relative to the baseline economy with the HSV tax function and $N = 10,000$. The key takeaway is that welfare in the equilibrium with HSV taxation (column 1) or at the first best (column 2) varies little with N. In contrast, in the Mirrlees economy (column 3), welfare gains increase substantially as N is reduced. For $N = 10,000$ these gains are 2.1 percent of consumption, compared with 20.1 percent with $N = 10$.

The intuition for this result is straightforward. At the Mirrleesian optimum, the local downward incentive constraints are binding. Suppose one solves for the constrained efficient allocation for a given grid on productivity. Now remove every other productivity value in the grid. At the original conjectured solution, none of the incentive constraints are now binding, because only the local ones were previously binding, and those constraints have been deleted. Thus, the planner can either reduce relative consumption of more productive

[16]Note that if the jump in the marginal tax rate at the threshold y_i^* were any smaller, then type θ_{i+1} would be be better off reducing earnings from y_{i+1}^* to a level just above y_i^*.

[17]For $N \geq 1,000$, the optimal marginal tax schedule obtained with this procedure is virtually identical to the schedule reported in figure 1A.

Table 1: Welfare Gains

	Welfare Gains (%,CEV)				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
# of grid points N	HSV	First Best	Mirrlees	Tax Formula	Ramsey
$10,000$	—	44.72	2.07	2.05	2.01
$1,000$	0.00	44.73	2.28	2.06	2.01
100	-0.01	44.81	4.40	2.13	2.01
50	-0.01	44.89	6.66	2.08	2.00
10	-0.21	46.07	20.13	-2.88	1.87

Note: Welfare gains are calculated relative to the economy with the HSV tax function and $N = 10,000$ grid points. For each alternative economy X, we compute the percentage amount by which all individuals consumption must be increased in the baseline economy for an individual to be indifferent between being dropped at random into economy X versus the baseline economy. The number of grid points varies from $N = 10,000$ to $N = 10$. The columns show results for (1) the equilibrium allocation with HSV taxation, (2) the no-private-information first best (decentralizable with type-specific lump-sum taxes), (3) the constrained efficient Mirrleesian allocation computed by the Mirrlees approach, (4) the constrained efficient Mirrleesian allocation computed by the tax formula approach, and (5) the equilibrium allocation with flexible Ramsey taxation.

types or increase their relative earnings, leading to a welfare superior allocation that is closer to the first best. In terms of the tax decentralization, the counterpart of incentive constraints being relaxed as the grid becomes more sparse is that the planner can increase redistribution while reducing distortions to labor supply by setting high marginal tax rates in between equilibrium income levels and low marginal rates at those equilibrium income levels, the pattern we found in figure 2B.

Our conclusion from this exploration is that Mirrleesian policy prescriptions based on analyses with a coarse grid are of little practical value if the underlying productivity distribution is, in reality, continuous. The resulting tax schedules will be far from the true optimum, and the implied welfare gains from tax reform will be vastly overstated. In practice, it might not be obvious ex ante how fine the grid needs to be to ensure robust optimal policy recommendations. One simple check that we recommend is to verify that the numerical solution is not materially changed when the number of grid points is increased.

4.2 Tax Formula Approach

The *tax formula approach* uses the DS formula to jointly solve for optimal taxes and the corresponding allocations (e.g., Mankiw et al. 2009; Brewer et al. 2010).[18] Versions of the formula have been derived in many extensions of the original static Mirrlees economy, and

[18] Appendix C.3 provides the details of the computation.

in these richer economies, it is sometimes considered easier to exploit the formula directly for computation rather than to adopt the Mirrleesian approach.

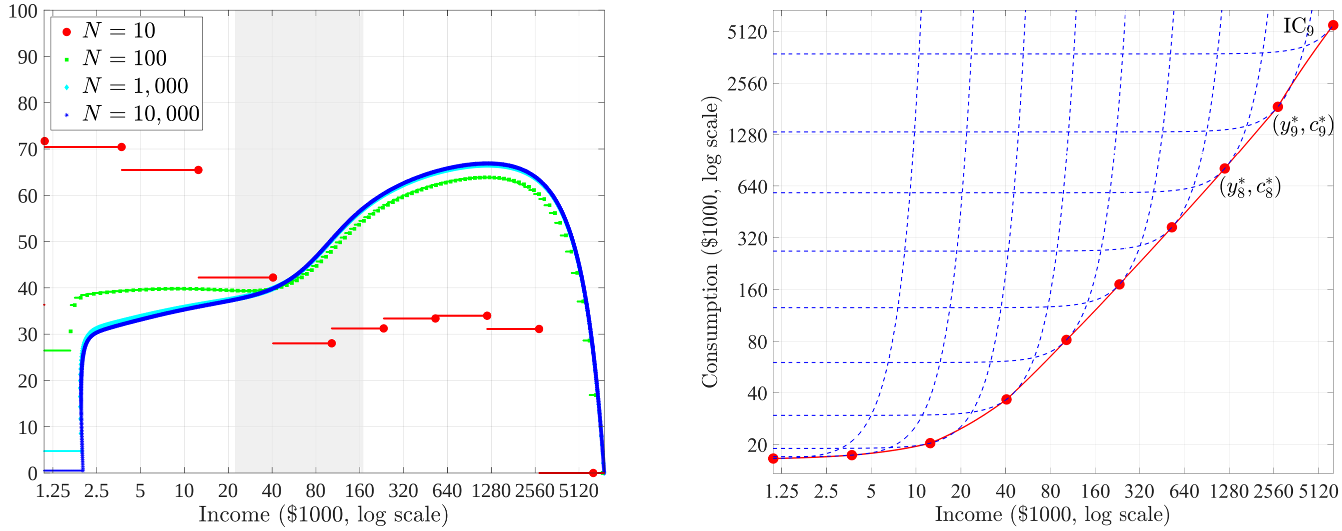
The formula that applies in our economy is eq. (13). Since it contains endogenous variables, it does not immediately pin down exact optimal tax rates. Rather, one must solve a fixed point problem to find T. The typical computational approach in the literature has been as follows. First, approximate T with a tax function $\tilde{T}$. Second, given $\tilde{T}$, compute the competitive equilibrium allocation for the discretized economy, $\{y_i, c_i\}_{i=1}^N$. Third, given this allocation, use the DS formula to compute a new guess for the optimal tax function, $\tilde{T}$. Repeat until the system converges to a tax function at which the corresponding equilibrium satisfies eq. (13).

Note that computing the equilibrium in the second step requires specifying a tax function $\tilde{T}$ for all possible values for income. A typical assumption is to assume that $\tilde{T}$ is a piecewise linear function, with as many segments as there are points on the productivity grid, and with kinks at the optimal income values chosen by each type.[19] Note, however, that when the grid on productivity is coarse, this piecewise linear assumption is grossly at odds with the true shape of the optimal tax schedule (e.g., figure 2B). Thus, as we will shortly show, the method will not work well in such cases.

Figure 3A is analogous to figure 1A; it plots the optimal marginal tax rates as N varies from 10 to $10,000$, using the tax formula approach. Note that unlike the previous case, the decentralizing tax schedule is pre-specified by the piecewise linear function $\tilde{T}$, which implies a corresponding marginal tax schedule that is a step function. When the grid is very fine, optimal tax rates are again insensitive to making the grid still finer. More importantly, they are also identical to those found using the Mirrlees approach plotted in figure 1A. Thus, with a fine grid, the tax formula approach works just fine. However, when the grid is coarse—for example, when $N = 10$—the tax formula approach delivers marginal rates that are much higher at the bottom but lower over the rest of the income distribution than the corresponding marginal rates at the true solution. Moreover, the corresponding candidate optimal allocations are quite different (Figure A1).

We now illustrate why the tax schedule that arises from the tax formula approach fails to deliver the constrained efficient allocation. Note that the tax formula approach specifies the entire tax schedule such that the government budget constraint is satisfied. To check whether these allocations constitute a solution to the original Mirrlees problem, it is necessary only to check whether incentive constraints are appropriately binding.

[19] Since the optimal income allocation y is not known ex ante, it is convenient to think of marginal tax rates as varying by wages, instead of income. Thus, each productivity type faces a constant type-specific marginal tax rate, which makes the individual maximization problem very simple. See, e.g., Mankiw et al. (2009).



(A) Marginal Tax Rate (%)

(B) Income versus Consumption

Figure 3: Optimal Tax Policy Using Tax Formula Approach. Panel A plots the optimal Mirrleesian marginal tax schedule with the number of grid points from 10 to 10,000, using the tax formula approach. Panel B plots the optimal income-consumption pairs when $N = 10$ (red dot) and the indifference curve for each wage type (blue dashed line). The red solid line is the income-consumption menu implied by the optimal tax schedule in panel A.

Figure 3B is analogous to figure 2A. The income-consumption menu off the grid points is pinned down by the red solid line. The figure shows that the local downward incentive constraints are not binding at the conjectured solution, as the indifference curve for type θ_i passes well above (y^*_{i-1}, c^*_{i-1}) (see, e.g., IC_9 and (y^*_8, c^*_8)). Therefore, there exists a feasible allocation that delivers higher social welfare than the candidate allocation $\{y^*_i, c^*_i\}_{i=1}^{10}$ computed following the tax formula approach.[20] Indeed, the candidate allocation generates welfare *losses* of -2.9 percent relative to the baseline HSV function, compared with the welfare gains of 20.1 percent under Mirrlees approach (table 1, column 4).[21]

Another problem with how the tax formula approach is typically implemented is that most papers start from the continuous productivity version of the DS formula (eq. 9) and then

[20] Another problem is that the marginal tax schedule in figure 3A does not even decentralize the candidate allocation $\{y^*_i, c^*_i\}_{i=1}^{10}$. In particular, look at the allocation for type θ_9—i.e., (y^*_9, c^*_9) in figure 3B and IC_9. Since the income-consumption menu goes above IC_9 between (y^*_9, c^*_9) and (y^*_{10}, c^*_{10}), type θ_9 has an incentive to increase income by working more. This problem arises because at the step of computing equilibrium, the algorithm conceptualizes each type as facing a type-specific but income-independent marginal tax rate. The agent with θ_9 therefore does not recognize that a non-marginal increase in earnings might be welfare improving, because it would push him into the top marginal tax bracket, where the marginal rate is zero.

[21] The candidate allocation generates large welfare losses partly because the piecewise linear function imposes zero marginal tax rates in a wide range at the top of the productivity distribution (see figure 3A). In Appendix D.1, we also consider a case in which $\tilde{T}$ is a piecewise quadratic function, so that $\tilde{T}'$ is a piecewise linear function. However, this specification still generates welfare losses of -1.1 percent.

take a discrete approximation to this formula for the purposes of numerical characterization. However, this approach does not deliver the correct optimal tax formula for a discrete productivity environment, which is the one derived from the FOCs to the discrete version of the Mirrlees problem (eq. 13).[22]

In sum, the tax formula approach delivers an accurate solution when the productivity grid is very fine, but it does not work well when the grid is coarse. The allocation to which the tax formula approach converges does not solve the original constrained planner's problem, and the gap in welfare terms is potentially large. This calls into question policy prescriptions based on computations using a tax formula approach and a coarse grid.

How, in practice, can one tell whether a given grid is sufficiently fine? The simple check we propose is to verify that the numerical solution is not materially changed when the number of grid points is increased. We apply this check to the economy in Mankiw et al. (2009). Figure 4 shows marginal tax rates when we use their computer code but make the grid 100 times finer. The resulting marginal tax rates at the bottom of the wage distribution are 10 to 20 percentage points lower than the rates they report, mirroring the finding in figure 3A.[23]

4.3 Differential Equations Approach

We have shown that the coarseness of the grid for productivity determines the extent to which private information constrains the set of feasible allocations. With a coarse grid, the planner faces only a small number of incentive constraints, and can achieve substantial redistribution without large efficiency costs, translating into high average but low marginal tax rates. However, if the true productivity distribution is continuous, then in reality the planner faces more constraints, and numerical allocations obtained assuming a coarse grid will not be feasible. In particular, the true constrained efficient allocation will be further from the first best.

The obvious solution to this problem is to compute the optimum using a very fine grid for productivity. In Sections 4.1 and 4.2, we showed that numerical solutions with $N \geq 1,000$ provide an accurate representation of optimal taxation in an economy with a continuous distribution of productivity, and that this result does not depend on which computational

[22] In particular, term $\tilde{A}$ in eq. (13) involves the parameter defining grid coarseness κ and thus cannot be interpreted purely in terms of labor supply elasticities. In Appendix D.1 we describe the marginal tax schedules that the tax formula approach delivers when we use the discrete approximation of eq. (9), as in Mankiw et al. (2009), as opposed to the correct equation (13).

[23] For the version with a finer grid, we take the baseline specification in Mankiw et al. (2009) with 144 grid points and add 99 points between each original grid point, so we end up with $(144 - 1) \times 100 + 1 = 14,301$ points. The density of each additional point is obtained by shape-preserving piecewise cubic interpolation. Bastani (2015) also studies how optimal marginal tax rates vary with the number of grid points using the economy in Mankiw et al. (2009).

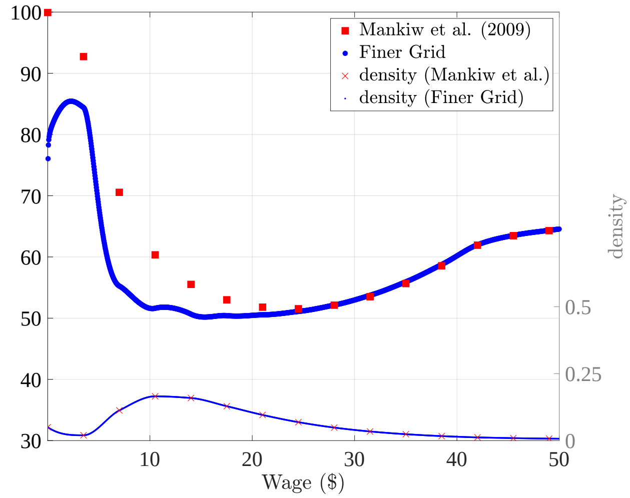


Figure 4: Optimal Tax Policy in Mankiw et al. (2009). The figure plots the optimal marginal tax schedule reported in Mankiw et al. (2009) (red dot) and that with a finer grid (blue line).

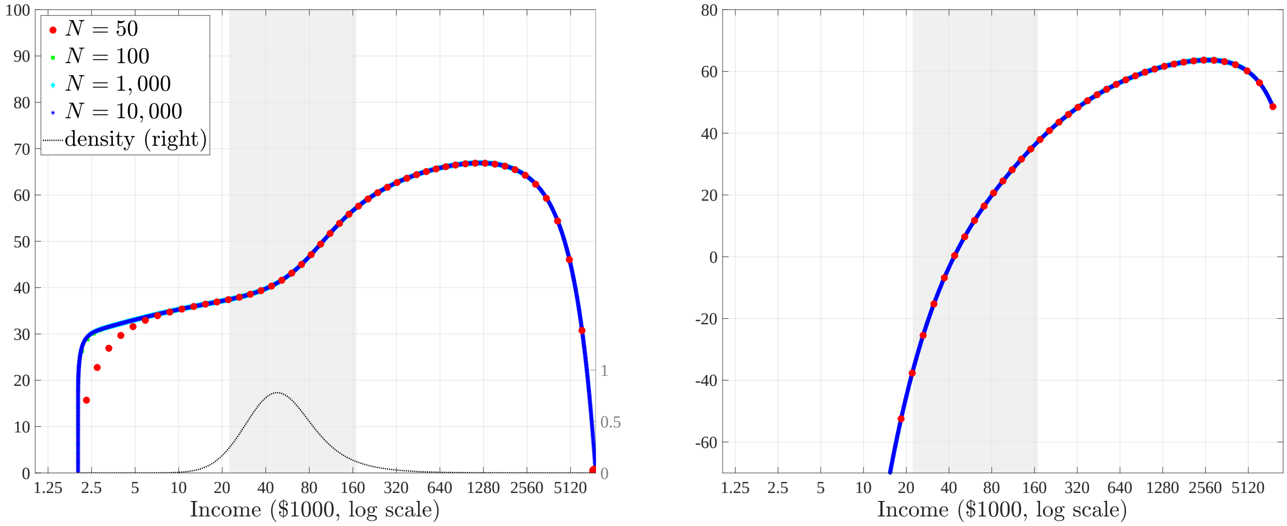
approach one takes. However, it is not always feasible to work with a very fine grid, especially in richer model environments.

Another practical solution is to solve the model without discretization using the system of differential equations (15-17). This *differential equations approach* was first introduced in Mirrlees' original paper and also used by Saez (2001), but the method is less popular than the other numerical methods we have discussed. We suspect that one reason for that is that the ODEs in these earlier papers are formulated in a way that lacks an obvious connection to the familiar Diamond-Saez equation. In contrast, our equation (17) follows directly from the Diamond-Saez equation and treats the familiar marginal tax ratio as the key unknown function. We hope that researchers will find this re-formulation more intuitive and accessible than earlier variants of the differential equations approach, even though it is very similar from a computational perspective.

Because this approach assumes a continuous true productivity distribution, the differential equations embed a continuous set of incentive constraints. And they pin down uniquely how marginal tax rates vary with income, in contrast to the discrete grid approach, where there are different possible ways to "fill in" tax rates in between grid points. We solve our system of ODEs using the 4th order Runge-Kutta method.[24] Figure 5 plots the optimal marginal and average tax rates computed this way.

Implementing the method does require one form of discretization, which is the choice of the step size to be used when evaluating the ODEs. In Figure 5 we use values for N, the number of steps, from 50 to $10,000$. The figure shows that the differential equations

[24] Appendix C.4 provides details on the computation.



(A) Marginal Tax Rate (%) **(B)** Average Tax Rate (%)

Figure 5: Optimal Tax Policy Using Differential Equations Approach. Panels A and B plot the optimal Mirrleesian marginal and average tax schedules with the number of grid points from 50 to 10,000. Panel A also plots the density of income distribution for 10,000 grid points. The area between the 5th and 95th percentiles is shaded gray.

approach works well even with small values for N. For example, the optimal marginal tax schedule is indistinguishable from the true tax schedule when $N = 100$, which is not the case for the previous two approaches (see Figures 1A and 3A). A deviation appears at the bottom of the income distribution when $N = 50$, indicating that the corresponding step size becomes too large to accurately approximate the differential equations. The numerical method breaks down for smaller grid sizes such as $N = 10$. But overall, the differential equations approach works better than the discrete grid methods described previously. Again, the reason is that the equations describing the solution to the Mirrlees problem are derived assuming a continuous productivity distribution and are derived *before* the numerical approximation step.

4.4 Flexible Ramsey Taxation as a Practical Alternative

The safest approach to characterizing the optimal tax and transfer schedule is to work with a very fine grid for productivity or to solve the system of ODEs. What if neither of these approaches is feasible? We want a computational approach that (i) is fast and easy to compute, (ii) delivers a policy prescription close to the true Mirrleesian optimum when the grid is very fine, and (iii) delivers a very similar policy prescription when the productivity distribution is (counterfactually) approximated using a coarse grid.

An approach that delivers on these desiderata is to search for a tax and transfer policy that is optimal within a flexible parametric class a la Ramsey. We focus on the following functional form for taxes net of transfers:

$$T(y) = \phi_0 + \sum_{i=1}^{M} \phi_i y \left(\log y\right)^{i-1},$$

which contains $M + 1$ parameters, $\{\phi_0, \phi_1, ..., \phi_M\}$. This class of tax functions allows for lump-sum taxes or transfers (ϕ_0) and features marginal tax rates that are a polylogarithmic function of income:

$$T'(y) = \sum_{i=1}^{M} \tau_i \left(\log y\right)^{i-1},$$

where the coefficients τ are given by $\tau_i = \phi_i + i\phi_{i+1}$ for $i = 1, ..., M-1$ and $\tau_M = \phi_M$.[25] Note that if the true Mirrleesian optimal marginal tax schedule is a continuous function, then by the Stone-Weierstrass theorem it will be possible to approximate it arbitrarily well, given a large enough choice for M.

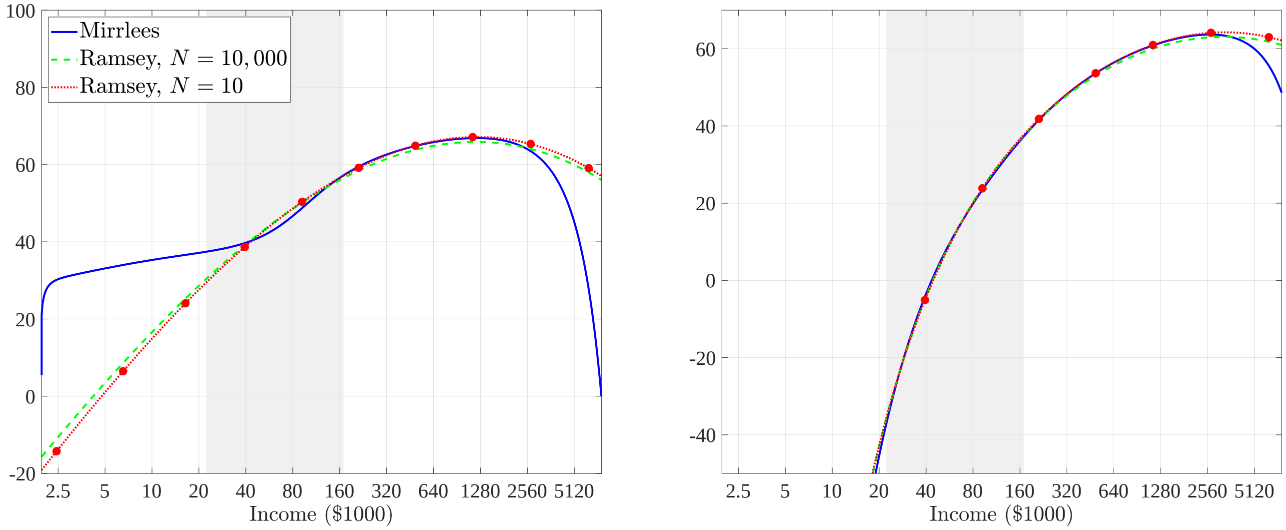
We now consider an example with $M = 4$. In this case, solving for the optimal schedule amounts to searching for the five parameters $\{\phi_0, \tau_1, \tau_2, \tau_3, \tau_4\}$ that maximize social welfare subject to government budget balance.[26]

Figure 6 compares optimal marginal and average tax rates under Ramsey taxation to those from the Mirrlees taxation problem. It shows that the Ramsey tax policy is a very good approximation of that in the Mirrleesian optimum, especially over the shaded range of productivity values where the vast majority of the population is located. More importantly, even with a coarse grid of $N = 10$, the best Ramsey policy remains very close to the true Mirrleesian policy. The welfare gains from switching to the Ramsey optimum, relative to the HSV baseline, are 2.0 and 1.9 percent of consumption for $N = 10,000$ and $N = 10$, respectively, compared with 2.1 percent under the true Mirrleesian optimal policy (table 1, column 5).[27]

[25] An alternative would be to specify taxes as a polynomial function of *level* income. However, income can take very high values, and marginal tax rates would tend to explode unless the coefficients on higher order level terms were near zero.

[26] Appendix C.5 provides the details of the computation. The optimal tax parameters are given by $\phi_0 =$ \$10,609 and $\tau = (0.489, 0.119, -0.018, -0.001)$ for the case of $N = 10,000$, and $\phi_0 =$ \$10,153 and $\tau = (0.490, 0.126, -0.019, -0.001)$ for the case of $N = 10$.

[27] Appendix D.2 shows that the consumption and earnings under the Ramsey tax policy are very close to those under the Mirrlees policy.



(A) Marginal Tax Rate (%) **(B)** Average Tax Rate (%)

Figure 6: Optimal Ramsey Tax Policy. Panels A and B compare the optimal Ramsey marginal and average tax schedules with the number of grid points of 10 and 10,000 to the optimal Mirrlees tax schedules.

5 Conclusions

We considered the canonical static Mirrleesian economy and described how to decentralize the optimal allocation via a tax function. We showed that in this class of models with incentive constraints, the representation of the productivity distribution is an important part of the model environment, because it controls the strength of information frictions. Standard approaches to computing the optimal tax schedule do not deliver useful policy guidance if they assume too coarse a grid for productivity values. To accurately characterize the optimal tax and transfer schedule, one should either work with a very fine productivity grid or alternatively solve the model as a system of ordinary differential equations that presumes a continuous productivity distribution. If neither of these options is feasible, we recommend searching for the optimal tax policy within a parametric polynomial class.

References

Bastani, S. (2015): "Using the Discrete Model to Derive Optimal Income Tax Rates," *FinanzArchiv: Public Finance Analysis*, 71, 106–117.

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

Blundell, R. and A. Shephard (2012): "Employment, Hours of Work and the Optimal Taxation of Low-Income Families," *Review of Economic Studies*, 79, 481–510.

Boar, C. and V. Midrigan (2021): "Markups and Inequality," mimeo.

Brewer, M., E. Saez, and A. Shephard (2010): "Means Testing and Tax Rates on Earnings," in *Dimensions of Tax Design: the Mirrlees Review*, Oxford University Press, 90–173.

Chang, Y. and Y. Park (2020): "Optimal Taxation with Private Insurance," *Review of Economic Studies*.

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

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

Golosov, M., M. Troshkin, and A. Tsyvinski (2016): "Redistribution and Social Insurance," *American Economic Review*, 106, 359–386.

Heathcote, J., K. Storesletten, and G. L. Violante (2017): "Optimal Tax Progressivity: An Analytical Framework," *Quarterly Journal of Economics*, 132, 1693–1754.

Heathcote, J. and H. Tsujiyama (2021): "Optimal Income Taxation: Mirrlees Meets Ramsey," *Journal of Political Economy*, 129, 3141–3184.

Homburg, S. (2001): "The Optimal Income Tax: Restatement and Extensions," *FinanzArchiv / Public Finance Analysis*, 58, 363–395.

Huggett, M. and J. C. Parra (2010): "How Well Does the U.S. Social Insurance System Provide Social Insurance?" *Journal of Political Economy*, 118, 76–112.

Kocherlakota, N. R. (2010): *The New Dynamic Public Finance*, Princeton University Press.

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

Mankiw, N. G., M. Weinzierl, and D. Yagan (2009): "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, 23, 147–174.

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

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

Practical Optimal Income Taxation

Appendix (For Online Publication Only)

Jonathan Heathcote and Hitoshi Tsujiyama

Contents

A Diamond-Saez Formulae

In this section, we derive the Diamond-Saez formula (9) and show how to derive its discrete-state version (13).

A.1 Derivation of Diamond-Saez Formula (9)

We derive the Diamond-Saez formula for our economy with a continuous productivity distribution. Reproducing the Mirrlees planner's problem from eqs. (4-6), we have

$$\begin{cases} \max\limits_{\{c(\theta),y(\theta)\}_{\theta\in\theta}} & \int \left[\log\left(c(\theta)\right) - \frac{1}{1+\sigma}\left(\frac{y(\theta)}{\theta}\right)^{1+\sigma}\right] dF\left(\theta\right) \\ \text{s.t.} & \log\left(c(\theta)\right) - \frac{1}{1+\sigma}\left(\frac{y(\theta)}{\theta}\right)^{1+\sigma} \geq \log\left(c(\tilde{\theta})\right) - \frac{1}{1+\sigma}\left(\frac{y(\tilde{\theta})}{\theta}\right)^{1+\sigma} \qquad \text{for all } \theta \text{ and } \tilde{\theta}, \\ & \int \left[y(\theta) - c(\theta)\right] dF\left(\theta\right) - G \geq 0. \end{cases}$$

The IC constraints state

$$U(\theta) \equiv \log\left(c(\theta)\right) - \frac{1}{1+\sigma}\left(\frac{y(\theta)}{\theta}\right)^{1+\sigma} = \max_{\tilde{\theta}} \log\left(c(\tilde{\theta})\right) - \frac{1}{1+\sigma}\left(\frac{y(\tilde{\theta})}{\theta}\right)^{1+\sigma}.$$

We have

$$\begin{aligned} U'(\theta) &= c(\theta)^{-1}c'(\theta) - \left[\frac{1}{\theta^{1+\sigma}}y(\theta)^{\sigma}y'(\theta) - \frac{1}{\theta^{2+\sigma}}y(\theta)^{1+\sigma}\right] \\ &= \frac{1}{\theta^{2+\sigma}}y(\theta)^{1+\sigma}, \end{aligned}$$

where the second line uses the envelope condition:

$$c(\theta)^{-1}c'(\theta) - \frac{1}{\theta^{1+\sigma}}y(\theta)^{\sigma}y'(\theta) = 0.$$

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

$$\begin{cases} \max\limits_{\{U(\theta),y(\theta)\}} & \int U(\theta) dF\left(\theta\right) \\ \text{s.t.} & U'(\theta) = \frac{1}{\theta^{2+\sigma}}y(\theta)^{1+\sigma} \quad \text{for all } \theta, \\ & \int \left[y(\theta) - c(\theta; U, y)\right] dF\left(\theta\right) - G \geq 0, \end{cases}$$

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

$$\mathcal{H} \equiv \{U(\theta) + \zeta [y(\theta) - c(\theta; U, y) - G]\} f(\theta) + \mu(\theta) \frac{1}{\theta^{2+\sigma}} y(\theta)^{1+\sigma}.$$

By optimal control, the following equations must hold

$$\begin{cases} 0 = \zeta \left[1 - c(\theta) \frac{1}{\theta^{1+\sigma}} y(\theta)^{\sigma}\right] f(\theta) + \mu(\theta) \frac{1+\sigma}{\theta^{2+\sigma}} y(\theta)^{\sigma}, \\ -\mu'(\theta) = [1 - \zeta c(\theta)] f(\theta), \\ \mu(0) = \mu(\infty) = 0. \end{cases} \tag{A1}$$

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

$$\mu(\theta) = \int_{\theta}^{\infty} [1 - \zeta c(s)] \, dF(s). \tag{A2}$$

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

$$\zeta = \frac{1}{\int c(s) dF(s)}. \tag{A3}$$

We now consider the decentralization via income taxes. Using the FOC (8), the first equation in (A1) can be written as

$$0 = \zeta T'(y(\theta)) f(\theta) + \mu(\theta) [1 - T'(y(\theta))] \frac{c(\theta)^{-1}}{\theta} (1 + \sigma), \tag{A4}$$

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

$$\frac{T'(y(\theta))}{1 - T'(y(\theta))} = (1 + \sigma) \frac{1 - F(\theta)}{\theta f(\theta)} \int_{\theta}^{\infty} \left[1 - \frac{\int c(x) \, dF(x)}{c(s)}\right] \frac{c(s)}{c(\theta)} \frac{dF(s)}{1 - F(\theta)}.$$

A.2 Proof of Proposition 1

All local downward incentive constraints are binding at the optimum (Carroll, 2012). The discretized version of the planner's problem is

$$\begin{cases} \max_{\{c_i, y_i\}_{i=1}^{N}} & \sum_i \pi_i \left[\log(c_i) - \frac{1}{1+\sigma} \left(\frac{y_i}{\theta_i}\right)^{1+\sigma}\right] \\ \text{s.t.} & \log(c_i) - \frac{1}{1+\sigma} \left(\frac{y_i}{\theta_i}\right)^{1+\sigma} = \log(c_{i-1}) - \frac{1}{1+\sigma} \left(\frac{y_{i-1}}{\theta_i}\right)^{1+\sigma} \qquad \text{for } i = 2, \cdots, N, \\ & \sum_i \pi_i c_i + G = \sum_i \pi_i y_i. \end{cases}$$

Denoting by μ_i and ζ the corresponding multipliers, FOCs with respect to c_i and y_i are given by

$$\begin{aligned} c_i^{-1}\left(\pi_i + \mu_i - \mu_{i+1}\right) &= \zeta \pi_i, && \text{(A5)} \\ \frac{y_i^{\sigma}}{\theta_i^{1+\sigma}}\left(\pi_i + \mu_i\right) - \frac{y_i^{\sigma}}{\theta_{i+1}^{1+\sigma}}\mu_{i+1} &= -\zeta \pi_i. && \text{(A6)} \end{aligned}$$

Note that $\mu_1 = 0$ (no incentive constraint for the least productive type) and $\mu_{N+1} = 0$.

First we solve for multipliers. Adding up eq. (A5) gives

$$\zeta \sum_i c_i \pi_i = \sum_i \left(\pi_i + \mu_i - \mu_{i+1}\right) = 1.$$

This leads

$$\zeta = \frac{1}{\sum_i c_i \pi_i} = \frac{1}{\mathbb{E}\left[c_i\right]}.$$

We also rewrite eq. (A5) as

$$\mu_i = \left(\zeta c_i - 1\right)\pi_i + \mu_{i+1}.$$

We can iterate this equation forwards to get

$$\begin{aligned} \mu_i &= \left(\zeta c_i - 1\right)\pi_i + \left(\zeta c_{i+1} - 1\right)\pi_{i+1} + \mu_{i+2} \\ &= \cdots \\ &= \sum_{s=i}^{N} \pi_s \left(\zeta c_s - 1\right). \end{aligned}$$

Next, combining eqs. (A5-A6), we have

$$\begin{aligned} c_i^{-1}\left(\pi_i + \mu_i - \mu_{i+1}\right) &= -\frac{y_i^{\sigma}}{\theta_i^{1+\sigma}}\left(\pi_i + \mu_i\right) + \frac{y_i^{\sigma}}{\theta_{i+1}^{1+\sigma}}\mu_{i+1}, \\ &= -\frac{y_i^{\sigma}}{\theta_i^{1+\sigma}}\left(\pi_i + \mu_i - \mu_{i+1}\right) - \left(\frac{y_i^{\sigma}}{\theta_i^{1+\sigma}} - \frac{y_i^{\sigma}}{\theta_{i+1}^{1+\sigma}}\right)\mu_{i+1}, \\ c_i^{-1}\left[1 - \frac{\frac{y_i^{\sigma}}{\theta_{i+1}^{1+\sigma}} - \frac{y_i^{\sigma}}{\theta_i^{1+\sigma}}}{c_i^{-1}\left(\pi_i + \mu_i - \mu_{i+1}\right)}\mu_{i+1}\right] &= -\frac{y_i^{\sigma}}{\theta_i^{1+\sigma}}, \\ c_i^{-1}\left[1 - \left(\frac{y_i^{\sigma}}{\theta_{i+1}^{1+\sigma}} - \frac{y_i^{\sigma}}{\theta_i^{1+\sigma}}\right)\frac{\mu_{i+1}}{\zeta \pi_i}\right] &= -\frac{y_i^{\sigma}}{\theta_i^{1+\sigma}}. \end{aligned}$$

The FOC of the decentralized economy is given by

$$c_i^{-1}\left(1-T_i^{*\prime}\right)=-\frac{y_i^{\sigma}}{\theta_i^{1+\sigma}}.$$

Thus the marginal tax rate ratio for type θ_i is

$$\begin{aligned}\frac{T_i^{*\prime}}{1-T_i^{*\prime}} &= \frac{\left(\frac{y_i^{\sigma}}{\theta_{i+1}^{1+\sigma}}-\frac{y_i^{\sigma}}{\theta_i^{1+\sigma}}\right)\frac{\mu_{i+1}}{\zeta\pi_i}}{-c_i\frac{y_i^{\sigma}}{\theta_i^{1+\sigma}}} \\ &= \left[1-\left(\frac{\theta_{i+1}}{\theta_i}\right)^{-(1+\sigma)}\right]c_i^{-1}\frac{\mu_{i+1}}{\zeta\pi_i} \\ &= \left[1-\kappa^{-(1+\sigma)}\right]c_i^{-1}\frac{\mu_{i+1}}{\zeta\pi_i}.\end{aligned}$$

where $\kappa=\theta_{i+1}/\theta_i$.

Substituting the expressions for μ_i and ζ, we have

$$\begin{aligned}\frac{T_i^{*\prime}}{1-T_i^{*\prime}} &= \left[1-\kappa^{-(1+\sigma)}\right]c_i^{-1}\frac{\mathbb{E}\left[c_i\right]}{\pi_i}\sum_{s=i+1}^{N}\pi_s\left(\frac{c_s}{\mathbb{E}\left[c_i\right]}-1\right) \\ &= \left[1-\kappa^{-(1+\sigma)}\right]\frac{1}{\pi_i}\sum_{s=i+1}^{N}\pi_s\left(1-\frac{\mathbb{E}\left[c_i\right]}{c_s}\right)\frac{c_s}{c_i}.\end{aligned}$$

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

B Construction of the System of ODEs

In this section, we provide the details of the derivation for Section 2.5.

Derivation of the System of ODEs (15) and (17). We take utility $U(\theta)$ as a state and labor supply $h(\theta)=y(\theta)/\theta$ as a control. The incentive constraint (6) states

$$U(\theta)=\max_{\tilde{\theta}}\left\{\log\left(c(\tilde{\theta})\right)-\frac{1}{1+\sigma}\left(\frac{y(\tilde{\theta})}{\theta}\right)^{1+\sigma}\right\}.$$

Using the envelope condition, we have eq. (15):

$$U'(\theta)=\frac{1}{\theta}h(\theta)^{1+\sigma}.$$

Next, we take the DS formula (9) and define the optimal marginal tax ratio $Q(\theta)=$

$\frac{T'(\theta)}{1-T'(\theta)}$:

$$\begin{aligned} Q(\theta) &= (1+\sigma)\frac{1}{\theta f(\theta)}\int_\theta^\infty \left(1-\frac{\zeta}{c(s)}\right)\frac{c(s)}{c(\theta)}dF(s) \\ &= (1+\sigma)\frac{1}{\theta f(\theta)}\frac{1}{c(\theta)}\left[\int_\theta^\infty c(s)dF(s) - \frac{1}{\zeta}(1-F(\theta))\right]. \end{aligned} \tag{A7}$$

Taking the derivative, we obtain eq. (17):

$$Q'(\theta) = -Q(\theta)\left[\frac{1}{\theta}+\frac{f'(\theta)}{f(\theta)}+\frac{c'(\theta)}{c(\theta)}\right]+\frac{1+\sigma}{\theta}\left(\frac{1}{\zeta c(\theta)}-1\right). \tag{A8}$$

The consumption function and its derivative are implicitly given by

$$\begin{aligned} c(\theta) &= \exp\left\{U(\theta)+\frac{1}{1+\sigma}h(\theta)^{1+\sigma}\right\}, \quad \text{and} \\ c'(\theta) &= c(\theta)\left[U'(\theta)+h(\theta)^\sigma h'(\theta)\right]. \end{aligned}$$

To use eq. (A8) in the computation, we want to eliminate the endogenous variable $h'(\theta)$. Taking the derivative of eq. (16), we have

$$Q'(\theta) = \frac{1}{c(\theta)h(\theta)^\sigma} - \sigma\frac{\theta}{c(\theta)h(\theta)^{\sigma+1}}h'(\theta) - \frac{\theta}{c(\theta)h(\theta)^\sigma}\frac{c'(\theta)}{c(\theta)}.$$

Simple algebra yields

$$h'(\theta) = \frac{-Q'(\theta)c(\theta)h(\theta)^\sigma + 1 - \theta U'(\theta)}{\theta\left(\frac{\sigma}{h(\theta)}+h(\theta)^\sigma\right)}. \tag{A9}$$

Substituting these expressions into eq. (A8), we have

$$Q'(\theta) = \left[-Q(\theta)\left(\frac{1}{\theta}+\frac{f'(\theta)}{f(\theta)}+\frac{1+\sigma}{\theta\left(\frac{\sigma}{h(\theta)^{\sigma+1}}+1\right)}\right)+\frac{1+\sigma}{\theta}\left(\frac{1}{\zeta c(\theta)}-1\right)\right]\left(\frac{\sigma+h(\theta)^{\sigma+1}}{\sigma+\frac{1}{\theta}c(\theta)h(\theta)^{2\sigma+1}}\right). \tag{A10}$$

Alternative Formulations. We derived eq. (A7) from the DS formula, but it can also be obtained directly from the planner's optimality condition (A1). Using eqs (A1-A2), we have

$$0 = \zeta\left[1-c(\theta)\frac{1}{\theta}h(\theta)^\sigma\right]f(\theta)+\frac{1+\sigma}{\theta^2}h(\theta)^\sigma\int_\theta^\infty\left[1-\zeta c(s)\right]dF(s). \tag{A11}$$

Using eq. (16) and rearranging terms yields eq. (A7).

There are alternative formulations that can also be used in the computation. We have verified that using them is equally stable. Here we derive expressions analogous to those in Mirrlees (1971) and Saez (2001).

Mirrlees (1971) derives a similar expression in a somewhat ad hoc way. Define

$$Q_M(\theta) \equiv \frac{1 - \frac{1}{\theta} c(\theta) h(\theta)^{\sigma}}{(1+\sigma)\, h(\theta)^{\sigma}}.$$

Substituting this into eq. (A11) and rearranging terms,

$$Q_M(\theta) = \frac{1}{\theta^2 f(\theta)} \left[\int_{\theta}^{\infty} c(s) dF(s) - \frac{1}{\zeta} (1 - F(\theta)) \right].$$

Taking the derivative, we obtain

$$Q_M'(\theta) = -\frac{Q_M(\theta)}{\theta} \left[\frac{\theta f'(\theta)}{f(\theta)} + 2 \right] - \frac{1}{\theta^2} \left(1 - \frac{1}{\zeta} \right), \tag{A12}$$

which is similar to eq. (A8) and analogous to eq. (50) in Mirrlees (1971). Although there is no clear economic interpretation for Q_M, the ODE is simpler than eq. (A10), as it does not involve the consumption function. So we will use it in the computation in Section 4.3.

Likewise, Saez (2001) defines

$$Q_S(\theta) \equiv \frac{\theta - c(\theta) h(\theta)^{\sigma}}{(1+\sigma)\, h(\theta)^{\sigma}}.$$

Substituting this into eq. (A11) and rearranging terms,

$$Q_S(\theta) = \frac{1}{\theta f(\theta)} \left[\int_{\theta}^{\infty} c(s) dF(s) - \frac{1}{\zeta} (1 - F(\theta)) \right].$$

Taking the derivative, we obtain

$$Q_S'(\theta) = -\frac{Q_S(\theta)}{\theta} \left[\frac{\theta f'(\theta)}{f(\theta)} + 1 \right] - \frac{1}{\theta} \left(c(\theta) - \frac{1}{\zeta} \right), \tag{A13}$$

which is analogous to the expression on p. 228 in Saez (2001).

C Computational Method

This section explains the details of the computational methods for each numerical approach.

C.1 Units

We define hours worked in the model by y/θ. We denote by Y and H average earnings and average hours worked in the baseline model. For the purpose of comparing model to data it is convenient to rescale model units. We target average annual earnings in the 2007 SCF household sample, which is $\bar{Y} = \$77,326$, and average household hours, which is $\bar{H} = 3,075$ in 2007 in the Current Population Survey. When plotting model allocations we scale model earnings, consumption and taxes by a factor $\bar{Y}/Y$, and wages by $(\bar{Y}/Y)/(\bar{H}/H)$.

C.2 Mirrlees Approach

With the preference class we consider, the local incentive compatibility constraints are necessary and sufficient for the global incentive compatibility constraints (12) (Carroll, 2012). We thus replace them with local incentive compatibility constraints:

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

These are downward incentive constraints (DIC) and upward incentive constraints (UIC), respectively. It is well-known that the DICs are binding at optimum; otherwise the planner can improve the welfare by transferring a small amount of consumption goods from a high-productivity agent to a low-productivity agent.

Next, we show that income is non-decreasing in i if and only if the UICs are satisfied. Suppose the UIC for type θ_i is satisfied. Since the DIC for type θ_i is binding, we have

$$\log(c_{i-1}) - \frac{\left(\frac{y_{i-1}}{\theta_{i-1}}\right)^{1+\sigma}}{1+\sigma} + \log(c_i) - \frac{\left(\frac{y_i}{\theta_i}\right)^{1+\sigma}}{1+\sigma} \geq \log(c_i) - \frac{\left(\frac{y_i}{\theta_{i-1}}\right)^{1+\sigma}}{1+\sigma} + \log(c_{i-1}) - \frac{\left(\frac{y_{i-1}}{\theta_i}\right)^{1+\sigma}}{1+\sigma}.$$

Rearranging terms, we have the weak monotonicity of income $y_i \geq y_{i-1}$. Conversely, suppose $y_i \geq y_{i-1}$. If $y_i = y_{i-1}$, then the UIC for type θ_i is trivially satisfied with equality. If instead $y_i > y_{i-1}$, then the UIC for type θ_i is slack, because

$$\begin{aligned} \frac{\left(\frac{y_{i-1}}{\theta_i}\right)^{1+\sigma}}{1+\sigma} + \frac{\left(\frac{y_i}{\theta_{i-1}}\right)^{1+\sigma}}{1+\sigma} &> \frac{\left(\frac{y_{i-1}}{\theta_{i-1}}\right)^{1+\sigma}}{1+\sigma} + \frac{\left(\frac{y_i}{\theta_i}\right)^{1+\sigma}}{1+\sigma}, \\ \log(c_i) + \log(c_{i-1}) + \frac{\left(\frac{y_{i-1}}{\theta_i}\right)^{1+\sigma}}{1+\sigma} + \frac{\left(\frac{y_i}{\theta_{i-1}}\right)^{1+\sigma}}{1+\sigma} &> \log(c_i) + \log(c_{i-1}) + \frac{\left(\frac{y_{i-1}}{\theta_{i-1}}\right)^{1+\sigma}}{1+\sigma} + \frac{\left(\frac{y_i}{\theta_i}\right)^{1+\sigma}}{1+\sigma}, \\ \log(c_{i-1}) - \frac{\left(\frac{y_{i-1}}{\theta_{i-1}}\right)^{1+\sigma}}{1+\sigma} &> \log(c_i) - \frac{\left(\frac{y_i}{\theta_{i-1}}\right)^{1+\sigma}}{1+\sigma}. \end{aligned}$$

Thus the planner's problem can be written as

$$\begin{cases} \max\limits_{\{c_i,y_i\}_{i=1}^N} & \sum_i \pi_i \left[\log(c_i) - \frac{\left(\frac{y_i}{\theta_i}\right)^{1+\sigma}}{1+\sigma}\right] & \\ \text{s.t.} & \log(c_i) - \frac{\left(\frac{y_i}{\theta_i}\right)^{1+\sigma}}{1+\sigma} \geq \log(c_{i-1}) - \frac{\left(\frac{y_{i-1}}{\theta_i}\right)^{1+\sigma}}{1+\sigma}, & \text{for } i = 2, ..., N \\ & y_i \geq y_{i-1}, & \text{for } i = 2, ..., N \\ & \sum_i \pi_i y_i \geq \sum_i \pi_i c_i + G. & \end{cases}$$

We use forward iteration (forward from θ_1 to θ_N) to search for an allocation that satisfies all the first-order conditions, the incentive constraints, and the resource constraint. Income is non-decreasing in wages, and thus the resulting allocation is optimal given that our utility function exhibits the single-crossing property.

Finally, we find the optimal marginal tax rate $T_i^{*\prime}$ for each i, using eq. (18).

C.3 Tax Formula Approach

Given the Diamond-Saez formula (13), we solve the fixed point problem to find T.

Following Mankiw et al. (2009), we approximate T with a piecewise linear tax function $\tilde{T}$ and search for an equilibrium allocation $\{y_i^*, c_i^*\}_{i=1}^N$ that satisfies eq. (13) along with $\tilde{T}$. Since the optimal income allocation y^* is not known ex ante, it is convenient to consider $\tilde{T}$ as a function of wage, instead of income. This way also allows bunching in the optimal allocation, allowing for different marginal tax rates for the same income but different wage levels.

The piecewise linear tax function is characterized by $\tilde{T}(0)$ and tax rates $\tilde{T}_i'$ for each wage grid point i. Given $\tilde{T}$, we find (y_i^*, c_i^*) for any i such that the individual FOC and the budget constraint are satisfied:

$$\begin{aligned} c_i^{*-\gamma}\theta_i\left(1 - \tilde{T}_i'\right) &= \left(\frac{y_i^*}{\theta_i}\right)^{\sigma}, \\ c_i^* &= y_i^* - \tilde{T}(y_i^*). \end{aligned}$$

We then adjust $\tilde{T}(0)$ so that the government budget constraint is satisfied: $\sum_{i=1}^N \pi_i \tilde{T}(y_i^*) = G$.

Given the equilibrium allocation $\{y_i^*, c_i^*\}_{i=1}^N$, we compute new tax rates for each grid using the Diamond-Saez formula (13). We iterate this process until the new tax rates get sufficiently close to the rates in the previous iteration.

In Appendix D.1, we also approximate T with a piecewise quadratic tax function so that

T' is a piecewise linear function.

C.4 Differential Equations Approach

The system of differential equations is given by eqs. (15) and (17). We can solve this, taking utility U as the state and labor supply h as the control. Since eq. (17) contains an endogenous variable ($h'(\theta)$), we must use the modified expression (A10).

The boundary conditions are given by $Q(\theta) \to 0$ as $\theta \to 0$, and $Q(\theta)\theta f(\theta) \to 0$ as $\theta \to \infty$, which correspond to the third condition in eq. (A1).

Here is the algorithm. We first guess the value of ζ and the labor supply of the least productive type $h(\theta_1)$. Using the definition of Q and the boundary condition above, we can solve for $U(\theta_1)$ and $c(\theta_1)$. We then solve the system of ODEs using the 4th order Runge-Kutta method. To address the potential for bunching, we check at each step whether the income is non-decreasing. If not, we set consumption and income constant in the range until income starts to increase. Finally, we adjust the initial guesses to satisfy the other constraint and the resource constraint.

When we apply the Runge-Kutta method, we first denote eqs. (15), (17) and (A9) as

$$\begin{aligned} D_U(\theta, h) &= \frac{dU}{d\theta}, \\ D_Q\left(\theta, Q, U, h, \frac{dU}{d\theta}, \frac{dh}{d\theta}\right) &= \frac{dQ}{d\theta}, \\ D_h\left(\theta, U, h, \frac{dQ}{d\theta}, \frac{dU}{d\theta}\right) &= \frac{dh}{d\theta}. \end{aligned}$$

Using this notation, for $x \in \{Q, U\}$ we use

$$x\left(\theta + d\theta\right) = x\left(\theta\right) + \frac{d\theta}{6}(k_1^x + 2k_2^x + 2k_3^x + k_4^x),$$

where $d\theta$ is the step size, and

$$\begin{aligned}
k_1^Q &= D_Q(\theta, Q, U, h), \\
k_2^Q &= D_Q\left(\theta + \frac{d\theta}{2}, Q + \frac{k_1^Q}{2}d\theta, U + \frac{k_1^U}{2}d\theta, h + \frac{k_1^h}{2}d\theta\right), \\
k_3^Q &= D_Q\left(\theta + \frac{d\theta}{2}, Q + \frac{k_2^Q}{2}d\theta, U + \frac{k_2^U}{2}d\theta, h + \frac{k_2^h}{2}d\theta\right), \\
k_4^Q &= D_Q\left(\theta + d\theta, Q + k_3^Q d\theta, U + k_3^U d\theta, h + k_3^h d\theta\right), \\
k_1^U &= D_U(\theta, h), \\
k_2^U &= D_U\left(\theta + \frac{d\theta}{2}, h + \frac{k_1^h}{2}d\theta\right), \\
k_3^U &= D_U\left(\theta + \frac{d\theta}{2}, h + \frac{k_2^h}{2}d\theta\right), \\
k_4^U &= D_U\left(\theta + d\theta, h + k_3^h d\theta\right), \\
k_1^h &= D_h\left(\theta, U, h, \frac{dQ}{d\theta}, \frac{dU}{d\theta}\right), \\
k_2^h &= D_h\left(\theta + \frac{d\theta}{2}, U + \frac{k_1^U}{2}d\theta, h + \frac{k_1^h}{2}d\theta, k_1^Q, k_1^U\right), \\
k_3^h &= D_h\left(\theta + \frac{d\theta}{2}, U + \frac{k_2^U}{2}d\theta, h + \frac{k_2^h}{2}d\theta, k_2^Q, k_2^U\right).
\end{aligned}$$

To facilitate comparison with other numerical approaches, the step size $d\theta$ is determined by the grid points specified in the calibration.

We can also use an alternative expression, (A12) or (A13), instead of eq. (17). In particular, eq. (A12) is the simplest as it does not include the consumption function, so we use it in Section 4.3. We have verified that using a different expression gives the same result and is equally stable.

C.5 Ramsey Taxation

Consider a third-order polylogarithmic marginal tax function

$$T'(y) = \tau_1 + \tau_2 \log(y) + \tau_3 \log(y)^2 + \tau_4 \log(y)^3.$$

The tax function is given by

$$T(y) = \phi_0 + \phi_1 y + \phi_2 y (\log y) + \phi_3 y (\log y)^2 + \phi_4 y (\log y)^3,$$

where

$$\begin{aligned}
\tau_1 &= \phi_1 + \phi_2, \\
\tau_2 &= \phi_2 + 2\phi_3, \\
\tau_3 &= \phi_3 + 3\phi_4, \\
\tau_4 &= \phi_4.
\end{aligned}$$

With polynomial tax systems, the households' FOCs are not sufficient in general. However, it is possible to prove that marginal utility is decreasing in income at sufficiently high income levels. Hence, for a given tax system, equilibrium allocations can be found by evaluating all roots of the household first-order necessary conditions in the range $[0, y]$ with y sufficiently large.

Solving for the optimal schedule amounts to searching for the five parameters $(\phi_0, \tau_1, \tau_2, \tau_3, \tau_4)$ that maximize social welfare. We search for these parameters using the Nelder-Mead simplex method. Note that ϕ_0 is chosen to close the government budget constraint. We check that the social welfare maximizing policy is independent of the initial set of tax parameters used to start the search process.

D Additional Results

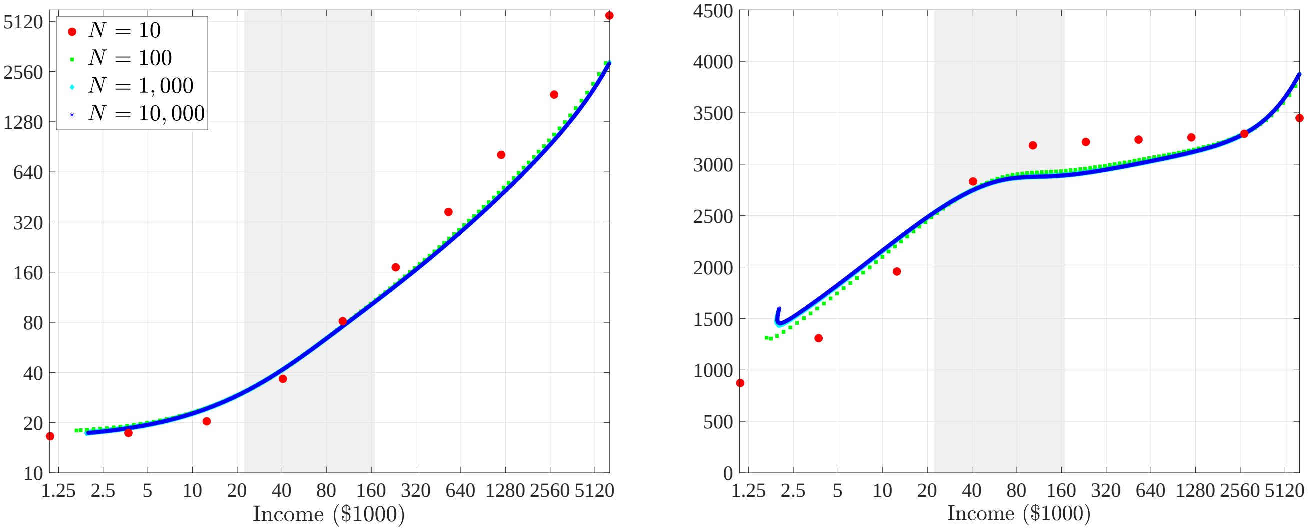
D.1 Additional Results of Tax Formula Approach

Figure A1 plots the optimal allocation of consumption and hours worked, i.e., y/θ, as the number of grid points varies from 10 to 10,000, using the tax formula approach.

In Section 4.2 we computed optimal tax schedules by the tax formula approach using the discrete-state version of the DS formula (13). In the following, we present additional results of the tax formula approach using different approximations to the DS formula.

Mankiw et al. Approximation. Following Mankiw et al. (2009), we approximate the continuous DS formula (9), approximating only term B in eq. (9) by term $\tilde{B}$ in eq. (14), while using term A in eq. (9). As we discussed, the value for term $\tilde{A}$ in eq. (14) falls drastically as the number of grid points is reduced. Therefore, this approximation would find an optimal marginal tax schedule that is very different from the true one.

Figure A2A plots optimal marginal tax rates. Compared to figure 3A, we find that the marginal tax rates are much higher, indicating that how one approximates the DS formula matters a lot quantitatively.



(A) Consumption ($1,000)

(B) Hours Worked

Figure A1: Optimal Allocation Using Tax Formula Approach. Panels A and B plot the optimal allocation (consumption c in panel A and hours worked y/θ in panel B) under the Ramsey optimal policy with the number of grid points of 10 and 10,000, using the tax formula approach.

Piecewise Quadratic Approximation. We also consider a case in which $\tilde{T}$ is a piecewise quadratic function so that $\tilde{T}'$ is a piecewise linear function. The piecewise linear marginal tax function is characterized by tax rates $\tilde{T}'_i$ for each wage grid point i. We assume that for the income values between zero and y_1^*, the marginal tax rate is constant at $\tilde{T}'_1$.

Figure A2B is analogous to figure 3A, plotting the optimal marginal tax rates as the number of grid points varies from 10 to $10,000$. Again, with a fine grid, the tax formula approach works just fine. Also, when $N = 10$, the tax formula approach delivers much higher marginal rates than the true solution. The welfare gains in this case change from 2.04% when $N = 10,000$ to -1.14% when $N = 10$.

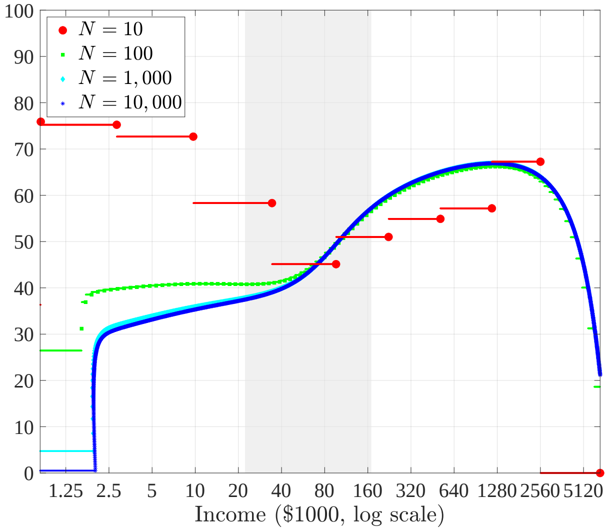
D.2 Allocation under Ramsey Policy versus Mirrlees policy

Figure A3 shows that consumption and hours worked under the Ramsey optimal policy are very close to those under the Mirrlees optimal policy.

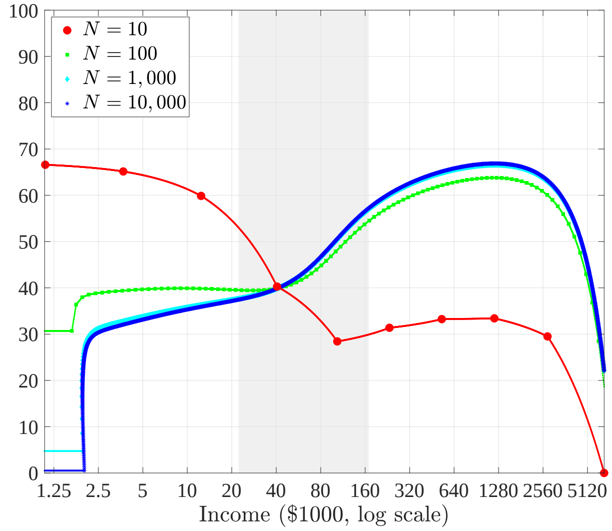
References

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

Mankiw, N. G., M. Weinzierl, and D. Yagan (2009): "Optimal Taxation in Theory



(A) Marginal Tax Rate (%)



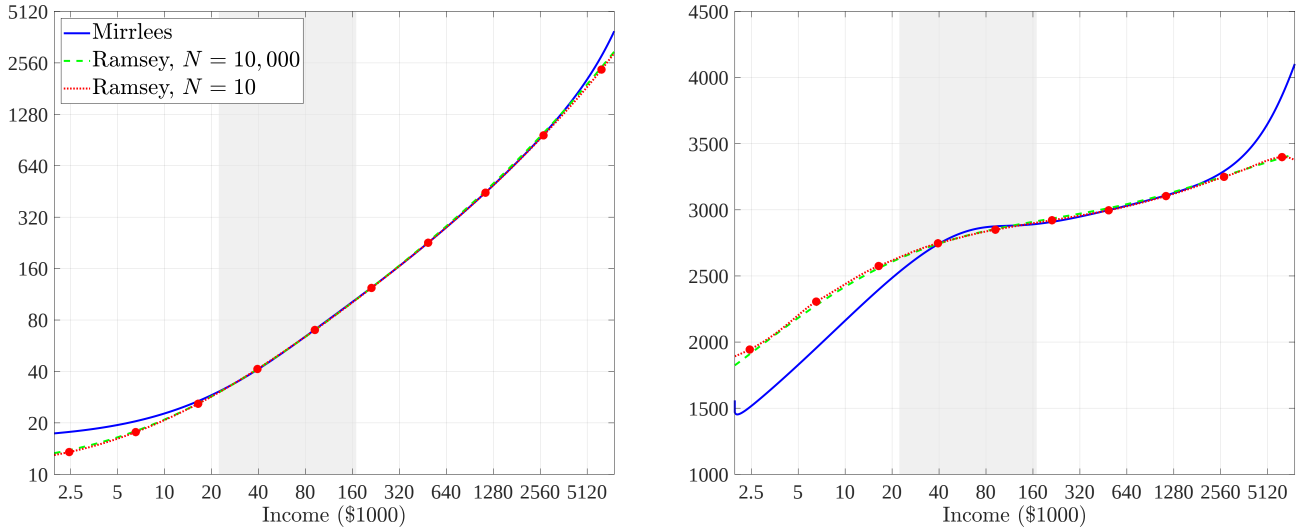
(B) Marginal Tax Rate (%)

Figure A2: Optimal Tax Policy Using Tax Formula Approach. The figure plots the optimal Mirrleesian marginal tax schedule with the number of grid points from 10 to 10,000, using the tax formula approach. Panel A uses the approximation to the DS formula as in Mankiw et al. (2009). Panel B uses a piecewise linear approximation to the marginal tax function in the DS formula.

and Practice," *Journal of Economic Perspectives*, 23, 147–174.

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

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



(A) Consumption ($1,000) **(B)** Hours Worked

Figure A3: Optimal Allocation Under Ramsey Tax Policy. Panels A and B compare the optimal allocation (consumption c in panel A and hours worked y/θ in panel B) under the Ramsey optimal policy with the number of grid points of 10 and 10,000 to those under the Mirrlees optimal policy.