Robust bounds on optimal tax progressivity

Anmol Bhandari[*] Jaroslav Borovička[†] Yuki Yao[‡]

[*] University of Minnesota and NBER

[†] New York University and NBER

[‡] University of Kent

June 18, 2025

Preliminary draft.

Abstract

We study the problem of a robust planner who designs an optimal taxation scheme for a heterogeneous population in presence of uncertainty about the shape of the distribution of underlying types. Low-income workers are well insured under the optimal scheme, and so concerns about the left tail of the type distribution are negligible. On the other hand, the planner fears misspecification of the right tail of the type distribution emerging from budgetary concerns. Even when the tail of the distribution is Pareto, arbitrarily small misspecification concerns lead to zero marginal taxes at the top. A quantitative calibration shows that a plausible degree of uncertainty leads to an optimal tax scheme with substantially reduced marginal tax rates for high-income earners and a peak marginal tax rate much lower than in the model without uncertainty.

We thank Marco Bassetto, Jonathan Heathcote, Nicolas Werquin, Eric Zwick, and seminar participants at Chicago BUMP 2022, the 5th PHBS Workshop in Macroeconomics and Finance 2022, BI-SHoF Conference 2023, Information and Expectations in Macroeconomics Workshop at the St. Louis Fed 2023, Chicago Blue Collar Working Group 2024, T2M in Amsterdam 2024, LAEF on the Lakes in Madison 2024, the Sargent Alumni Reading Group, Midwest Macro 2024 in Richmond, IMF-CARF-TCER-WASEDA Conference 2024, Macroeconomics and Social Insurance Workshop at BSE Summer Forum 2024, SED in Barcelona 2024, NBER SI EFG 2024, MAFE Workshop on Risk Mitigation in Berlin 2024, EEA in Rotterdam 2024, the 3rd XAmsterdam Macroeconomic Workshop, USC, NYU, Minnesota, Kent, UAH and UNC Charlotte for valuable feedback. Bhandari acknowledges support from the Heller—Hurwicz Economic Institute.

1 Introduction

It is well-known from the theory of optimal taxation (Mirrlees (1971)) that the shape of the marginal tax curve crucially depends on the shape of the underlying labor productivities and labor supply preferences.[1] Despite increasingly available surveys and administrative micro-data on income, estimating a joint distribution of skills and preferences that will realize over a particular fiscal planning horizon remains a complex task. The problem is more severe in the tails, where sample sizes are extremely small relative to the time-series variation in counts and incomes of high earners. Moreover, the problem gets even more difficult if we want to jointly estimate productivity and labor supply elasticities for such individuals.

In this paper, we study optimal taxation of labor income when the government acknowledges uncertainty about the underlying distribution of skills and preferences. We find that these concerns generally lead to lower progressivity. More strikingly, we find that the celebrated Diamond (1998) and Saez (2001)'s high top marginal tax rate result is overturned. By introducing a minimal degree of uncertainty to the Diamond–Saez setting, we prove that the top marginal tax approaches zero at a rate that is bounded away from zero and independent of structural parameters.

Methodologically, we build on Hansen and Sargent (2008)'s notion of "robustness" which aims to construct policies that work well not just for a single distribution but across a set of distributions. The concerns about the shape of the type distribution manifest as a max-min game between a government that chooses a nonlinear tax schedule to maximize a given welfare criterion and its alter ego that adversely perturbs the joint distribution of skills and preferences away from a given benchmark distribution subject to a penalty. Motivated by measurement concerns, we use a penalty that scales in a statistical measure of distance between the two distributions, i.e., relative entropy. The max-min problem allows the government to design a tax schedule that is *robust* with respect to a set of distributions that are hard to distinguish from each other using an available finite data sample.

While our work has elements in common with public finance and Hansen–Sargent robustness literature, there are several distinct differences. Similar to the classic Mirrlees problem, we examine a static environment where individuals vary in skills and preferences, and they supply labor given a non-linear tax schedule. The government devises this schedule to maximize a social welfare criterion while adhering to a budget constraint. Our approach differs in the explicit modeling of the uncertainty the government faces regarding the type distribution. Instead of assuming certainty as in the traditional Mirrlees framework, our government uses the max-min formulation discussed previously.[2] If

[1] See, for example, Golosov, Troshkin, and Tsyvinski (2016).

[2] There is a small number of papers that bring statistical concerns to an optimal taxation problem. See Lock-

we set the penalty in the max-min problem to be large enough, our problem converges to the standard Mirrlees problem in which the planner perfectly knows the shape of the type distribution—a feature that gives us a convenient point of departure. We label the problem involving an infinite penalty as the "baseline problem", and we refer to the type distribution in this baseline as the "benchmark" distribution. Similarly, we use terms like "robust problem" and "worst-case distribution" to denote the planner's problem with a finite penalty and the adverse distribution chosen by the minimizing agent, respectively.

Compared to the robustness or more generally the literature on ambiguity[3], we focus on a different type of uncertainty. While the existing literature primarily deals with the distribution of alternative potential outcomes or states of nature a decision-maker might face under uncertainty, our government is concerned about the uncertainty about the shape of the cross-sectional distribution. Households who have private information about their types of skills and preference face no uncertainty. Moreover, our formulation expands upon the typically used one-parameter penalty specification, allowing us to express varying degrees of uncertainty in different segments of the underlying distribution. For example, we can formally articulate when the government is more uncertain about the extreme ends ("tails") of the distribution than the central part.[4]

Our formalization allows us to examine the optimal nonlinear tax schedule using standard mechanism design techniques and use the revelation principle. This enables us to recast the problem of choosing a nonlinear tax schedule equivalently as selecting an optimal allocation, subject to truth-telling constraints. To compare our results with existing literature, we begin with a scenario where the type distribution is one-dimensional (skills), and the penalty is scalar. We then broaden our scope to more complex situations. For this simplified setting, the Diamond–Saez results show that in the baseline problem, the marginal tax rate for a particular skill level is determined by the so-called ABC formula. This formula consists of three components: A) a term dependent on labor supply elasticities, B) a term tied to the hazard rate at a given skill level, and C) a term that relates to the shape of the distribution above that skill level. These three factors strike a balance be-

wood, Sial, and Weinzierl (2021) and Chang and Wu (2023). Our work differs from these studies in modeling statistical concerns over an infinite dimensional object—the entire type distribution—using a non-Bayesian robust control approach, while they model uncertainty over a finite set of parameters that are handled with Bayesian techniques. Those papers found that parametric uncertainty in their environment leads to a more progressive tax system while we find the optimal tax schedule to be less progressive relative to the baseline.

[3]See, for example, Adam and Marcet (2011), Hansen and Sargent (2012), Karantounias (2013), Bhandari, Borovička, and Ho (2024), and Hansen and Miao (2022).

[4]There is a separate but related literature on mechanism design with moral hazard when the planner has limited knowledge of agents' action set. See, for example Carroll (2015). Vairo (2024) applies such a framework to study optimal taxation when agents' action sets consists of how risky their income profiles are. She finds that the optimal tax schedule should be uniformly progressive and in some cases will have higher taxes at the top as compared to a baseline in which the action sets are known. In her setting, a decreasing marginal tax rate encourages socially undesirable risk-taking. Our setting does not feature such moral hazard considerations.

tween the efficiency costs of labor supply distortion caused by the marginal tax imposed at a given skill level, and the benefits derived from redistributing additional income collected from workers above that skill level in a non-distortionary fashion.

We illustrate that in our context, the marginal tax rate adheres to a modified Diamond–Saez ABC formula. The difference lies in the replacement of the hazard rates and distributions with the worst-case distribution, which is determined as part of the max-min problem. This ex-post Bayesian representation helps us analyze the differences in the optimal tax schedule with and without concerns for misspecification of the type distributions by breaking it down into two logical components: 1) what influences the shape of the worst-case distribution, and 2) how the tax schedule is pinned down by the Diamond–Saez ABC formula given the shape of the worst-case distribution. Since the worst-case distribution is an endogenous object that relies on the shape of the tax function, the solution to the robust optimal tax problem is a fixed-point that simultaneously characterizes the optimal tax schedule and the worst-case distribution.

Our first theoretical result is the analysis of the top tax rate in a familiar setting with quasilinear preferences, Rawlsian welfare weights, and a benchmark distribution that has a Pareto tail. In the baseline, when the penalty is infinite and the planner is therefore certain that the benchmark distribution is the correct type distribution, the Diamond–Saez ABC formula shows that the top tax rate approaches a positive finite value that depends on the labor supply elasticity and the Pareto tail parameter. This tax rate is quantitatively quite large—around 70–75% for reasonable elasticity and tail parameters. Intuitively, the thick tail of the productivity distribution means that the government can always collect sufficient revenues from the right of any given threshold productivity to offset the cost of distortions due to an increase in the marginal tax rate at that threshold. For similar reasons, for distributions with bounded support or, more generally, with thin tails, the top tax rate approaches zero.

Now consider the problem faced by a government which is concerned that the underlying benchmark Pareto distribution is misspecified. We first show that the ratio of densities of the worst-case distribution and the benchmark Pareto distribution has an exponential tilting expression familiar from the robustness literature. The worst-case density shifts mass away from worker types who are valuable to the government but the magnitude of this reweighting is disciplined by the entropy penalization of the statistical discrepancy between the benchmark and the worst-case distributions. In our context, an individual's value to the government is determined by two components: first, their utilitarian contribution to the welfare objective function, which depends on the individual's indirect utility under the optimal allocation, and second, their contribution to easing the government budget constraint, which hinges on the net tax revenue the government collects from that individual weighted by the marginal social value of a unit of consumption.

Moving mass away from the right tail of the productivity distribution is costly to the planner and hence desired by the adverse player in the max-min problem because each high-productivity individual generates substantial tax revenue that a redistributive government can use to provide transfers to low-skilled individuals. While tax revenues raised from a particular productivity type rely on the product of productivity level and the mass of agents at that productivity level and decrease as we move further into the right tail, the entropy cost of shifting mass from the right tail diminishes much more rapidly because it scales only with the density. Consequently, in the minimization part of the max-min problem, gains from skimming the density in the right tail grow as we move further into the tail, regardless of the penalty value. We demonstrate that the optimal top tax rate gradually reduces to zero, with the worst-case distribution approaching a distribution with an exponentially decaying density, or a thin tail, in spite of the benchmark distribution being a Pareto that has a polynomial decay in the tail, or, more generally, any other fat-tailed distribution. Thus, the mechanism formalizes the practical concern of policymakers to deal with welfare-relevant aspects of tail behavior that are very hard to detect.

The asymptotic top tax rate is necessarily a statement about limits, and may not be relevant if marginal tax rates approach the zero limit very slowly. To investigate that, we study the elasticity of the marginal tax rate with respect to income. We show that in the limit as productivity becomes large, the worst-case distortion is dominated by planner's concerns about the amount of tax revenue raised from those types. The slower the decay in the marginal tax rate, the higher the revenue collected from the very productive individuals. But higher revenue means these individuals are more valuable to the government, and the adverse agent in the max-min problem has stronger incentives to shift the worst-case density away from them. As their mass becomes smaller, the redistributive gains from taxing high-productivity individuals decline, and the marginal tax rate dictated by the Diamond–Saez formula falls faster. We show that this fixed point is resolved in a unique value such that the elasticity of the marginal tax rate with respect to income equals minus one half. Remarkably, this asymptotic decay rate of the marginal tax rate is independent of any primitive parameters, such as preferences of the household or the government or those characterizing the distribution of skills.

After establishing the main results in the tractable quasilinear-Rawlsian setting, we extend our analysis to more general environments. We study extensions that allow more general utility functions that feature concave preferences over consumption and more general welfare weights. We show that the robust optimal top tax rate is still zero and the limiting rate of convergence is bounded below by negative one half. We then generalize the robust analysis so that the penalty function belongs to a class of power divergences that go beyond just relative entropy. We show that although in some cases the robust optimal top tax rate can be positive, it is always strictly lower than the baseline optimal top tax rate.

To examine the complete tax schedule rather than just the top tax rate, we resort to a numerical solution. The benchmark distribution is calibrated as an exponentially modified Gaussian (EMG) distribution for the logarithm of productivity. Although the top tax rate is not influenced by the specific value of the penalty parameter that controls the degree of misspecification concerns about the shape of the productivity distribution, the overall shape of the tax function is.

In order to quantify the magnitude of model misspecification concerns, we think of a planner who designs a tax schedule and commits to it for some planning horizon, say five years. We use time-series variation in income distributions over such planning horizons to construct a set of alternative distributions that the planner considers as plausible. This set is rich and contains all distributions in the smallest "entropy ball" that encloses the observed distributions in a given 5 year interval. The radius of this set maps to the penalty parameter in our max-min problem and the center maps to the benchmark distribution. By varying the timing and size of the window, we study the sensitivity of the optimal tax schedule to the degree of misspecification concerns.

We evaluate our findings against the baseline scenario, in which the penalty parameter is infinite. We observe significant effects of the presence of misspecification concerns on the marginal tax rate. For instance, households earning more than \$500,000 face a baseline tax rate exceeding 70% in the absence of misspecification concerns, but under our preferred calibration, the optimal tax rate peaks at 57.5% for earnings around \$375,000. This rate falls to 50% for households with incomes of \$1 million, ultimately diminishing to zero, as suggested by our theory. These lower tax rates result in an approximately 2% increase in output but also lead to slightly more than 8% reduction in transfers to the lowest-income households.

Finally, we consider a case in which the government is uncertain not just about the distribution of labor productivity but the joint distribution of labor productivity and labor supply elasticities. The robust planner entertains a family of alternative joint densities in which the conditional distribution of the elasticity of taxable income may vary with producivity. In the worst-case, the elasticity of taxable income is positively correlated with income exacerbating a familiar trade-off: revenues are disproportionately collected from the top of the distribution, yet these same taxpayers may generate the largest deadweight losses if their labor supply proves highly elastic. Anticipating this risk, the planner lowers marginal tax rates on high earners to hedge against the possibility of a sharp revenue shortfall. Calibrating the degree of misspecification to match recent estimates for taxpayers earning roughly five million dollars a year, we find that the optimal top marginal rate falls from 66.5 percent in the rational benchmark to 14.0 percent—a reduction of roughly 52.5 percentage points. Importantly, the income distribution itself remains virtually unchanged; what drives the result is not tail thickness but the feared interaction between tail income and tail elasticity. The mechanism thus parallels our earlier one-dimensional analysis: whenever a large share of public resources is expected to come from a small set of taxpayers, even modest doubts about their behavioral response can rationalize substantially lower optimal taxes at the top.

The rest of the paper is structured as follows. Section 2 describes the simple model with one-dimensional types and scalar penalty. Section 3 contains our main theoretical results about the top marginal tax rate and the rate of convergence. Section 4 uses a calibrated economy to study the full tax schedule. Section 5 extends the environment to multi-dimensional types. Section 7 concludes.

2 Model

The economy is populated by a continuum of workers indexed by their productivity z distributed according to density $f(z)$ with a continuous support $[\underline{z}, \bar{z}] \subseteq \mathbb{R}_+$ where $\bar{z}$ may be infinite. Productivity types are private information of the worker. A worker with productivity z supplying labor n produces income $y = zn$. The worker solves the utility-maximization problem

$$\max_{c,n} U(c, n) \quad \text{s.t. } c = zn - T(zn)$$

where $U(c, n)$ is a strictly concave and differentiable utility function representing worker's preferences over consumption and hours worked, and $T(y)$ is the tax levied on income y. Taking the tax function as given, worker's optimal choice of labor supply yields the condition

$$U_c(c, n)\left(1 - T'(zn)\right) z + U_n(c, n) = 0. \tag{1}$$

Denote the optimal choice of consumption and labor $\mathcal{C}(z; T)$ and $\mathcal{N}(z; T)$, respectively, the resulting output $\mathcal{Y}(z; T) = z\mathcal{N}(z; T)$, and the associated indirect utility function $\mathcal{U}(z; T)$. For the rest of the paper, we drop T as an explicit argument of functions $\mathcal{C}(\cdot)$, $\mathcal{N}(\cdot)$, $\mathcal{Y}(\cdot)$, and $\mathcal{U}(\cdot)$.

The government is in charge of choosing the tax schedule T as a function of observed income y. Taxes are levied for redistribution purposes and to pay for government expenditure G. The government welfare objective is given by

$$\mathbb{E}[\psi\mathcal{U}] + V(G) = \int_{\underline{z}}^{\bar{z}} \psi(z)\,\mathcal{U}(z)\, f(z)\, dz + V(G)$$

where $\psi(z)$ is a Negishi (1960) welfare weighting function that satisfies $\mathbb{E}[\psi] = 1$, and $V(\cdot)$ is a concave differentiable function.[5] For example, $\psi(z) = 1$ implies a utilitarian planner, while $\psi(z) = \delta_{\underline{z}}(z) / f(z)$ where $\delta_{\underline{z}}(z)$ is the Dirac delta function yields the Rawlsian welfare criterion. In the absence of model misspecification concerns, the government chooses the tax schedule so as to maximize the welfare objective subject to the budget constraint

$$G = \mathbb{E}[T(\mathcal{Y})].$$

We study the optimal taxation problem in a situation when the government is concerned that the underlying distribution of productivity types $f(z)$ is misspecified. In the spirit of Hansen and Sargent (2001a,b), the government contemplates a set of alternative type distributions $\widetilde{f}(z)$ that are statistically close to the 'benchmark' distribution $f(z)$. We denote $m(z) = \widetilde{f}(z) / f(z)$ the likelihood ratio between the benchmark and the alternative distribution, and $\widetilde{\mathbb{E}}[\cdot]$ the expectation operator under the distribution $\widetilde{f}(z)$. By construction, $\mathbb{E}[m] = 1$. For any integrable function $X(z)$, the Radon–Nikodým theorem implies

$$\widetilde{\mathbb{E}}[X] = \int_{\underline{z}}^{\bar{z}} X(z)\widetilde{f}(z)\,dz = \int_{\underline{z}}^{\bar{z}} X(z)\,m(z)\,f(z)\,dz = \mathbb{E}[mX]. \tag{2}$$

The degree of statistical distinguishability of the two distributions $f(z)$ and $\widetilde{f}(z)$ is represented by their relative entropy

$$\mathcal{E}\left(f,\widetilde{f}\right) = \mathbb{E}[m\log m] = \int_{\underline{z}}^{\bar{z}} m(z)\log m(z)\,f(z)\,dz.$$

The relative entropy is nonnegative, and is equal to zero if and only if $m \equiv 1$ with probability one. Alternative distributions $\widetilde{f}(z)$ that are statistically easier to distinguish from $f(z)$ yield a larger relative entropy.

The government desires to choose a tax function that would perform well across the set of alternative type distributions $\widetilde{f}(z)$ that are statistically not too distinct from the benchmark distribution $f(z)$. We restrict this set by an entropy bound κ:

$$\mathcal{F}(f,\kappa) = \left\{\widetilde{f} : \mathcal{E}\left(f,\widetilde{f}\right) \leq \kappa\right\}.$$

A larger value of κ represents stronger model misspecification concerns. This leads to the maxmin problem for the government:

$$\max_{T} \min_{\widetilde{f}\in\mathcal{F}(f,\kappa)} \widetilde{\mathbb{E}}[\psi\mathcal{U}] + V\left(\widetilde{\mathbb{E}}[T(\mathcal{Y})]\right). \tag{3}$$

[5] The results carry over to the case when G is an exogenous amount of government expenditures that the government must raise.

The first term in the objective function is equal to the government welfare $\widetilde{\mathbb{E}}\left[\psi\mathcal{U}\right]$ evaluated under the alternative distribution $\widetilde{f}(z)$. The second term are the total net tax revenues, again evaluated under the alternative distribution, that are raised to pay for government expenditures $G = \widetilde{\mathbb{E}}\left[T\left(\mathcal{Y}\right)\right]$.

The government problem (3) leads to an optimal tax function that is robust to misspecifications of the type distribution that adversely affect the government objective. The problem can be interpreted as a two-player game in which the government faces a malevolent nature that chooses alternative distributions with the most adverse welfare consequences for the contemplated tax function. In line with the literature, we call the distribution $\widetilde{f}(z)$ that solves the minimax problem in (3) the worst-case distribution.

It will be more convenient in our analysis to work with an equivalent penalized version of the government problem:

$$\max_{T} \min_{\substack{m>0 \\ \mathbb{E}[m]=1}} \mathbb{E}\left[m\psi\mathcal{U}\right] + V\left(\mathbb{E}\left[mT\left(\mathcal{Y}\right)\right]\right) + \theta\mathbb{E}\left[m\log m\right], \tag{4}$$

where we represent the alternative distributions $\widetilde{f}$ using their likelihood ratio m as in (2). Since $\widetilde{f}(z) = m(z) f(z)$, the likelihood ratio $m(z)$ plays the role of a weighting function that over- or underweighs the alternative distribution relative to the benchmark.

The first two terms of the objective are the same as in (3). The last term is an entropy penalty that penalizes distributions with a large statistical distance from the benchmark distribution. The degree of penalization is controlled by the parameter θ. This parameter can be interpreted as the Lagrange multiplier on the entropy constraint $\mathcal{E}\left(f, \widetilde{f}\right) \leq \kappa$ for a suitable value of κ. A larger value of θ implies a tighter entropy constraint with a smaller κ, leading to a worst-case distribution that is statistically closer to the benchmark. As $\theta \to \infty$, the entropy penalty becomes prohibitive, model misspecification concerns vanish, and we obtain $\widetilde{f} \equiv f$, equivalent to $\kappa = 0$.

The government problem implies that the malevolent nature exploits both the direct welfare impact as well as the budgetary consequences of adversely chosen distributions. On the one hand, it desires to impose a high $m(z)$ for types with low welfare impact $\psi(z)\,\mathcal{U}(z)$ to lower $\mathbb{E}\left[m\psi\mathcal{U}\right]$. On the other hand, it strives for adverse budgetary consequences by underweighing types for whom $T\left(\mathcal{Y}(z)\right)$ is positive (net tax payers) while, vice versa, overweighing those for whom $T\left(\mathcal{Y}(z)\right)$ is negative (net tax recipients), and thus lowering the net revenue $G = \mathbb{E}\left[mT\left(\mathcal{Y}\right)\right]$. At the same time, alternative adverse distributions $\widetilde{f}(z)$ chosen by nature cannot be too distinct from the benchmark so as not to incur a large penalty $\theta\mathbb{E}\left[m\log m\right]$.

2.1 Mirrleesian formulation

Rather than solving for the optimal tax function (4), we follow Mirrlees (1971), and characterize the constrained optimal allocation under the restriction that the tax function can only depend on worker's income. As usual, we cast the problem as a mechanism design problem, and rely on Myerson (1979), focusing on incentive-compatible mechanisms in which workers truthfully reveal their types. The social welfare function that the mechanism implements is given by the inner minimization problem in (4).

The government offers to workers a menu of allocations $(c(z), y(z))$ indexed by z. Worker of type z chooses a reporting strategy $\hat{z}$ that entitles to consumption $c(\hat{z})$ in exchange for providing output $y(\hat{z})$ that requires labor input $y(\hat{z})/z$. The reporting strategy therefore solves the announcement problem

$$\max_{\hat{z}} U\left(c(\hat{z}), \frac{y(\hat{z})}{z}\right).$$

Incentive-compatibility requires that the optimal report satisfies $\hat{z} = z$. The first-order necessary condition evaluated at $\hat{z} = z$ yields

$$U_c\left(c(z), \frac{y(z)}{z}\right) c'(z) + U_n\left(c(z), \frac{y(z)}{z}\right) \frac{y'(z)}{z} = 0. \tag{5}$$

Totally differentiating the utility function with respect to z at the allocations $(c(z), y(z))$ and plugging in the optimal reporting strategy condition derived in (5), we obtain

$$\frac{dU}{dz} = -U_n\left(c(z), \frac{y(z)}{z}\right) \frac{y(z)}{z^2}. \tag{6}$$

This is a condition on the utility gradient the menu $(c(z), y(z))$ has to satisfy to be locally incentive-compatible (IC). When this condition holds, the worker has no incentives to misrepresent the true type by an infinitesimal deviation.

The key obstacle is the complicated structure of the social welfare function involving the minimization problem over alternative distributions. However, we can apply the minimax theorem to exchange the order of optimization in (4).[6] The planner is thus solving

$$\min_{\substack{m>0 \\ \mathbb{E}[m]=1}} \max_{c,y} \int_{\underline{z}}^{\bar{z}} \psi(z) U\left(c(z), \frac{y(z)}{z}\right) m(z) f(z)\, dz + V(G) + \theta \int_{\underline{z}}^{\bar{z}} m(z) \log m(z) f(z)\, dz \tag{7}$$

[6]We provide a formal verification of the applicability of the minimax theorem for a special case in Appendix A. We verify more complicated cases using numerical algorithms.

subject to the IC constraint (6) and the budget constraint

$$G = \int_{\underline{z}}^{\bar{z}} \left(y\left(z\right) - c\left(z\right)\right) m\left(z\right) f\left(z\right) dz. \tag{8}$$

Given a fixed function $m(z)$, the inner maximization problem is now a standard constrained allocation problem. Assuming that the function

$$q\left(c,n\right) = -n\frac{U_n\left(c,n\right)}{U_c\left(c,n\right)} \tag{9}$$

is strictly increasing in n for each fixed c implies a single-crossing property under which the local IC constraint (6) also implies global incentive compatibility, and allocations $\left(c\left(z\right), y\left(z\right)\right)$ that satisfy incentive compatibility are also strictly increasing in z.

Since the IC constraint is type-by-type and does not depend on the underlying distribution, the model misspecification concern on the side of the planner does not alter its form. We can therefore rely on the convenient Hamiltonian formulation that characterizes the allocation given by the inner maximization problem in (7), and yields a modification of the Diamond (1998) and Saez (2001) elasticity formula for the marginal tax rate. Treating $\mathcal{U}$ as the state variable, λ as its co-state, and y and m as control variables, we form the constrained Hamiltonian

$$\begin{aligned} H\left(\mathcal{U},y,m,\lambda\right) &= \psi\left(z\right)\mathcal{U}\left(z\right)m\left(z\right)f\left(z\right) + \theta m\left(z\right)\log m\left(z\right)f\left(z\right) - \chi m\left(z\right)f\left(z\right) \\ &\quad -\lambda\left(z\right)U_n\left(c\left(z\right),\frac{y\left(z\right)}{z}\right)\frac{y\left(z\right)}{z^2} + \mu\left[y\left(z\right) - c\left(z\right)\right]m\left(z\right)f\left(z\right). \end{aligned} \tag{10}$$

Here, χ and μ are multipliers on the constraints $\mathbb{E}\left[m\right] = 1$ and (8), respectively, and $c(z)$ is defined implicitly from the definition of the utility function as $c\left(z\right) = C\left(\mathcal{U}\left(z\right), y\left(z\right)\right)$.

We derive a general characterization of the problem in Appendix B. Here we provide the analysis of a special case with quasilinear preferences and isoelastic labor disutility.

Assumption 1. *Workers' preferences are given by*

$$U\left(c,n\right) = c - \frac{n^{1+\gamma}}{1+\gamma}, \tag{11}$$

and the function $V(G)$ satisfies, for some $\underline{G} \geq -\infty$,

$$\lim_{G \searrow \underline{G}} V'\left(G\right) = \infty \qquad \lim_{G \nearrow \infty} V'\left(G\right) = 0. \tag{12}$$

The Inada conditions (12) guarantee an interior solution to the government problem.

The first-order condition with respect to $m(z)$ in (10) together with the restriction

$\mathbb{E}[m] = 1$ yield a characterization of the worst-case distortion in the form of an exponential tilting formula

$$m(z) = \bar{m} \exp\left(-\frac{1}{\theta} \left[\psi(z) \mathcal{U}(z) + \mu T(y(z)) \right] \right), \tag{13}$$

where $\bar{m}$ is a normalization constant that assures $\mathbb{E}[m] = 1$, and $T(y(z)) = y(z) - c(z)$ represents the effective tax the allocation imposes on worker of type z. The remaining optimality conditions then imply the formula for the marginal tax

$$\frac{T'(y(z))}{1 - T'(y(z))} = (1 + \gamma) \frac{\widetilde{\Psi}(z) - \widetilde{F}(z)}{1 - \widetilde{F}(z)} \frac{1 - \widetilde{F}(z)}{z \widetilde{f}(z)}, \tag{14}$$

where $\widetilde{F}(z)$ is the cumulative distribution function of the worst-case density, and $\widetilde{\Psi}(z)$ is planner's cumulative welfare weight

$$\begin{aligned} \widetilde{F}(z) &= \int_{\underline{z}}^{z} \widetilde{f}(\zeta)\, d\zeta = \int_{\underline{z}}^{z} m(\zeta) f(\zeta)\, d\zeta \\ \widetilde{\Psi}(z) &= \int_{\underline{z}}^{z} \frac{\psi(\zeta) \widetilde{f}(\zeta)}{\int_{\underline{z}}^{\bar{z}} \psi(\xi) \widetilde{f}(\xi)\, d\xi} d\zeta. \end{aligned}$$

The lump sum portion of the tax $T(y(\underline{z}))$ imposed on the least productive worker is then determined so as the whole tax scheme equalizes the marginal cost of public funds μ, i.e., the utility cost of raising an extra unit of tax revenue, to the marginal value of government expenditure

$$\mu = V'(G).$$

The optimal marginal tax formula is analogous to that of Diamond (1998) and Saez (2001), except that now, it depends on the endogenously determined distribution $\widetilde{f}(z)$ represented by the distortion $m(z)$ in (13). The first term on the right-hand side of (14) captures distortionary effects of taxation on the labor supply, indicating that marginal taxes should be lower when the inverse of the labor supply elasticity γ is low. The second term represents the desire for redistribution, and is bounded above by one. Marginal taxes will be strictly positive when $\widetilde{\Psi}(z) > \widetilde{F}(z)$, indicating a planner that puts higher welfare weights on lower worker types in the first-order stochastic dominance sense. Finally, the third term is determined by the shape of the tail of the type distribution, and it represents the tradeoff that an increase in the marginal tax $T'(y(z))$ causes at a particular z. This marginal tax has an adverse distortionary effect on the labor supply of all workers with type exactly at z, leading to a total output loss $z\widetilde{f}(z)$, while generating the benefit of raising extra revenue in lump sum fashion from all workers with type above z, whose mass is $1 - \widetilde{F}(z)$.

The form of the distortion (13) reveals that the model misspecification concerns of the robust planner have a redistributive and a budgetary component. The numerator of the expression for $m(z) = \widetilde{f}(z)/f(z)$ in (13) indicates that the robust planner underweighs worker types who, under the optimal tax policy, receive allocations with high weighted utility $\psi(z)\mathcal{U}(z)$ or those with high net contributions to the planner's budget, $\mu T(y(z))$. The Lagrange multiplier μ converts the tax revenue to utility units under the government welfare function.

The planner uses the tax policy to maximize the social welfare function. Since insurance is not perfect, the worst-case distribution that puts more weight on types with a low $\psi(z)\mathcal{U}(z)$ and less weight on types with a high $\psi(z)\mathcal{U}(z)$ reflects the concern that the chosen tax function achieves lower welfare $\widetilde{\mathbb{E}}[\psi\mathcal{U}]$ than that measured under the benchmark model, $\mathbb{E}[\psi\mathcal{U}]$.

At the same time, the government needs tax revenue to achieve the desired redistribution and spending. Underweighing worker types who deliver high tax revenue $\mu T(y(z))$ and overweighing those who deliver low tax revenue $\mu T(y(z))$ reflects concerns that worker types who contribute substantially to the budget are less abundant than under the benchmark model, making it more challenging to achieve the desired goals.

The parameter θ controls the entropy penalty in the planner's problem (7) and hence the degree of model misspecification concerns. A small value of θ reflects more substantial concerns, which leads to stronger exponential tilting in (13). As $\theta \to \infty$, model misspecification concerns vanish, and the worst-case density $\widetilde{f}(z)$ approaches the benchmark model density $f(z)$ in the statistical sense expressed by the entropy penalty $\mathbb{E}[m \log m]$.

Importantly, the worst-case distortion $m(z)$ in (13) and the tax function $T(y(z))$ in (14) are determined jointly as an outcome of the minimax problem (7). Given the tax function, the worst-case density delivers the lowest penalized objective in (7), and vice versa, taking the worst-case density as given, the tax function maximizes planner's welfare. The solution is a saddle point in the objective function that constitutes an equilibrium in a two-person game between the benevolent government and the malevolent nature.

3 Optimal marginal tax rates at the top

In this section, we provide an analytical characterization of the asymptotic behavior of the tax rate in (14) in the presence of model misspefication concerns.

When the planner is utilitarian with $\psi(z) \equiv 1$, then, due to the quasilinear form of preferences in (11), the motive for redistribution is absent. Equivalent to the case without misspecification concerns, we obtain that marginal taxes are zero, $T'(y(z)) = 0$. Redistributive concerns are therefore induced by a decreasing welfare weighting function $\psi(z)$.

Here we focus on the case in which there exists a $\hat{z}$ such that $\psi(z) = 0$ for all $z \geq \hat{z}$. The planner hence puts a zero welfare weight on the right tail of the worker distribution. In this case, $\widetilde{\Psi}(z) = 1$ for $z \geq \hat{z}$, and the tax formula becomes

$$\frac{T'(y(z))}{1 - T'(y(z))} = (1+\gamma)\frac{1 - \widetilde{F}(z)}{z\widetilde{f}(z)} \tag{15}$$

Since the second term in the tax formula (14) is bounded above by one, the current case yields the highest possible tax rate in the tail of the type distribution across all alternative welfare functions. The worst-case distortion for $z \geq \hat{z}$ then becomes

$$m(z) = \bar{m}\exp\left(-\frac{\mu}{\theta}T(y(z))\right), \tag{16}$$

corresponding to a special case of formula (13). The misspecification concerns in the right tail of the distribution therefore do not involve the welfare of high-type agents, and only reflect the concerns about the budgetary consequences of not having sufficiently many high-type workers who contribute to the budget.

In Section 3.4, we provide the analysis of the more general case that relaxes the assumption of quasilinear preferences in (11) and also treats more general welfare weighting functions so that the welfare concern term $\psi(z)\mathcal{U}(z)$ in (13) is not zero in the right tail. It turns out that in most cases, the budgetary concern $\mu T(y(z))$ dominates, and when it does not, the additional welfare concern further reinforces our results.

3.1 Zero marginal taxes at the top

In the absence of model misspecification concerns, $m(z) \equiv 1$, which implies that the worst-case distribution in (15) corresponds to the exogenously specified benchmark model. The limiting tax rate then depends on the shape of the benchmark distribution. When the distribution $f(z)$ is sufficiently thick-tailed, then the marginal tax rate determined (15) has a strictly positive limit. For example, in the case of the Pareto distribution with shape parameter α, we have $(1 - F(z)) / (zf(z)) = \alpha^{-1}$. On the other hand, bounded or thin-tailed distributions, such as normal or lognormal, imply a zero limit. This has lead to widely differing policy prescriptions about the range of recommended marginal tax rates the planner should impose on top incomes.[7]

When model misspecification concerns are present, we characterize the shape of the

[7] For example, Diamond and Saez (2011) find a mid-range estimate for the top marginal tax of 73%, based on labor supply elasticity $\gamma^{-1} = 0.25$ and a Pareto distribution of types z with shape parameter $\alpha = 1.875$ (under the given elasticity of labor supply, this translates to a Pareto distribution of incomes $y(z)$ with shape parameter $\alpha_y = 1.5$).

worst-case density

$$\widetilde{f}(z) = m(z) f(z) = \bar{m} \exp\left(-\frac{\mu}{\theta} T(y(z))\right) f(z). \tag{17}$$

The distribution $\widetilde{f}(z)$ remains continuous, which implies that $\lim_{z\to\bar{z}} \widetilde{F}(z) = 1$. When $\bar{z}$ is finite, the conclusion about the top marginal tax rate is the same as without model misspecification concerns, and $\lim_{z\to\bar{z}} T'(y(z)) = 0$. To see this, notice that in order for the top marginal tax rate in (15) to be different from zero, we need $\lim_{z\to\bar{z}} z\widetilde{f}(z) = 0$. In this case, we can apply L'Hôpital's rule to obtain

$$\lim_{z\to\bar{z}} \frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)} = \lim_{z\to\bar{z}} -\frac{1}{1+z\frac{d\log\widetilde{f}(z)}{dz}} = 0,$$

where the conclusion follows from the fact that $\lim_{z\to\bar{z}} z\widetilde{f}(z) = 0$ for a finite $\bar{z}$ implies $\lim_{z\to\bar{z}} \log\widetilde{f}(z) = \lim_{z\to\bar{z}} \frac{d}{dz}\log\widetilde{f}(z) = -\infty$. This is a contradiction and the limiting marginal tax at the top must be zero.

We therefore focus on the more interesting case when $\bar{z} = \infty$. It turns out that the marginal tax rate at the top still asymptotically converges to zero.

Assumption 2. *There exists a $\hat{z}$ such that the type density $f(z)$ under the benchmark distribution is continuously differentiable on $[\hat{z},\bar{z})$, and $zf(z)$ is strictly decreasing on $[\hat{z},\bar{z})$, with*

$$\lim_{z\to\bar{z}} \frac{d\log f(z)}{d\log z} < -1,$$

with the limit possibly being $-\infty$.

Theorem 3.1. *Assume that preferences satisfy Assumption 1, the type distribution satisfies Assumption 2 with $\bar{z} = \infty$, and $\theta < \infty$. Then the marginal tax rate vanishes to zero at the top:*

$$\lim_{z\to\infty} T'(y(z)) = 0.$$

We formally prove the theorem in Appendix C.1. The proof requires a technical treatment of the existence of the limit but conditional on its existence, the result is intuitive. Denote $T'_{rat}(y_{rat}(z))$ the optimal marginal tax rate in the model without model misspecification concerns, $\theta = \infty$. Then

$$\lim_{z\to\infty} \frac{T'_{rat}(y_{rat}(z))}{1-T'_{rat}(y_{rat}(z))} = (1+\gamma)\lim_{z\to\infty} \frac{1-F(z)}{zf(z)} = (1+\gamma)\lim_{z\to\infty} \frac{1}{-\frac{d\log f(z)}{d\log z} - 1} < \infty,$$

where the second equality follows from an application of L'Hôpital's rule, and the final inequality is implied by Assumption 2. This yields $\lim_{z\to\infty} T'_{rat}(y_{rat}(z)) < 1$.

Further, the single-crossing property (9) implies that the optimal incentive compatible scheme yields output $y(z)$ that is strictly increasing in worker's type z, and since the marginal tax is strictly positive, $m(z)$ in (17) is strictly decreasing. This implies that, for any $z \geq \hat{z}$,

$$\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)} < \frac{1-F(z)}{zf(z)}$$

and hence also $\lim_{z\to\infty} T'(y(z)) \leq \lim_{z\to\infty} T'_{rat}(y_{rat}(z))$.

The limiting tax rate under model misspecification cannot, however, be positive. If it were converging to $\tau_\infty > 0$, then the output function $y(z)$ implied by the optimal labor choice (5) for the case of quasilinear preferences (11) would asymptotically behave as

$$y(z) = \left(1 - T'(y(z))\right)^{\frac{1}{\gamma}} z^{\frac{1+\gamma}{\gamma}} \approx (1-\tau_\infty)^{\frac{1}{\gamma}} z^{\frac{1+\gamma}{\gamma}}, \tag{18}$$

and the worst-case distortion as

$$m(z) = \bar{m} \exp\left(-\frac{\mu}{\theta} T(y(z))\right) \approx \bar{m} \exp\left(-\frac{\mu}{\theta}\tau_\infty (1-\tau_\infty)^{\frac{1}{\gamma}} z^{\frac{1+\gamma}{\gamma}}\right). \tag{19}$$

In that case, an application of L'Hôpital's rule to the tax formula (15) yields

$$\lim_{z\to\infty} \frac{T'(y(z))}{1-T'(y(z))} = \lim_{z\to\infty} (1+\gamma) \frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)} = \lim_{z\to\infty} \frac{1+\gamma}{\frac{\mu}{\theta} z \frac{d}{dz} T(y(z)) - \frac{d\log f(z)}{d\log z} - 1} = 0, \tag{20}$$

because the first term in the denominator diverges to ∞. This is a contradiction to the assumption that $\lim_{z\to\infty} T'(y(z)) = \tau_\infty > 0$, and hence the tax rate has to converge to zero.

The striking result is that the marginal tax rate at the top converges to zero irrespective of the degree of model misspecification concerns. Since we abstracted from welfare concerns at the top of the income distribution, the planner only cares about the budgetary consequences associated with taxing top incomes. From this perspective, the robust planner is concerned that there are fewer high-productivity workers that can be taxed.

The form of the distortion in (16) indicates that the concerns grow proportionally with the marginal social value of the tax revenue $\mu T(y(z))$ that the worker with a given productivity z contributes to the budget. The distortion $m(z)$ therefore becomes more severe as z increases, effectively generating a thinner tail in the worst-case distribution $\widetilde{f}(z)$ in (17). This consequently implies a lower and vanishing optimal marginal tax as $z \to \infty$, since the tradeoff of an increase in the marginal tax $T'(y(z))$ at z that compares the extra benefit of taxing workers above z with the cost of distorting labor supply of workers at z becomes less favorable under the worst-case distribution.

While we have shown the vanishing marginal tax rate result for the case of quasilinear utility and no planner's welfare concerns in the right tail of the type distribution, these results carry over to more general cases. We show in Section 3.4 that the result also holds when utility from consumption is concave, and in the presence of welfare concerns at the top of the type distribution when $\Psi(z)$ only asymptotically converges to one.

The intuition for these generalizations is straightforward, as they yield the more general form of the distortion of the type distribution represented by expression (13). The case of concave utility leads to a different marginal social value of public funds μ and to a different optimal tax function trajectory $T(y(z))$ but since μ remains strictly positive, we only need to show that the tax revenue $T(y(z))$ under the optimal tax continues to diverge to ∞ as $z \to \infty$.

Adding welfare concerns corresponds to a nonzero $\psi(z)$ function as $z \to \infty$. Since $\mathcal{U}(z)$ is increasing in z, this can only lead to a more strongly decreasing distortion $m(z)$, as the budgetary and welfare concerns of not having sufficiently many workers with high types who contribute substantially both to the budget as well as to the welfare objective reinforce each other.

3.2 Rate of decay of the tax rate

Theorem 3.1 shows that the marginal tax rates at the top vanish to zero irrespective of the underlying type distribution. However, the theorem does not determine the rate of convergence. In quantitative applications, the rate of convergence matters because it sharpens information about the practical importance of the asymptotic behavior for finite values of productivity z.

In this subsection, we derive this rate of convergence. We first state it in the form of a theorem, and then provide a derivation of the result that leads to a specific differential equation that will be helpful in numerical implementation of the full characterization of the optimal tax scheme.

In order to simplify the formula, we strengthen Assumption 2 and assume that the tail of the benchmark density $f(z)$ is given by a Pareto distribution with shape parameter α. This is the prototypical choice that yields strictly positive asymptotic marginal taxes in absence of model misspecification. We characterize the following result directly in income space, treating $y = y(z)$ as the endogenous income of each worker. Previous results imply that $\lim_{z\to\infty} y(z) = \infty$, so that limits $z \to \infty$ and $y \to \infty$ are equivalent.

Theorem 3.2. *Assume that preferences satisfy Assumption 1, underlying productivity has a right tail that is Pareto distributed with shape parameter α, and $\theta < \infty$. Then the limiting tax rate* *satisfies*

$$\lim_{y\to\infty} \frac{\mu}{\theta} \left[T'(y)\right]^2 y = \gamma. \tag{21}$$

In particular, the limiting elasticity of the marginal tax with respect to income, if the limit exists, is equal to

$$\lim_{y\to\infty} \frac{d\log T'(y)}{d\log y} = -\frac{1}{2}. \tag{22}$$

The proof of the theorem is provided in Appendix C.3. Expression (21) depends on one endogenous object, the marginal social value of wealth μ, which needs to be determined separately. The expression implies that the marginal tax rate has to decay to zero at rate

$$\log T'(y) \approx -\frac{1}{2}\log y.$$

As a result, if the elasticity of the marginal tax rate with respect to income has a limit, this limit has to be equal to $-\frac{1}{2}$.

In Section 3.4 and Appendix D, we again show how this result generalizes when we relax the assumptions of the theorem. Broadly speaking, the elasticity expression (22) remains robust but the rate of decay of the marginal tax rate to zero may be even faster when, for example, planner's welfare concerns for the top income earners are sufficiently strong, or when the type distribution under the benchmark density is already sufficiently thin-tailed to begin with.

Remarkably, the elasticity of the marginal tax rate with respect to income (22) does not depend on any of the parameters of the model. In order to understand the result, assume that for incomes $y \geq \hat{y}$, the marginal tax rate $T'(y)$ is sufficiently well approximated by a constant elasticity function with elasticity ν. Taking a $\bar{y} \gg \hat{y}$, we can write

$$T(\bar{y}) = T(\hat{y}) + \int_{\hat{y}}^{\bar{y}} T'(y)\,dy \approx T(\hat{y}) + \int_{\hat{y}}^{\bar{y}} (1+\nu)\,y^{\nu} dy = T(\hat{y}) - \hat{y}^{1+\nu} + \bar{y}^{1+\nu}.$$

This means that asymptotically, for high-productivity types $\bar{z}$ with income $\bar{y} = y(\bar{z})$, the worst-case distortion behaves as

$$m(\bar{z}) = \bar{m}\exp\left(-\frac{\mu}{\theta}T(\bar{y})\right) \approx \widetilde{m}\exp\left(-\frac{\mu}{\theta}\bar{y}^{1+\nu}\right) \tag{23}$$

where $\widetilde{m}$ absorbs the contribution of terms $T(\hat{y}) - \hat{y}^{1+\gamma}$. At the same time, the output function $y(z)$ in (18) implies that as $T'(y) \to 0$, we can approximate $y(z) \approx z^{\frac{1+\gamma}{\gamma}}$. Applying L'Hôpital's rule as in (20), we obtain

$$\lim_{z\to\infty} \frac{T'(y(z))}{1-T'(y(z))} = (1+\gamma)\lim_{z\to\infty} \frac{1}{\frac{\mu}{\theta} z T'(y(z))\,y'(z) - \frac{d\log f(z)}{d\log z} - 1} \tag{24}$$

The denominator on the right-hand side is then dominated by the the first term,

$$\frac{\mu}{\theta} z T'(y(z))\, y'(z) \approx \frac{\mu}{\theta}(1+\nu)\frac{1+\gamma}{\gamma} z^{\frac{1+\gamma}{\gamma}(\nu+1)}.$$

Comparing the elasticities of the left-hand and right-hand side of (24) with respect to the type z yields

$$\frac{1+\gamma}{\gamma}\nu = -\frac{1+\gamma}{\gamma}(\nu+1),$$

from which we obtain $\nu = -\frac{1}{2}$.

Intuitively, expression (24) indicates that the optimal rate of decay of the tax rate balances two forces, the effect on the tax revenue $T(y(z))$ collected from high worker types z, and the effect this tax revenue has on the worst-case distortion $m(z)$. If the decay rate was higher (a more negative ν), then the tax revenue $T(y(z))$ would grow more slowly as $z \to \infty$. This would consequently diminish the budgetary concerns of the model misspecification in (23), the worst-case density $\widetilde{f}(z)$ would be less distorted with a thicker tail, and the optimal tax formula would indicate a more gradual decay of the tax rate. The chain of arguments is reversed if the decay rate was lower (a less negative ν).

The elasticity choice $\nu = -\frac{1}{2}$ thus uniquely identifies the asymptote of the saddle point between the maximization problem that seeks the optimal tax rate, and the minimization problem that finds the model misspecification with the most adverse consequences for the planner.

3.3 Phase diagram

We formalize the proof of Theorem 3.2 by deriving a differential equation for the marginal tax rate $T'(y)$. After differentiating the tax formula (15) with respect to the productivity z and a sequence of algebraic manipulations, we obtain the differential equation

$$-\frac{T''(y)\, y}{1-T'(y)} = -\left[2 - \frac{1+\gamma+\alpha}{1+\gamma} T'(y)\right]^{-1} \left[\frac{\mu}{\theta}\left[T'(y)\right]^2 y - \gamma + \gamma \frac{1+\gamma+\alpha}{1+\gamma} T'(y)\right]. \quad (25)$$

This equation depends on z only implicitly through $y(z)$, and we therefore can treat the equation directly as a function of income y. This is a first-order differential equation for the unknown marginal tax $T'(y)$, and the unique strictly positive solution is pinned down by the terminal condition $\lim_{y\to\infty} T'(y) = 0$. More detail concerning the analysis of this differential equation is provided in Appendix C.2.

The left-hand side of the differential equation is the elasticity of the take-home rate

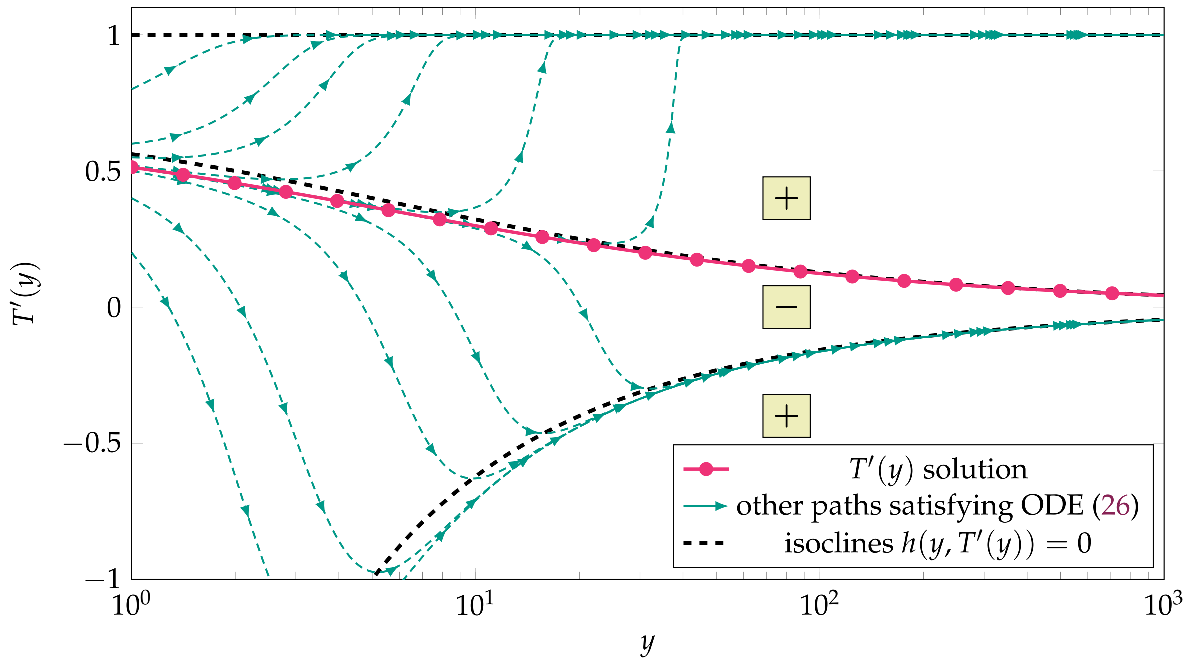

Figure 1: Phase diagram for differential equation (26). The dashed solid lines correspond to isoclines $h(y, T'(y)) = 0$, and the sign of $h(y, T'(y))$ in between these lines is depicted in the boxes. The magenta line with circles corresponds to the unique strictly positive solution that satisfies the terminal condition. Green dashed lines are other trajectories that satisfy (26). The parameters are $\mu = 1, \theta = 1, \alpha = 1.5, \gamma = 2$.

with respect to income

$$-\frac{T''(y)\,y}{1-T'(y)} = \frac{d\log\left(1-T'(y)\right)}{d\log y}.$$

This elasticity must converge to zero as $y \to \infty$. In order for the right-hand side of (25) to converge to zero, the last bracket has to converge to zero. Since $\lim_{y\to\infty} T'(y) = 0$, we obtain equation (21) in Theorem 3.2.

Let us denote

$$h(y,\tau) = \frac{1-\tau}{y}\left[2 - \frac{1+\gamma+\alpha}{1+\gamma}\tau\right]^{-1}\left[\frac{\mu}{\theta}\tau^2 y - \gamma + \gamma\frac{1+\gamma+\alpha}{1+\gamma}\tau\right].$$

Then the differential equation (25) can be rewritten as

$$T''(y) = h\left(y, T'(y)\right). \tag{26}$$

In Figure 1, we plot the phase diagram for this differential equation. The differential equation exogenously fixes the marginal social value of resources μ, which must be determined separately and jointly with the lump-sum tax on the lowest worker type $T(y(\underline{z}))$ so that the planner's budget constraint holds. We reiterate that this characterization holds for the right tail of the type distribution that has a Pareto density and for which the planner has no welfare concerns. This solution for the right tail can then be combined with that for the rest of the type distribution that possibly has a different shape and for which welfare concerns are nonzero, using the general expression for the marginal tax rate given by (13)–(14). The marginal social value of resources μ connects the solutions.

The black dashed lines are the isoclines for which the slope of the marginal tax curve is equal to zero. Since the solution must satisfy the terminal condition $\lim_{y\to\infty} T'(y) = 0$ and must be strictly positive, these isoclines bound the solution into the positive part of the region denoted with the minus sign.

The solution is depicted with the magenta line with circles. Taking this trajectory from the perspective of an initial value problem, the solution constitutes an unstable saddle path. Starting from any other initial condition, the trajectories satisfying equation (26) converge to one of two stable saddle paths visible in the graph, so they either converge to one, or become negative. This also verifies that the terminal condition $\lim_{y\to\infty} T'(y) = 0$ pins down a unique strictly positive solution. In addition, this solution for the marginal tax rate must be strictly decreasing.

3.4 Generalizations

The preceding analysis studies the case when workers have quasilinear preferences and the planner has no welfare concerns for high-type workers in the right tail of the productivity distributions. In this subsection, we briefly discuss generalizations of these results, with detailed calculations provided in Appendix D. The central insight is that the marginal tax converging to zero at an exponential rate equal to (at least) $-\frac{1}{2}$ is a robust result that holds in a range of extensions.

3.4.1 Concave separable preferences

We first consider the case of general separable isoelastic preferences of the form

$$U(c,n) = \frac{c^{1-\rho}-1}{1-\rho} - v\frac{n^{1+\gamma}}{1+\gamma} \tag{27}$$

where ρ is the inverse of the consumption elasticity, with logarithmic utility as the $\rho \to 1$ limit. We still focus on the right tail of the type distribution for which we assume no welfare concerns on the planner's side, $\psi(z) = 0$. In this case, the optimality conditions for the Hamiltonian (10) yield the optimal tax formula in the form

$$\frac{T'(y(z))}{1-T'(y(z))}=(1+\gamma)\frac{1-\widetilde{F}_\rho(z)}{z\widetilde{f}_\rho(z)}, \tag{28}$$

where

$$\widetilde{f}_\rho(z)=\bar{u}^{-1}(c(z))^\rho m(z) f(z) \tag{29}$$

is now the inverse marginal utility weighted density under the worst-case model, with normalization constant

$$\bar{u}=\int_{\underline{z}}^{\bar{z}}(c(\zeta))^\rho m(\zeta) f(\zeta)\, d\zeta=\widetilde{\mathbb{E}}[c^\rho],$$

and $c(z)=y(z)-T(y(z))$. The function $\widetilde{F}_\rho(z)$ is the corresponding cumulative distribution function associated with density $\widetilde{f}_\rho(z)$ defined in (29).

Theorem 3.3. *Assume that worker's preferences are given by the separable form (27), underlying productivity has a right tail that is Pareto distributed, and $\theta<\infty$. Then the marginal tax rate vanishes at the top,*

$$\lim_{y\to\infty}T'(y)=0.$$

More specifically, the limiting tax rate satisfies

$$\lim_{y\to\infty}\frac{\mu}{\theta}\left[T'(y)\right]^2 y=(\gamma+\rho)\,\widetilde{\mathbb{E}}[c^\rho], \tag{30}$$

and hence the limiting elasticity of the marginal tax with respect to income is equal to

$$\lim_{y\to\infty}\frac{d\log T'(y)}{d\log y}=-\frac{1}{2}.$$

This theorem generalizes Theorems 3.1 and 3.2 to the case of concave marginal utility of consumption. Setting $\rho=0$ recovers the special quasilinear case. As in the quasilinear case, the zero limiting marginal tax is preserved in cases when the density that characterizes the right tail of the distribution is thinner than Pareto, even though the decay rate may then be faster than $-\frac{1}{2}$.

Intuitively, decreasing marginal utility from consumption effectively reduces the elasticity of labor supply with respect to the productivity. But since the zero limiting marginal tax result does not depend on the labor supply elasticity to start with, it is also robust with respect to the introduction of more general separable utility form (27). The only difference is expression (30) that adjusts for the consumption elasticity, with the last term $\widetilde{\mathbb{E}}[c^\rho]$ converting goods to their marginal social value units.

3.4.2 Welfare concerns at the top

When the welfare function does not assign zero weights $\psi(z)$ to top productivity types, the determination of the limiting tax must also take into account the distortions under the worst-case model that are due to welfare concerns. To illustrate the consequences, we consider here the case of a utilitarian planner with $\psi(z) \equiv 1$. The worst-case distortion then takes the form

$$m(z) = \bar{m} \exp\left(-\frac{1}{\theta}\left[\mathcal{U}(z) + \mu T(y(z))\right]\right), \tag{31}$$

and the marginal tax formula can be expressed as

$$\frac{T'(y(z))}{1 - T'(y(z))} = (1+\gamma) \frac{\widetilde{\Psi}(z) - \widetilde{F}_{\rho}(z)}{z \widetilde{f}_{\rho}(z)}, \tag{32}$$

where $\widetilde{f}_{\rho}(z)$ is the inverse marginal utility weighted worst-case density (29), $\widetilde{F}_{\rho}(z)$ the corresponding cumulative distribution function, and the cumulative welfare weight $\widetilde{\Psi}(z)$ specializes to

$$\widetilde{\Psi}(z) = \int_{\underline{z}}^{z} \widetilde{f}(\zeta)\, d\zeta.$$

Comparing $\widetilde{F}_{\rho}(z)$ and $\widetilde{\Psi}(z)$, we have $\widetilde{\Psi}(z) > \widetilde{F}_{\rho}(z)$ whenever $\rho > 0$, which reflects the redistributive motives of the utilitarian planner. When $\rho = 0$, the redistributive motive is absent, and $\widetilde{\Psi}(z) = \widetilde{F}_{\rho}(z)$.

The worst-case distortion $m(z)$ in (31) now combines the contributions of welfare and budgetary concerns. The distortion from the utility term $\mathcal{U}(z)$ reflects the concern that there are fewer high-type workers in the distribution, which directly adversely affects the planner's objective function. Since both $\mathcal{U}(z)$ and $T(y(z))$ are strictly increasing in z, both concerns imply a strictly decreasing $m(z)$. Which of the two terms dominates depends on the curvature of the utility function.

Theorem 3.4. *Assume that worker's preferences are given by the separable form (27), underlying productivity has a right tail that is Pareto distributed, the planner is utilitarian with $\psi(z) \equiv 1$, the curvature of the utility function is $\rho > 0$, and $\theta < \infty$. Then*

$$\lim_{y \to \infty} T'(y) = 0,$$

and

$$\lim_{y \to \infty} \frac{d \log T'(y)}{d \log y} = \min\left(-\frac{1}{2}, \rho - 1\right).$$

The theorem shows that the direct welfare concern dominates when the preferences are sufficiently elastic, $\rho < \frac{1}{2}$. In this case, the limited curvature of the utility function implies that the utility term $\mathcal{U}(z)$ grows faster than the tax revenue that determines the budgetary concern. However, the decay rate $-\frac{1}{2}$ derived for the benchmark model constitutes the slowest rate of decay we can anticipate.

3.4.3 Power divergence functions

The objective function (7) of the robust planner penalizes deviations from the benchmark using an entropy penalty, also known as the Kullback–Leibler divergence. While entropy is a natural penalty choice from a statistical perspective, the tendency toward lower progressivity holds for more general divergence functions. Here, we consider the Cressie and Read (1984) class of power divergence functions analyzed, for example, in Almeida and Garcia (2017) or Borovička et al. (2016). The class of divergences is given by $\mathcal{E}_\eta(m) = \mathbb{E}\left[\phi_\eta(m)\right]$ with

$$\phi_\eta(m) = \frac{m^{1+\eta} - 1}{\eta(1+\eta)},$$

where $m(z) = \widetilde{f}(z)/f(z)$ and $\eta \in \mathbb{R}$. Divergences $\mathcal{E}_\eta(m)$ for $\eta \in \{-1, 0\}$ are constructed by appropriate limiting arguments, yielding the entropy $\mathcal{E}_0(m) = \mathbb{E}[m \log m]$ in the limit as $\eta \to 0$. Relative to entropy, power divergences $\mathcal{E}_\eta(m)$ for $\eta > 0$ penalize relatively more the deviations in the left tail of the distribution, while for $\eta < 0$ they penalize more strongly the right tail.

Appendix D.5 provides more detail on power divergences. It also shows that replacing $\theta\mathcal{E}_0(m)$ with $\theta\mathcal{E}_\eta(m)$ in the planner's objective function (7) leads, in the case when utilitarian concerns are absent in the right tail of the productivity distribution, to the same optimal tax formula (15) but the worst-case distortion (16) is now given by

$$m(z) = \left[\frac{\eta}{\theta}\left(\chi - \mu T(y(z))\right)\right]^{\frac{1}{\eta}} \tag{33}$$

where χ is the Lagrange multiplier on the constraint $\mathbb{E}[m] = 1$. As in the entropy case, the worst-case distortion is decreasing because the expression for the optimal marginal tax implies that the marginal tax is positive.

Specializing to the case of quasilinear utility and Pareto-distributed productivity under the benchmark distribution, the differential equation for the optimal marginal tax (25) now takes the form

$$-\frac{T''(y)\,y}{1-T'(y)} = -\left[2 - \frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right]^{-1}\left[-\frac{\mu}{\lambda}\frac{\left[T'(y)\right]^2 y}{\mu T(y) - \chi} - \gamma + \gamma\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right].$$

This leads to the following result.

Theorem 3.5. *Assume that worker's preferences are given by the quasilinear form (11), the type distribution satisfies Assumption 2 with $\bar{z} = \infty$, the divergence penalty in the planner's problem is $\theta \mathcal{E}_\eta(m)$, and $\theta < \infty$. Then the optimal marginal tax rate $T'(y(z))$ for an agent with type z under the robust planner is lower than under the planner without model misspecification concerns.*

Moreover, assume in addition that the underlying productivity has a right tail that is Pareto distributed with shape parameter α. When $\eta \geq 0$, then the marginal tax rate at the top satisfies

$$\lim_{y \to \infty} T'(y) = 0. \tag{34}$$

When $\eta < 0$, then the marginal tax rate at the top is given by

$$\lim_{y \to \infty} T'(y) = \tau_\eta = \frac{1+\gamma}{1+\gamma+\widetilde{\alpha}} \qquad \text{with } \widetilde{\alpha} = \alpha - \frac{1+\gamma}{\gamma}\frac{1}{\eta} > \alpha. \tag{35}$$

Expressions (34) and (35) show that the asymptotic top marginal tax is continuous in the parameter η that indexes the power divergences. As $\eta \nearrow 0$, it also converges to the entropy case with zero asymptotic marginal tax.

Details of the derivation are provided in Appendix D.5. To provide intuition, consider first the case $\eta > 0$. In this case, the worst-case distortion formula (33) implies that $\chi - \mu T(y) > 0$. This means that $T(y)$ must be bounded from above, and since the marginal tax $T'(y)$ has to be positive, it must converge to zero.

When $\eta < 0$, expression (33) implies that $\chi - \mu T(y) < 0$, so $T(y)$ can be unbounded. Conjecturing that the limiting marginal tax $\tau_\eta \in (0, 1)$, an asymptotic approximation analogous to that in (18) and (19) implies that the worst-case distortion (33) asymptotically behaves as a power function of z. This means that the worst-case density $\widetilde{f}(z) = m(z) f(z)$ behaves asymptotically as a Pareto density, with an adjusted shape parameter $\widetilde{\alpha}$. This shape parameter becomes arbitrarily large as $\eta \nearrow 0$, when the divergence function approaches entropy, and, in this case, the asymptotic marginal tax rate approaches zero.

4 Quantitative application

In Section 3, we provided a theoretical characterization of the tail behavior of the optimal tax function when the planner is concerned about misspecification of the type distribution. We now focus on a quantitative evaluation of the whole tax function. Specifically, we are interested in a plausible calibration of the magnitude of the misspecification concerns, and implications for the relative distortions across the type distribution.

4.1 Model calibration

We base our benchmark model calibration on the results in Heathcote and Tsujiyama (2021), who use labor income data from the Survey of Consumer Finances (SCF) to infer the productivity distribution. Heathcote and Tsujiyama (2021) argue that the SCF provides substantially more information about the right tail of the productivity distribution than other household surveys like the Current Population Survey (CPS).

They also show that the labor income distribution in SCF is very well approximated by using the exponentially modified Gaussian (EMG) distribution for the logarithm of the productivity $x = \log z$. The EMG distribution describes the sum of a normal and an exponential random variable, with density given by

$$\begin{aligned} f_x\left(x; \bar{\mu}, \sigma, \alpha\right) &= \frac{\alpha}{2} e^{\frac{\alpha}{2}\left(2\bar{\mu}+\alpha\sigma^2-2x\right)} \operatorname{erfc}\left(\frac{\bar{\mu}+\alpha\sigma^2-x}{\sqrt{2}\sigma}\right) \\ \operatorname{erfc}\left(x\right) &= \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt. \end{aligned}$$

This distribution implies that productivity $z = \exp\left(x\right)$ has support $\left(0, \infty\right)$, the left tail follows the log-normal distribution with parameters $\bar{\mu}, \sigma$, and the right tail is asymptotically Pareto distributed with shape parameter α. We can therefore invoke the theoretical results derived in Section 3 for the case of the Pareto-distributed right tail.[8] We denote the class of EMG distributions as $\boldsymbol{F}_{emg}$. As in Heathcote and Tsujiyama (2021), we choose $\alpha = 2.2$ and $\sigma^2 = 0.142$. The first moments of the logarithm and the level of the productivity distribution are given by

$$E\left[\log z\right] = \bar{\mu} + \frac{1}{\alpha} \qquad E\left[z\right] = \exp\left(\bar{\mu} + \frac{1}{2}\sigma^2\right)\frac{\alpha}{\alpha-1}.$$

We set $\bar{\mu}$ to normalize the average productivity to $E\left[z\right] = 1$.

We further assume a utilitarian planner endowed with a concave separable isoelastic utility function (27), with parameters $\rho = 1$ implying logarithmic utility from consumption, $\psi = 1$, and $\gamma = 2$. We choose $V\left(G\right) = \bar{v}G$, and set the marginal value of government expenditures $\bar{v}$ so that $G = 0$ under the optimal tax scheme, implying that the government only taxes for redistributive purposes.

Under the isoelastic utility function (27), as long as the asymptotic optimal marginal tax satisfies $\lim_{z\to\infty} T'\left(z\right) < 1$, the right tail of the income distribution inherits the Pareto property, with shape parameter $\frac{\rho+\gamma}{1+\gamma}\alpha$. Combined with logarithmic utility, $\rho = 1$, the income distribution is asymptotically Pareto distributed with shape parameter α.

[8] Truncating the Gaussian component of the distribution $f_x\left(x\right)$ for a sufficiently high value of x so that the tail has an exact Pareto distribution is quantitatively inconsequential.

4.2 Quantifying model misspecification concerns

In order to quantify the magnitude of model misspecification concerns, we use the following strategy for the calibration of the entropy ball parameter κ. Since changes in the income tax schedule are infrequent, we envision that the planner designs the tax schedule and commits to it for some foreseeable future, say five years. We then ask how much uncertainty in the income distribution can the planner plausibly anticipate over this period.

We proceed as follows. We download data on quantiles of the annual U.S. income distributions for years 1966–2015 from the World Inequality Database.[9] For each year t we fit an EMG distribution f_t with parameters $(\bar{\mu}_t, \sigma_t, \alpha_t)$ to the logarithm of the income distribution described by these quantile data. We then conservatively normalize the parameters $\bar{\mu}_t$ so that all estimated income distributions have unit mean in levels, to abstract from differences in the distributions associated with aggregate growth in the economy.

Subsequently, for each 5-year window $\{t, \ldots, t+4\}$, we construct the smallest entropy ball $\mathcal{F}\left(\overline{f}_t, \kappa_t\right)$ that includes all the estimated distributions $f_t, \ldots, f_{t+4}$:

$$\kappa_t = \min\left\{\kappa : \exists \overline{f}_t \in \boldsymbol{F}_{emg} \text{ s.t. } f_{t+i} \in \mathcal{F}\left(\overline{f}_t, \kappa\right), i = 0, \ldots, 4\right\}.$$

We treat the EMG distribution $\overline{f}_t$ at the center of the entropy ball as the benchmark distribution. Given the knowledge of this benchmark distribution, the planner contemplates any of the distributions f_{t+i}, $i = 0, \ldots, 4$, as plausible. However, the set $\mathcal{F}\left(\overline{f}_t, \kappa_t\right)$ also contains other distributions in the EMG parameteric class as well as a much larger nonparameteric set of distributions $\widetilde{f}$ that all satisfy $\mathcal{E}\left(\overline{f}_t, \widetilde{f}\right) \leq \kappa_t$.

Figure 2 displays the outcome of the procedure. Each circle represents the estimated parameters (α_t, σ_t) for a given year of the data. The data points capture a visible trend that reflects increasing inequality—the circles move over time from the bottom right to the top left corner, reflecting increases in the thickness of the right tail (lower α_t) and increases in the dispersion of the left tail (higher σ_t).

The orange curves capture three examples of smallest entropy balls $\mathcal{F}\left(\overline{f}_t, \kappa_t\right)$ constructed using five-year periods of data, with centers $\overline{f}_t$ captured by the solid red dots. For example, the entropy ball for period 1966–1970 covers all five distributions for this period, with distributions for years 1969 and 1970 exactly at the boundary. Taking the distribution $\overline{f}_t$ as the benchmark distribution, the planner views all EMG distributions within the orange curve are viewed as plausible. However, as mentioned above, many other distributions within $\mathcal{F}\left(\overline{f}_t, \kappa_t\right)$ are not represented in the figure as they fall outside the EMG parameteric class.

[9] https://wid.world

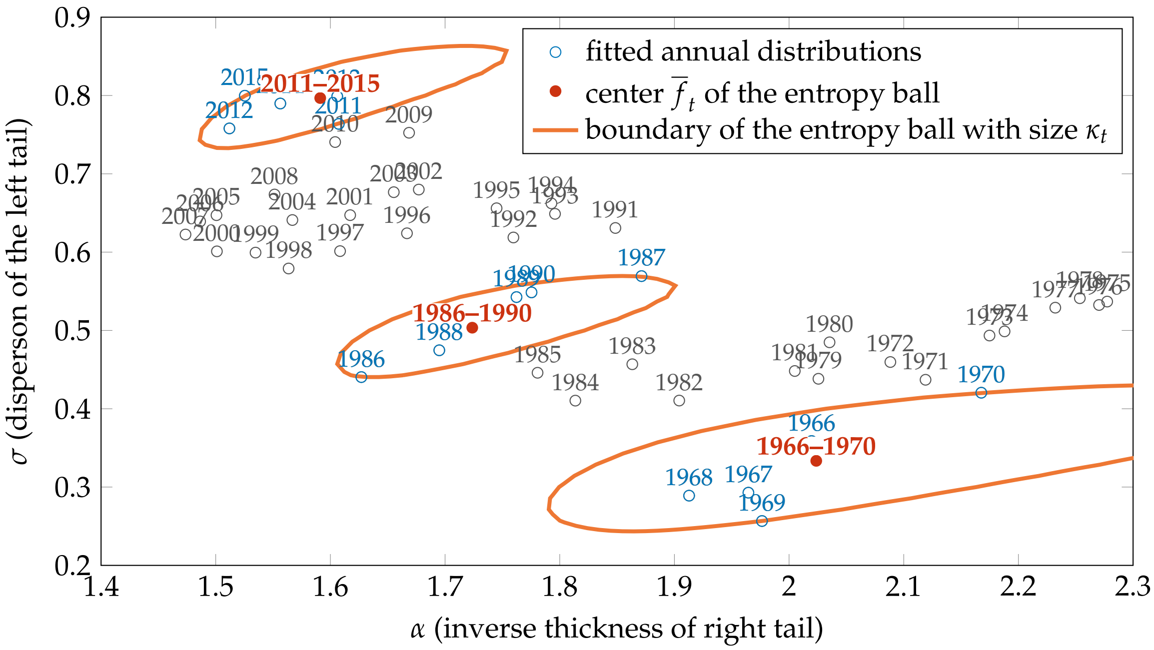

Figure 2: Estimated income distributions for U.S. data, 1966–2015. Each point shows the combination of parameters (α_t, σ_t) of the estimated EMG distribution for the given year. The orange curves represents three example entropy balls, centered at the solid red dots, for periods 1966–1970, 1986–1990, and 2011–2015.

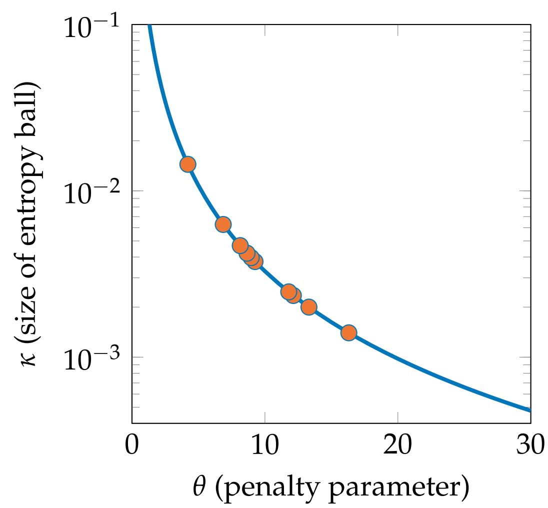

Figure 3: Relationship between the penalty parameter θ and entropy κ for the model calibrated in Section 4.1.

Figure 3 converts the entropy constraint κ into the corresponding penalty parameter θ for the model calibration from Section 4.1. In our results, we will highlight implications for the median and smallest size of the entropy ball, corresponding to $\theta = 9.254$ and $\theta = 16.312$, respectively. In each of the economies, we recalibrate $\bar{v}$ so that $G = 0$ under the worst-case model.

4.3 Asymptotic behavior of the marginal tax rate

We start with the following result that characterizes the marginal tax rate for the left and right tail of the productivity distribution.

Lemma 4.1. *When the logarithm of productivity z follows the exponentially modified Gaussian with parameters $(\bar{\mu}, \sigma, \alpha)$, the following results hold.*

1. *In the rational model ($\theta = \infty$), the asymptotic marginal tax rate is given by*

$$\lim_{y \to \infty} T'(y) = \frac{(\gamma + \rho)(1 + \gamma)}{\alpha(\gamma + \rho) + \gamma(1 + \gamma)}.$$

2. *Under model misspecification concerns ($\theta < \infty$),*

$$\lim_{z \to \infty} T'(y) = 0. \tag{36}$$

$$\lim_{y \to \infty} \frac{d \log T'(y)}{d \log y} = \min\left(-\frac{1}{2}, \rho - 1\right) \tag{37}$$

3. *For the left tail of the productivity distribution,*

$$\lim_{y \to 0} T'(y) = 0,$$

irrespective of the value of θ.

A sketch of the proof is provided in Appendix D. The first result evaluates the tax formula (32) that reflects planner's utilitarian concerns for the workers in the right tail of the productivity distribution, and the concave shape of the utility function. The limit aligns with the case when the tail is exactly Pareto distributed, since the contribution of the log-normal component in the right-tail vanishes. The second result restates results shown in Theorem 3.4.

Finally, the third result follows from the log-normal shape of the productivity distribution at zero. To understand the reason why the marginal tax rate is zero irrespective of the presence of misspecification concerns, we note that because the marginal utility of consumption is infinite at zero consumption, the planner optimally provides a finite, strictly positive transfer $T(\underline{z})$ to workers with zero productivity $\underline{z} = 0$. Such a transfer has infinite marginal social value against a finite social marginal cost of resources. Consequently, $\mathcal{U}(\underline{z})$ is finite, and the distortion $m(z)$ in (31) is bounded and bounded away from zero in the

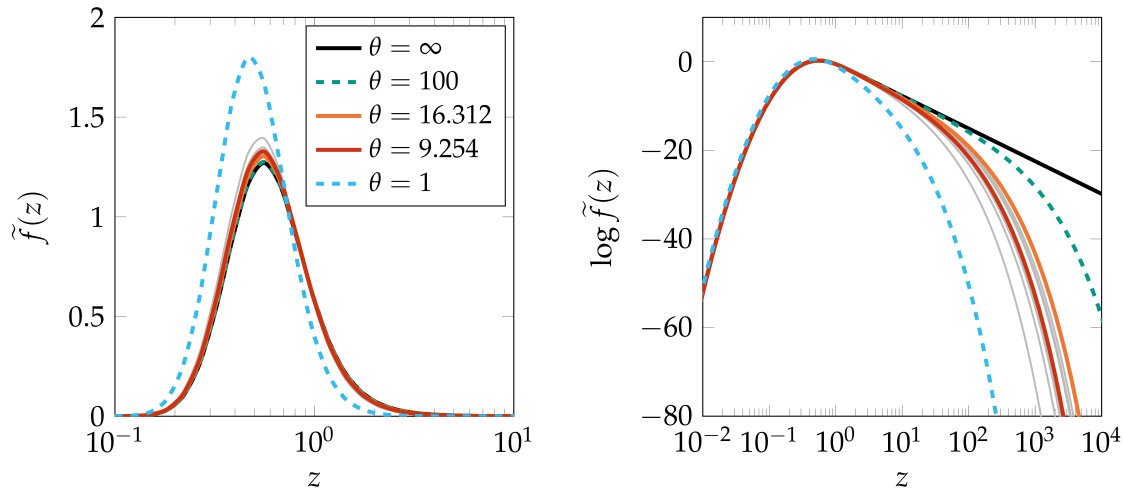

Figure 4: Worst-case distributions $\widetilde{f}(z)$ for alternative levels of misspecification concerns given by θ. The black solid line for $\theta = \infty$ corresponds to the rational benchmark for which $\widetilde{f}(z) = f(z)$. Thin grey lines represent the densities for the sizes of entropy balls constructed using the empirical procedure from Section 4.2 for alternative time periods. The orange line for $\theta = 16.312$ is the case with the smallest misspecification concern, while the red line for $\theta = 9.254$ represents the median case.

neighborhood of $\underline{z} = 0$. Since, as we show in the appendix, the limiting tax for the left tail is zero in the rational model, a perturbation of the tax rate formula (32) by a finite $m(z)$ will not alter the zero limiting marginal tax rate in the model with misspecification concerns.

4.4 Worst-case distributions and marginal tax rates

Figure 4 plots the worst-case productivity distributions $\widetilde{f}(z)$ across different choices of the parameter θ. The black line with an infinite entropy penalty $\theta = \infty$ corresponds to the rational case for which $\widetilde{f}(z) = f(z)$, with subsequent lines representing increasing misspecification concerns as θ decreases.

The left panel shows the distribution in the proximity of the mean under the benchmark model, which is equal to $\mathbb{E}[z] = 1$, while the right panel focuses on the broad range of productivities z and plots the density in logarithms to highlight distortions in the tail. In the log-log plot in the right panel, the straight black line in the right tail reflects the Pareto shape of the right tail under the benchmark distribution $f(z)$.

As is apparent from the right graph, misspecification concerns lead to sharp distortions in the right tail of the distribution. This is in line with theoretical results from Section 3 that show that an arbitrarily small amount of misspecification concerns leads to a thin-tailed worst-case distribution. The parameter θ determines the range of incomes at which the concerns start to manifest themselves in a more pronounced way.

quantiles \ θ	∞	100	16.312	9.254	1
$\widetilde{q}(0.01)$	0.262	0.261	0.259	0.258	0.227
$\widetilde{q}(0.05)$	0.349	0.348	0.345	0.343	0.297
$\widetilde{q}(0.25)$	0.537	0.536	0.530	0.525	0.438
$\widetilde{q}(0.50)$	0.748	0.746	0.734	0.724	0.581
$\widetilde{q}(0.75)$	1.098	1.093	1.067	1.045	0.784
$\widetilde{q}(0.95)$	2.317	2.287	2.158	2.062	1.267
$\widetilde{q}(0.99)$	4.815	4.678	4.158	3.823	1.864
$\widetilde{q}(0.999)$	13.713	12.657	9.776	8.391	2.993

Table 1: Quantiles of the distribution of productivity z under the worst-case distribution for alternative values of θ. The case $\theta = \infty$ corresponds to the rational case.

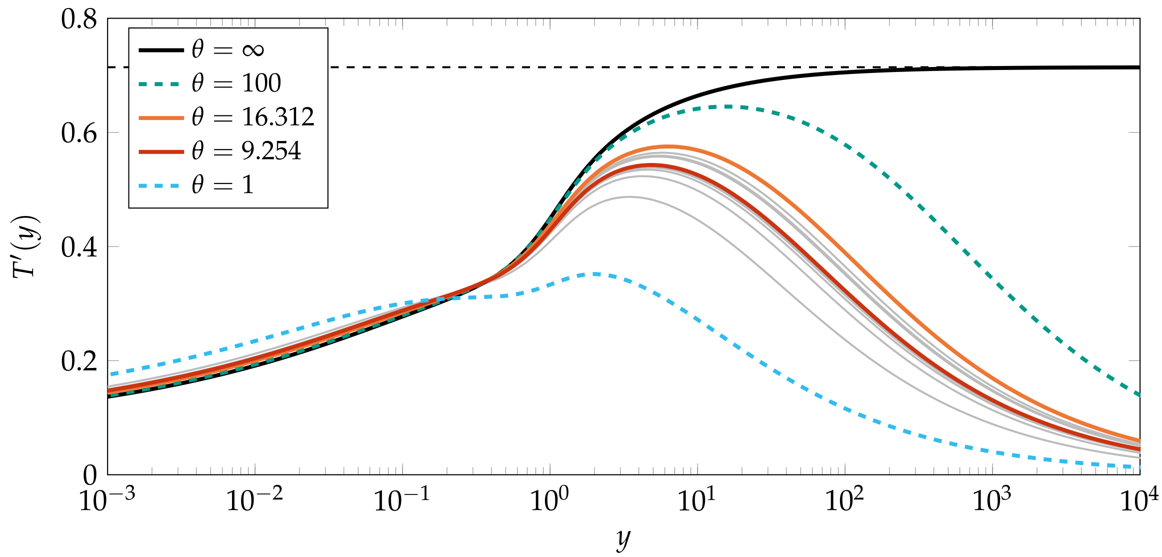

Figure 5: Optimal marginal tax schedules for alternative levels of misspecification concerns. The dashed line corresponds to the limiting marginal tax rate for the rational case.

At the same time, the center of the distribution gets noticeably distorted only as misspecification concerns become substantial ($\theta = 1$), and the left tail remains essentially undistorted. These conclusions are also confirmed in Table 1 that tabulates the quantiles of the productivity z under the alternative worst-case distributions $\widetilde{f}(z)$.

Figure 5 shows the optimal marginal tax rate schedule for alternative levels of the model misspecification concerns. The solid black line represents the marginal tax rate for the rational case. In line with the literature, since the underlying productivity distribution exhibits a Pareto tail, the asymptotic tax rate $\lim_{z\to\infty} T'(y(z))$ is positive and quantitatively large, at 71.4%.[10] The tax rate asymptotes to zero as $z \to 0$, in line with Lemma 4.1.

When misspecification concerns are present, the shape of the optimal tax schedules

[10] In Heathcote and Tsujiyama (2021), the computed marginal tax rates in the right tail asymptote to zero because they truncate the distribution and focus on numerical solutions for the truncated case.

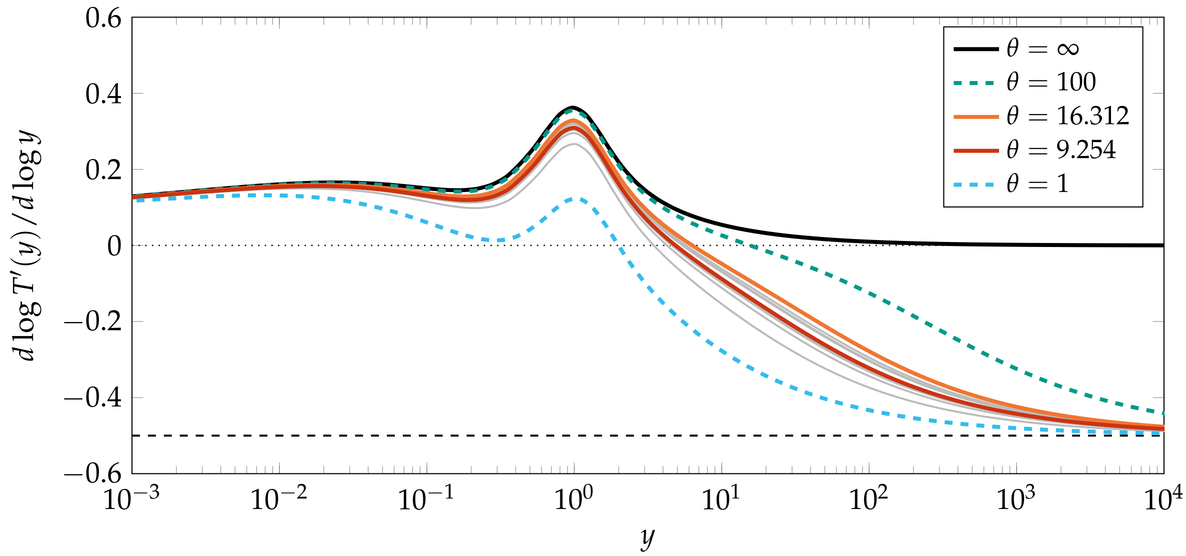

Figure 6: Growth rates of optimal marginal tax schedules, $d \log T'(y) / d \log y$, for alternative levels of misspecification concerns. The dashed black line corresponds to the theoretical limit, equal to $-1/2$.

looks notably different. While for wage levels around the mean ($\mathbb{E}[z] = 1$), the optimal marginal tax looks similar to that under the rational case, it starts departing quickly for higher wage levels. For the benchmark case of $\theta = 16.312$, the marginal tax peaks at 57.5% for wages equal to 7.5 the average wage ($\arg\max_y T'(y) / \mathbb{E}[y]$), and starts declining thereafter.

Theoretical results from Section 3 show that not only should marginal tax rates decline to zero as $z \to \infty$, but the asymptotic rate of decline is also pinned down. From Lemma 4.1, given that $\rho = 1$, we have that the rate of decline should converge to $-\frac{1}{2}$. Figure 6 verifies this result numerically. The solid black line, representing the rational case, is above zero and converges to zero, reflecting an increasing marginal tax rate schedule converging to a positive limiting tax rate. On the other hand, across all levels of misspecification concerns, the decay rate indeed asymptotically converges to the theoretically predicted value.

4.5 Insurance provision and budgetary concerns

The decision problem of the robust planner trades off utilitarian and budgetary concerns, reflected in the worst-case distortion

$$m(z) = \bar{m} \exp\left(-\frac{1}{\theta}\left[\mathcal{U}(z) + \mu T(y(z))\right]\right). \tag{38}$$

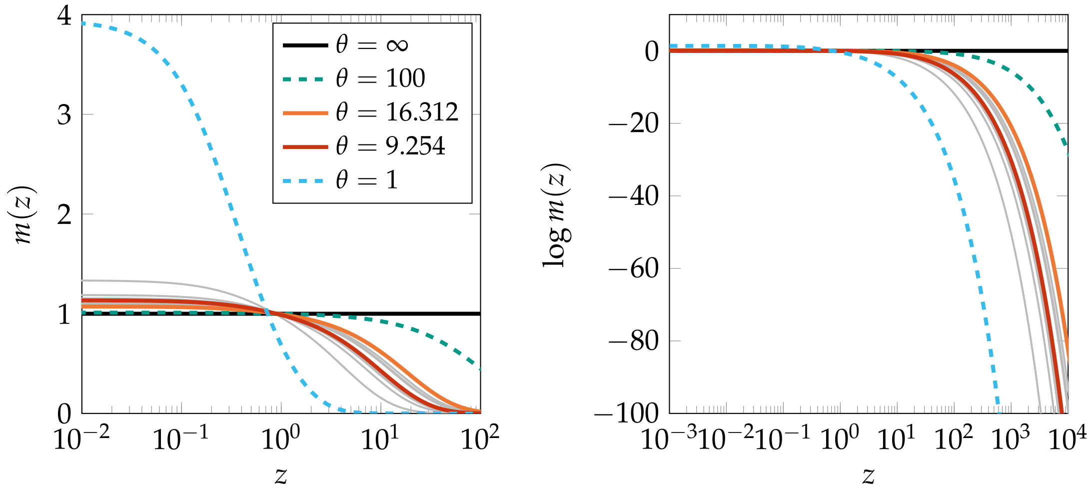

Figure 7: The likelihood ratio $m(z) = \widetilde{f}(z)/f(z)$ representing the distortion of the worst-case distribution relative to the benchmark distribution, plotted for alternative levels of misspecification concerns given by θ. The case $\theta = \infty$ corresponds to the rational benchmark for which $m(z) = 1$.

On the one hand, the planner is concerned that there are more agents in the left tail of the productivity distribution, who receive low utility $\mathcal{U}(z)$ and generate low (in fact negative) net tax revenue $T(y(z))$. The planner can diminish the utilitarian concerns by providing more insurance, thus raising $\mathcal{U}(z)$ and lowering $m(z)$. This insurance comes in the form of transfers, and hence at the cost of a lower net tax revenue $T(y(z))$. On the margin, the optimal tax schedule designed by the planner trades off a unit of tax revenue at the marginal social value μ against a unit of consumption transferred to the low-productivity agent at marginal value $\mathcal{U}'(z)$.

Any transfers provided to low-productivity workers must come from those in upper parts of the productivity distribution. As shown in Section 3, when the utility function is sufficiently concave ($\rho > \frac{1}{2}$), budgetary concerns dominate the shape of the distortion (38) in the right tail of the productivity distribution. Following the tax formula (32), the planner chooses the marginal tax rate for a particular z as a result of a tradeoff between the tax distortion imposed on the $\widetilde{f}(z)$ agents at z against the social benefit of raising lump sum revenue from the (marginal utility weighted) mass $1 - \widetilde{F}(z)$ of agents with productivities above z. Since the misspecification concern increases with z, the planner fears that the mass of agents above z is lower relative to agents residing exactly at z who are distorted by the marginal tax at z, and hence opts for lower marginal tax rates in the tail.

This desire to lower marginal taxes at the top combined with concerns about the higher prevalence of low-productivity and lower prevalence of high-productivity workers increases the marginal social value of a unit of tax revenue μ, pushing toward lower overall redistribution.

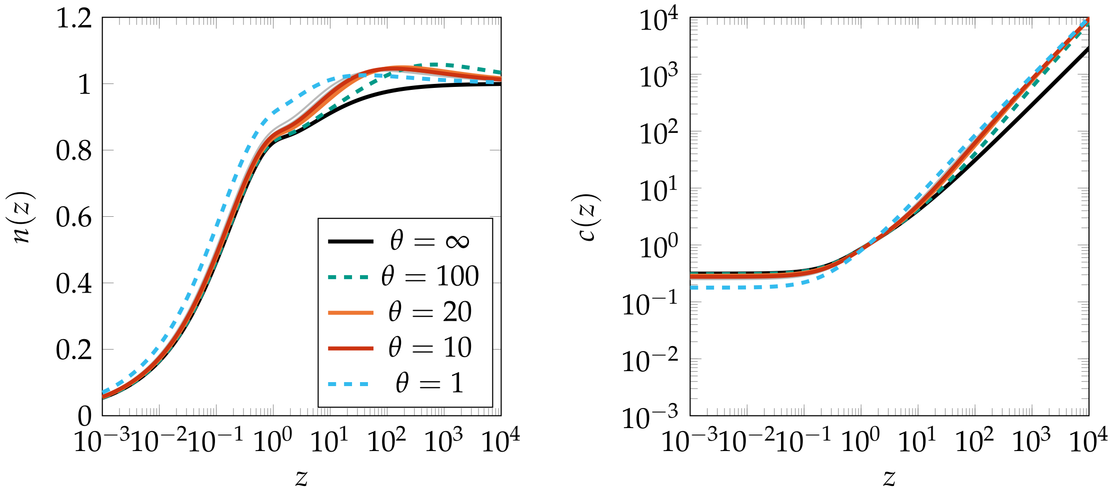

Figure 8: Labor supply under optimal tax schedules (left panel), and consumption as a function of income (right panel) under alternative levels of misspecification concerns.

Figure 7 represents these implications quantitatively by plotting the shape of the distortion $m(z) = \widetilde{f}(z) / f(z)$. The plots reveal that in the left tail of the productivity distribution, the utilitarian and budgetary concerns reflected in the shape of $m(z)$ are minimal for the benchmark choice $\theta = 16.312$. Without any redistribution scheme, $\lim_{z \to 0} \mathcal{U}(z) = -\infty$, and consequently $\lim_{z \to 0} m(z) = \infty$, as the utilitarian concerns about the low-productivity workers dominate. However, the optimal tax schedule insures the low-productivity workers sufficiently, leading to bounded and quantitatively modest distortions of the left tail. Only when the misspecification concerns are substantial ($\theta = 1$), the planner's concerns about insufficient tax revenue to insure low-productivity workers start increasing more notably.

On the other hand, in the right tail of the productivity distribution, the concerns about loss of tax revenue from high-productivity workers are severe. The planner, concerned about the distortion of the labor supply at the top, lowers marginal tax rates asymptotically to zero, yet the marginal tax rate declines sufficiently slowly to make the tax revenue $T(y(z))$ from each high-productivity worker grow without bound as $z \to \infty$. Consequently, the planner's concerns that these high-productivity workers are less prevalent than assumed under the benchmark distribution also grow without bound, leading to $\lim_{z \to \infty} m(z) = 0$.

The left panel in Figure 8 depicts the labor supply under the optimal tax schedule for alternative levels of the misspecification concerns. The lower marginal tax rates when misspecification concerns are present generally increase labor supply in the right tail of the productivity distribution. The increase in labor supply combined with a lower tax burden translates to higher consumption leves for high-productivity workers, as shown in the

moments \ θ	∞	100	16.312	9.254	1
$\mathbb{E}[z]$	1.000	1.000	1.000	1.000	1.000
$\widetilde{\mathbb{E}}[z]$	1.000	0.987	0.944	0.914	0.657
$\mathbb{E}[y]$	0.823	0.827	0.841	0.850	0.918
$\widetilde{\mathbb{E}}[y]$	0.823	0.815	0.787	0.768	0.579
μ	1.215	1.227	1.270	1.303	1.727
T_0	-0.315	-0.308	-0.289	-0.276	-0.178
$\max_y T'(y)$ (%)	71.4	64.5	57.5	54.3	35.2
$\arg\max_y T'(y)$	∞	15.549	6.305	4.856	1.993
$\mathbb{E}[T]$	0.000	0.007	0.027	0.039	0.110
$\widetilde{\mathbb{E}}[T]$	0.000	0.000	0.000	0.000	0.000
$\mathbb{E}[T]/\mathbb{E}[y]$	0.000	0.008	0.032	0.046	0.120
$\widetilde{\mathbb{E}}[T]/\widetilde{\mathbb{E}}[y]$	0.000	0.000	0.000	0.000	0.000

Table 2: Moments under the benchmark distribution $f(z)$ and the worst-case distribution $\widetilde{f}(z)$ under optimal tax schedules for alternative values of θ.

right panel of Figure 8.

In the left tail of the productivity distribution, the effects on labor supply are much more modest. The increase in labor supply, which manifests itself only for high levels of misspecification concerns ($\theta = 1$), is driven by the wealth effect of a lower lump sum transfer $T_0 = T(y(\underline{z}))$. At the same time, the right panel shows that low productivity workers who produce little output are well insured by the optimal tax scheme, regardless of the level of misspecification concerns.

Table 2 summarizes these insights in the form of moments under the benchmark and worst-case distributions. The benchmark mean $\mathbb{E}[z]$ is identical and normalized to one across parameterizations since the distribution $f(z)$ is exogenously specified by the calibrated EGM distribution in Section 4.1. The worst case means $\widetilde{\mathbb{E}}[z]$ decrease as misspecification concerns increase, reflecting shifts toward more adversely slanted productivity distributions subjectively perceived by the robust planner.

Objective means of the income distribution $\mathbb{E}[y(z)] = \mathbb{E}[zn(z)]$ increase with increasing misspecification concerns, as lower marginal taxes increase labor supply, plotted in the left panel of Figure 8. Nevertheless, the mean of the income distribution under the worst case distribution, $\widetilde{\mathbb{E}}[y]$, declines.

As mentioned above, since increases in misspecification concerns are manifested in increases in the marginal social value of tax funds μ, reflecting planner's fear about more severe scarcity of tax revenue.

Finally, the bottom part of Table 2 reports statistics for the tax schedule. The lump sum transfer to the lowest productivity worker, $T_0 = T(y(\underline{z}))$, expressed as a share of average income $\mathbb{E}[y]$, is roughly 33–37%, and rather stable across a range of the values of the parameter θ. Only when misspecifications concerns and the fear of lack of tax revenue worsen severely, the transfer is reduced more substantially.

While the lump sum transfer to the lowest productivity worker is rather insensitive to the degree of misspecification concerns for a range of values of θ, the peak marginal tax rate changes substantially. For the case without misspecification concerns, the top marginal tax rate asymptotes at 71.4%, for the benchmark calibration $\theta = 16.312$, the marginal tax rate peaks at 57.5% at an income $\arg\max_y T'(y)$ corresponding to 7.5 of the average income, and starts declining thereafter.

Since the planner's optimization problem involves running a balanced budget under the endogenously determined subjective distribution, and we chose $G = 0$, this implies $\widetilde{\mathbb{E}}[T] = 0$ for all choices of θ. However, the tax revenue under the benchmark distribution generally differs from zero. Since the worst-case distribution is pessimistically biased, then, compared to the benchmark distribution, the planner underestimates the amount of tax revenue the given tax schedule raises under that benchmark distribution. For the preferred choice $\theta = 16.312$, the extra surplus generated by the tax policy is about 3.2% of total labor income in the economy.

If we interpret the benchmark distribution as the true productivity distribution under the data-generating measure, then the optimal tax policy of the robust planner indeed generates this surplus. This raises the question of what the planner does with the surplus resources. Our model is static so it does not speak to intertemporal tradeoffs but it is plausible to envision a dynamic extension of this model in which the planner also manages the accumulated debt or assets over time. Hansen and Sargent (2012, 2015), Kwon and Miao (2017), Ferriere and Karantounias (2019), or Karantounias (2023) are important contributions in this direction that continue optimal dynamic policies in representative agent frameworks in which the planner is ambiguous about the stochastic path of the aggregate economy. Introducing dynamic debt management into our framework faces novel challenges but is a natural way of moving forward.

5 Multidimensional Robust Taxation

A crucial input for the design of optimal taxes, in addition to the skill distribution, is the value of labor supply elasticity. However, it is a well-known in the public finance literature that substantial variation exists in the values of labor supply elasticities and the elasticity of taxable income employed across various applications. [11] These estimates vary based on the samples considered and the econometric methods applied. Also very little is known

[11] Saez, Slemrod, and Giertz (2012) provide a survey on the elasticity of taxable income.

about how elasticities vary with income.[12]

In this section, we formalize this lack of knowledge as the planner's concerns regarding the joint distribution of labor supply elasticities and productivity, and design nonlinear taxes that are robust to such misspecifications. We proceed in three steps. First, we propose a modification to the entropy penalty utilized to construct a set of alternative one-dimensional distributions. Second, we slightly modify preferences to ensure that we can isolate the uncertainty over labor supply responses to tax reforms from uncertainty in the level of hours worked. Finally, we study an application to revisit the quantitative relevance of this source of uncertainty.

5.1 Modified penalty

Household type $s = (\gamma, z)$ is two-dimensional, comprising a parameter γ that influences labor supply and another parameter z that determines the hourly wage rate. Types are distributed according to a joint density $f(\gamma, z)$, which remains private information to the households. The government is concerned that the joint distribution $f(\gamma, z)$ is misspecified and is contemplating a set of alternative distributions $\tilde{f}(\gamma, z)$. As previously discussed, we can characterize the set using a penalty function.

Factor the joint density $f(\gamma, z)$ into a marginal for skills $f^z(z)$ and conditional $f^{\gamma|z}(\gamma)$. Using two positive scalars, $(\theta_\gamma, \theta_z)$, we postulate a penalty $\mathcal{P}\left(f, \tilde{f}|\theta_\gamma, \theta_z\right)$ for the multidimensional case as follows:

$$
\begin{aligned}
\mathcal{P}\left(f, \tilde{f}|\theta_\gamma, \theta_z\right) &= \theta_\gamma \int \mathcal{E}\left(f^{\gamma|z}(\cdot \mid z), \tilde{f}^{\gamma|z}(\cdot \mid z)\right) m(z) f^z(z)\, dz + \theta_z \mathcal{E}\left(f^z, \tilde{f}^z\right) \\
&= \theta_\gamma \int \left(\int m(\gamma \mid z) \ln m(\gamma \mid z) f^{\gamma|z}(\gamma \mid z)\, d\gamma \right) m(z) f^z(z)\, dz + \theta_z \int m(z) \ln m(z) f(z)\, dz,
\end{aligned}
$$

where $\mathcal{E}(\cdot, \cdot)$ is the relative entropy.

We denote $m(\gamma, z) = \tilde{f}(\gamma, z) / f(\gamma, z)$, $m(\gamma \mid z) = \tilde{f}^{\gamma|z}(\gamma \mid z) / f^{\gamma|z}(\gamma \mid z)$, $m(z) = \tilde{f}^z(z) / f^z(z)$ as the likelihood ratios btween the benchmark and the alternative distributions of (γ, z), γ given z, and z, respectively.

Let $\mathcal{U}(\gamma, z; T)$, and $\mathcal{Y}(\gamma, z; T)$ be the indirect utility and the income of household type (γ, z) under tax policy T. Using the penalty $\mathcal{P}$, the minimizing problem given tax policy T is

[12]A wide range of values is reported in the literature. Chetty et al. (2011) conduct a meta-analysis on the values of labor supply elasticity in both micro and macro studies, finding that a point estimate of the Frisch labor supply elasticity from micro studies is 0.54 for intensive margin and a point estimate for macro studies for aggregate hours is 2.84. Neisser (2021) also conducts a meta-analysis on the value of the elasticity of taxable income and finds estimates vary across regression techniques, sample restrictions, countries and time. Lockwood, Sial, and Weinzierl (2021) elicit the belief on the distribution of the value of the elasticity of taxable income among public economists by running a survey.

$$\min_m \int \psi(\gamma,z)\,\mathcal{U}(\gamma,z;T)\,m(\gamma,z)\,f(\gamma,z)\,d(\gamma,z) + V\left(\int \mathcal{Y}(\gamma,z;T)\,m(\gamma,z)\,f(\gamma,z)\,d(\gamma,z)\right) + \mathcal{P}\left(f,\tilde{f}|\theta_\gamma,\theta_z\right),$$

$$\text{s.t.} \int m(z)\,f^z(z)\,dz = 1 \tag{39}$$

$$\int m(\gamma \mid z)\,f^{\gamma|z}(\gamma \mid z)\,d\gamma = 1 \quad \forall z.$$

The optimal tax problem is defined as the maximization of (39) over tax functions T. The conditions on the likelihood ratios $m(\gamma \mid z)$ and $m(z)$ imply that the likelihood ratio $m(\gamma,z)$ satisfies $\int m(\gamma,z)\,f(\gamma,z)\,d(\gamma,z) = 1$.

The modified penalty term $\mathcal{P}\left(f,\tilde{f}|\theta_\gamma,\theta_z\right)$ satisfies the following properties.

Lemma 5.1. *1. When $\theta_\gamma = \theta_z = \theta$, $\mathcal{P}\left(f,\tilde{f}|\theta_\gamma,\theta_z\right)$ reduces to an entropy penalty over the joint distribution over (γ,z)*

$$\mathcal{P}\left(f,\tilde{f}|\theta,\theta\right) = \theta\mathcal{E}\left(f,\tilde{f}\right).$$

2. When $\theta_\gamma \to \infty$, the minimizing probability measure is not distorted conditional on z, $m(\gamma \mid z) = 1$ for all (γ,z). Similarly, as $\theta_z \to \infty$, the minimizing probability measure is not distorted in terms of z, $m(z) = 1$ for all z.

Proof. See the Appendix. ■

Lemma 5.1 shows that the modified penalty term $\mathcal{P}\left(f,\tilde{f}|\theta_\gamma,\theta_z\right)$ generalizes the usual relative entropy penalty by allowing for flexible weights θ_z and θ_γ on the marginal distribution of productivity z, and the conditional distributions $f^{\gamma|z}$, respectively. This flexibility allows for different degrees of concerns over different parts of the joint distribution. For instance, a high value of θ_z and a low value of θ_γ describes a planner who is more confident about wage data but less sure about estimates of elasticities. In the application, we will use a limit $\theta_z \to \infty$ to understand how much the uncertainty about labor supply elasticities affects the optimal tax schedule.

5.2 Hours versus elasticity

Consider a variant of the preferences discussed in Section 4, $\frac{1}{1-\rho}\left(c - \psi\frac{n^{1+\gamma}}{1+\gamma}\right)^{1-\rho}$. For a differentiable tax function T, the optimal labor supply is determined by

$$\psi n^\gamma = z\left(1 - T'(zn)\right)$$

From the previous expression, we can see that the parameter γ matters for both the level of hours worked and the responsiveness of hours to changes in taxation. Therefore, uncertainty about the parameter γ confounds uncertainty about the responsiveness of labor supply and the level of hours. We are particularly interested in the responsiveness of labor supply to tax changes; and propose an adjustment to the preferences for labor supply to isolate that concern.

Specifically, we consider the following thought experiment. Imagine a government contemplating a tax reform ΔT which changes the tax function $T^0 \to T^0 + \Delta T$ and is concerned about how the labor supply of various individuals will respond due to uncertainty about the distribution of γ. A desirable feature of how one can use the robust approach to model this concern is a requirement that the size of worst-case distortions should be small if the size of the reform is small, or for any level of z,

$$\Delta T \approx 0 \to m(\gamma|z) \approx 1.$$

To do that, we postulate the following utility function of worker with type (γ, z).

$$\frac{1}{1-\rho}\left(c - \Psi(\gamma, z)\frac{n^{1+\gamma}}{1+\gamma}\right)^{1-\rho} + \Delta(\gamma, z). \tag{40}$$

Our goal is to derive restrictions on the shifters Ψ and Δ that guarantee $m(\gamma|z) = 1$ when $\Delta T = 0$. The following lemma summarizes the outcomes for the separable constant elasticity formulation that we throughout in paper.

Lemma 5.2. *Suppose the utilitarian planner is endowed with a concave GHH utility utility function (40), the status quo tax policy T^0 and let $\bar{c}(z;T), \bar{n}(z;T)$ the optimal choices of workers with utility*

$$\bar{u}(z;T) = \max_{c \le zn - T(zn), n} \frac{1}{1-\rho}\left(c - \bar{\psi}\frac{n^{1+\bar{\gamma}}}{1+\bar{\gamma}}\right)^{1-\rho} \tag{41}$$

for some positive scalars $\bar{\psi}$ and $\bar{\gamma}$. There exists labor disutility shifter $\Psi(\gamma, z)$ and the utility shifter $\Delta(\gamma, z)$ so that when $T = T^0$, we have the following:

1. *The optimal choices of workers given the tax policy T, $\mathcal{C}(\gamma, z; T)$, $\mathcal{N}(\gamma, z; T)$ and the optimal level of utility $\mathcal{U}(\gamma, z; T)$ do not depend on γ, and for all γ', z,*

$$\mathcal{C}\left(\gamma', z; T^0\right) = \bar{c}\left(z; T^0\right)$$
$$\mathcal{N}\left(\gamma', z; T^0\right) = \bar{n}\left(z; T^0\right)$$
$$\mathcal{U}\left(\gamma', z; T^0\right) = \bar{u}\left(z; T^0\right).$$

2. *The minimizing probability measure given productivity level z is not distorted, and for all*

γ, z

$$m(\gamma|z) = 1.$$

Proof. See the Appendix. ■

6 Application

Utilizing the modified penalty and labor supply preferences, we assess the impact of uncertainty regarding labor supply responsiveness on the shape of the optimal tax schedule. The key choices are the specification of the benchmark distribution $f(\gamma, z)$ and the values of penalty parameters $(\theta_z, \theta_\gamma)$.

6.1 Calibration

We use the modified GHH preference introduced in the previous section with a curvature parameter ρ and the inverse of the elasticity of taxable income γ [13]. The curvature ρ is kept to 1 as in the one-dimensional case with the separable utility.

Given the lack of evidence on the correlation between productivity and labor supply elasticity, we assume that in the benchmark distribution the elasticity of taxable income γ^{-1} is independent of z, i.e., $f(\gamma, z) = f^{\gamma|z}(\gamma \mid z) f^z(z)$ and $f^{\gamma|z}(\gamma \mid z)$ does not depend on z. The benchmark marginal distribution of productivities $f^z(z)$ is modeled as exponentially modified Gaussian (EMG) with the same parameters as in Section 4.

The benchmark distribution of the elasticity of taxable income is obtained from the survey of Lockwood, Sial, and Weinzierl (2021) who conduct a survey on the belief on the value of the uncompensated elasticity of taxable income among economists. The survey reveals the range of values from less than 0.1 to more than 2.0, indicating the long-standing substantial disagreement on the value of the labor supply elasticity. We use a discretized version of their histogram of survey responses and construct $f^{\gamma|z}(\gamma)$ which is reported in Figure 9.

Two penalty parameters θ_z and θ_γ are set as follows. For the exercise in this section, we set $\theta_z \to \infty$, and from part 2 of Lemma 5.1, any alternative distribution $\tilde{f}$ satisfies $\tilde{f}^z(z) = f^z(z)$. By doing so, we focus on studying concerns about labor supply responsiveness and isolate results from those in Section 4 that studied the consequences of distorting the productivity distribution.

[13] With the GHH prefernce, $1/\gamma$ is the elasticity of taxable income (ETI) if the marginal tax rate is constant. In general, the ETI also depends on the slope of the marginal tax schedule.

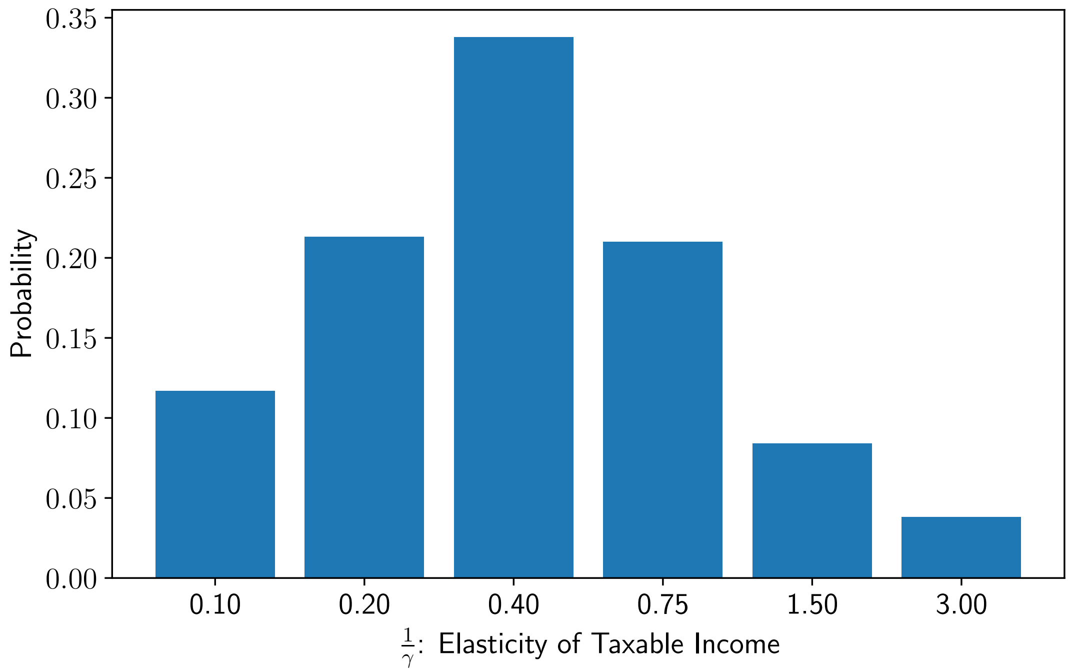

Figure 9: Benchmark distribution of elasticity of taxable income. We use a discretized distribution that is constructed using the histogram reported in Lockwood et al. (2021)

Given the remaining parameters, our model suggests that the worst-case joint density is skewed so that high-income individuals are more likely to have higher elasticity of taxable income, and higher average values of γ^{-1}. The degree of this distortion is influenced by the value of parameter θ_γ. We calibrate it to ensure that the worst-case model produces the elasticity for high-income earners reported in Rauh and Shyu (2024).[14] We also discuss results for higher and lower values of θ_γ surrounding our preferred estimate. Figure 10 plots the average of the γ^{-1} given productivity z under the planner's worst-case model against the mean of before-tax earnings y given productivity z. As discussed, in the worst-case model, the average value of the elasticity is higher than in the rational model (represented by the bold blue line) for higher-income earners. Rauh and Shyu (2024) estimates the elasticity of taxable income between 2.6 and 3.0 for taxpayers with an annual income of five million USD. Targeting those values, our calibration strategy implies $\theta_\gamma = 5$.

To compute the non-linear tax schedule, we restrict the space of tax functions. Characterizing a fully nonlinear tax function within a multidimensional private information framework using the Mirrleesian approach presents significant challenges (see, for example, Golosov and Krasikov (2023)). Instead, we search for a tax schedule T within a flexible

[14] Rauh and Shyu (2024) is one of the few studies that estimates labor supply elasticities for high-income earners with credible identification strategies. They use micro data from the state of California and study the labor supply response to Proposition 30, a 2012 measure that increased California marginal tax rates by up to 3 percentage.

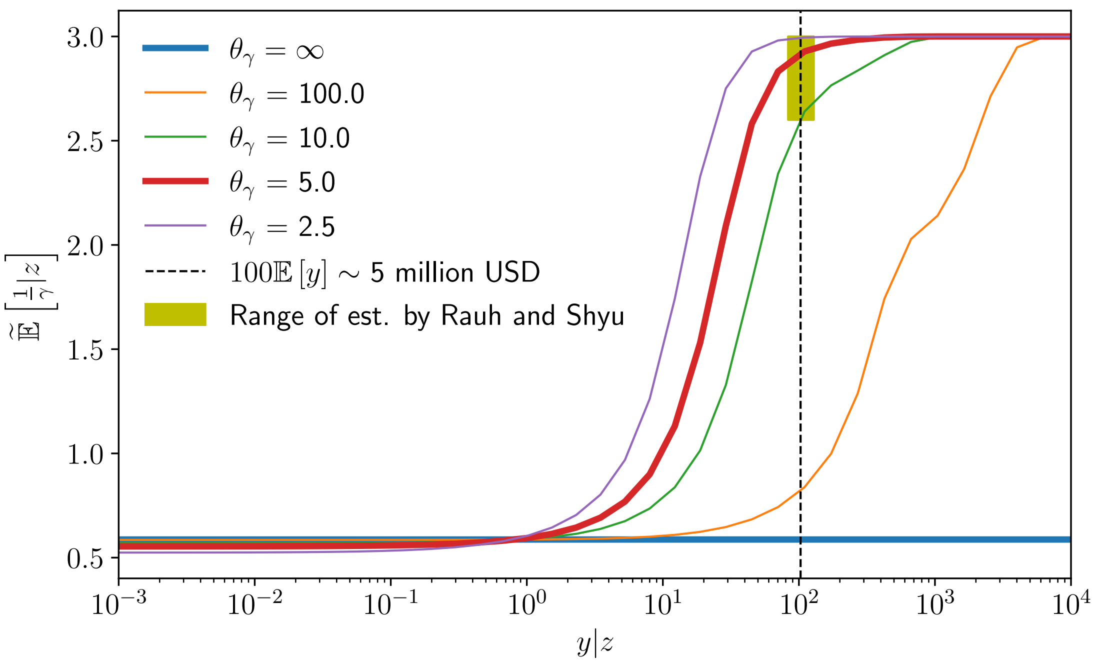

Figure 10: Mean elasticity of taxable income in worst-case models for alternative levels of multidimensional misspecification concerns

parametric class of functions such that the marginal tax rate $T'(y)$ is expressed through cubic basis functions of $\ln y$, featuring N knots $\{(\ln y_i, \tau_i)\}_{i=1,\cdots,N}$, where $T'(y)$ remains constant outside the range of the knots. We choose $N = 3$. In Appendix F, we discuss the trade-offs involved in choosing N, as well as benchmarking the spline approximation against the Mirlees solution for the one-dimensional case.

The remaining parameters are consistent with those outlined in Section 4. We set $T^0 = 0$ as a baseline and present results for alternative choices of the status-quo tax function in the Appendix G.

6.2 Results

Figure 11 plots the optimal marginal tax schedules for various values of θ_γ. Low values indicate higher level of concern about the conditional distributions of γ. In the figure, we observe that increased concerns result in lower marginal tax rates for high-income earners. At the calibrated value $\theta_\gamma = 5$, the top marginal tax rate $T'(y)$ at optimum is 14.0%, whereas in the rational model it is about 66.5%, leading to a substantial 52.5% decrease in the marginal tax rate at the top.

The economic mechanisms leading to lower tax rates in this context are analogous to those observed in the case of one-dimensional uncertainty regarding the productivity distribution, as discussed in Section 4. In that scenario, the robust planner was concerned about the shape of the right tail of the distribution and the ability to raise adequate revenues. As a result, the planner placed less emphasis on imposing distortions on high earners and reduced the top tax rate. In the situation examined here – where uncertainty pertains to the distribution of the elasticity of taxable income – the robust planner becomes wary that the costs of distorting high-income earners may be too large. To mitigate this risk, the planner adopts a cautious approach and lowers the top tax rate. In both cases, the economic mechanism is fundamentally similar, relying on the idea that tax revenues are concentrated, and concerns about misspecification arise from the possibility that those revenues may be difficult to secure.

Contrasting the worst-case distribution of $\gamma^{-1}|z$ with the benchmark further clarifies this intuition. Recall that in the benchmark distribution, we assumed that γ^{-1} was independently distributed with respect to productivities z. In contrast, the worst-case joint density features an endogenous positive correlation between productivities and elasticities. Interestingly, the resulting income distributions under the benchmark and the worst-case distributions do not differ significantly. As shown in Figure 12, the Pareto tail of the income distributions remains largely unaffected by the value of θ_γ. This occurs because income distributions are primarily driven by the distribution of productivities, and we have set $\theta_z = \infty$. Consequently, the robust planner preserves the marginal productivity distribution f^z, and the income distribution retains its core properties. A key takeaway from this example is that one does not need to rely on a thin-tailed worst-case income distribution to justify lower taxes.

In Table 3, we report additional moments and summarize how simulated outcomes vary across different values of θ_γ. As the penalty weight θ_γ decreases from 100 to 1, the mean of the elasticity under the worst-case scenario for highly productive workers $\widetilde{\mathbb{E}}\left[\gamma^{-1} \mid z \geq \bar{z}\right]$ also goes up. This concern is particularly evident among highly productive workers. Other outcomes such as average output, total transfers, and the marginal value of public goods remain largely unaffected.

7 Conclusion

The design of optimal tax schedules crucially depends on assumptions about the underlying distribution of types of taxed individuals. Despite affluently available data, the distribution of types over a fiscal planning horizon remains a complex task, especially in the tails where sample sizes are extremely small relative to the time-series variation.

In this paper, we tackle this uncertainty about the underlying type distribution by studying optimal taxation in the presence of concerns that the underlying distribution

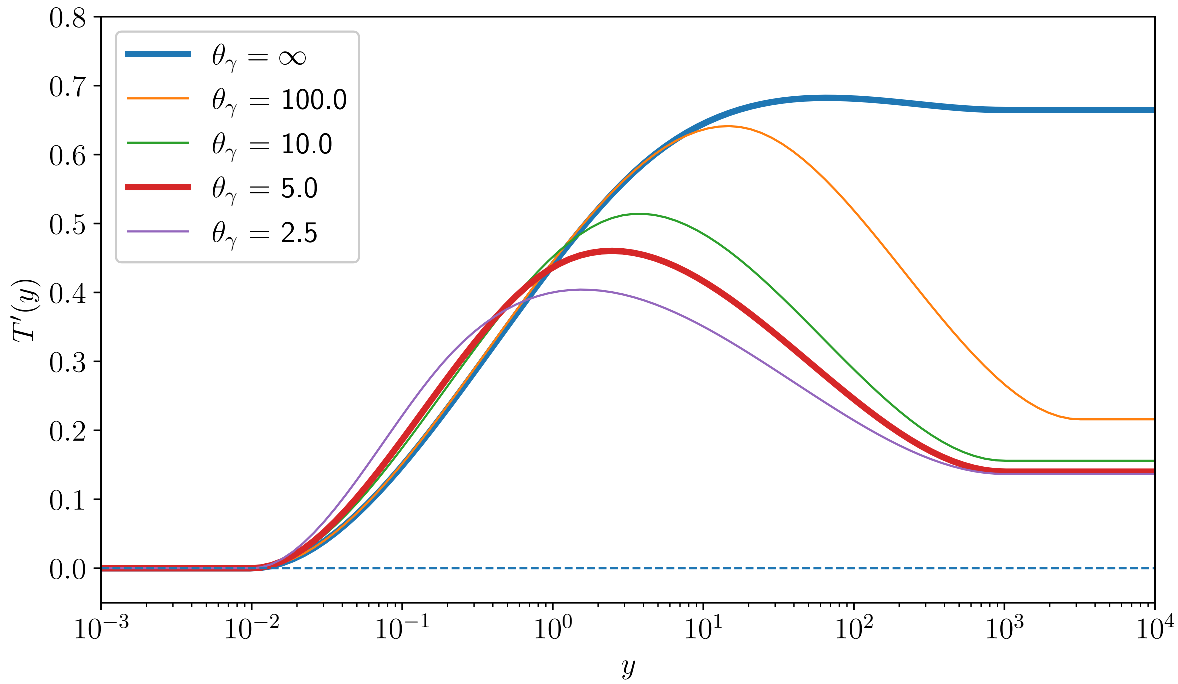

Figure 11: Optimal marginal tax schedules for alternative levels of multidimensional misspecification concerns. The status quo tax policy is no-tax-no-transfer, $T^0(y) = 0$.

moments $\backslash\theta_\gamma$	∞	100	10	5	2.5
$\widetilde{\mathbb{E}}[z]$	1.00	1.00	1.00	1.00	1.00
$\widetilde{\mathbb{E}}[\gamma^{-1}]$	0.59	0.59	0.59	0.59	0.60
$\mathbb{E}[\gamma^{-1} \mid z \geq \bar{z}]$	0.59	0.59	0.59	0.59	0.59
$\widetilde{\mathbb{E}}[\gamma^{-1} \mid z \geq \bar{z}]$	0.59	1.12	2.75	2.96	3.00
$\mathbb{E}[y]$	1.03	1.07	1.12	1.15	1.18
$\widetilde{\mathbb{E}}[y]$	1.03	1.01	1.00	1.00	0.99
μ	1.37	1.44	1.55	1.61	1.68
T_0	-0.42	-0.39	-0.36	-0.34	-0.32
$\mathbb{E}[T]$	-0.00	0.02	0.04	0.04	0.05
$\widetilde{\mathbb{E}}[T]$	-0.00	-0.00	-0.00	0.00	-0.00
$\mathbb{E}[T]/\mathbb{E}[y]$	-0.00	0.02	0.03	0.04	0.04
$\widetilde{\mathbb{E}}[T]/\widetilde{\mathbb{E}}[y]$	-0.00	-0.00	-0.00	0.00	-0.00

Table 3: Moments under the benchmark and worst-case. The productivity level $\bar{z}$ is minimum z such that $y \mid z \geq 100\mathbb{E}[y]$ given the benchmark distribution with $T^{0\prime} = 0$.

is misspecified. The robust approach we employ allows us to avoid making parametric assumptions about the structure of uncertainty, and instead focus on nonparametric misspecifications of the underlying type distribution that the planner fears to be most consequential for the welfare function.

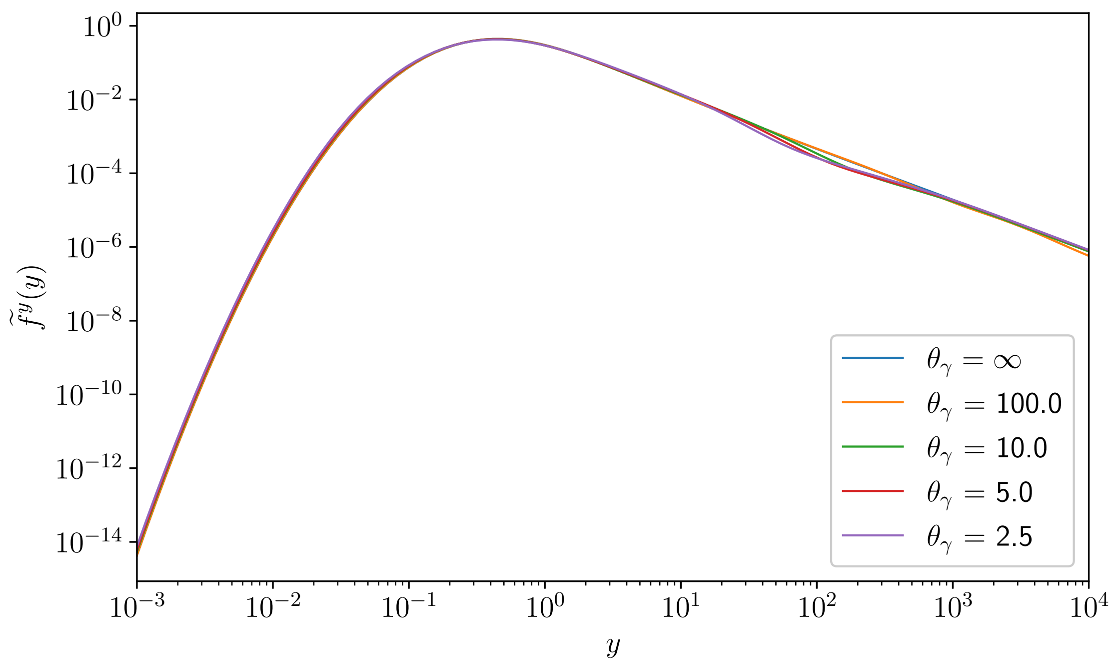

Figure 12: Worst-case distributions of before-tax income $\tilde{f}^y(y)$ for alternative levels of multidimensional misspecification concerns. The status quo tax policy is no-tax-no-transfer, $T^0(y) = 0$.

The uncertainty about the distribution of productivity leads to a stark difference in the optimal tax schedule. We find that marginal taxes should optimally decline to zero, even when the underlying benchmark distribution has a fat right tail, like in the Pareto case. Misspecification concerns do not only have a theoretical asymptotic impact, they decrease marginal tax rates of high-income individuals in quantitatively important ways, against a modest decrease in redistribution toward low-productivity workers.

We further study the implications of uncertainty about the elasticity of taxable income in designing the optimal tax schedule reform. When the robust planner faces additional concerns about the elasticity of taxable income, the worst-case scenario features a positive correlation between the elasticity and productivity: the revenue is disproportionately collected from high-income earners, and the tax reform may fail to collect extra revenue from the high-income earners due to potential high responsiveness to the tax reform. Anticipating this risk, the planner optimally reduces an increase in the marginal tax rates for high-income earners, even without concerns about the distribution of productivity.

Important directions are left for follow-up work. A natural candidate involves a dynamic extension where optimal taxation of heterogeneous workers is combined with an optimal debt management problem. Another prominent application in which uncertainty about the underlying distribution is substantial involves taxation of wealth. Insights developed in this work will help analyze such cases as well.

Appendix

A Application of the minimax theorem

In this appendix, we show the validity of exchanging the order of optimization in a version of the planner's problem from Section 2.1. We focus on the discrete-type formulation with I types, equal Pareto weights, separable isoelastic household utility, and linear utility from government expenditures

$$\max_{\{c_i, y_i\}_{i=1}^I \in \mathbb{X}} \min_{\{m_i\}_{i=1}^I \in \mathbb{Y}} \sum_{i=1}^I \pi_i m_i \left(u\left(c_i\right) - v\left(\frac{y_i}{z_i}\right)\right) + \bar{v} G + \theta \sum_{i=1}^I \pi_i m_i \log m_i, \tag{42}$$

subject to the incentive compatibility constraints

$$u\left(c_i\right) - v\left(\frac{y_i}{z_i}\right) \geq u\left(c_j\right) - v\left(\frac{y_j}{z_i}\right), \qquad i, j \in \{1, \cdots, I\},$$

and Radon–Nikodým derivative constraints

$$\sum_{i=1}^I \pi_i m_i = 1, \qquad m_i \geq 0, \qquad i \in \{1, \ldots, I\}, \tag{43}$$

where $\{\pi_i\}_{i=1}^I$ is the benchmark probability distribution, and government expenditures G are equal to the net tax revenue

$$G = \sum_{i=1}^I \pi_i m_i \left(y_i - c_i\right).$$

The functions $u\left(c\right)$ and $v\left(y/z\right)$ have the isoelastic form from (27). To apply a version of the minimax theorem (Sion (1958)), we need to establish suitable compact and convex subsets $\mathbb{X}$ and $\mathbb{Y}$ of $\mathbb{R}^I$ for the choice variables, convexity in the minimizing variables, and concavity in the maximizing variables.

It is convenient to write the problem as a function of maximizing controls u_i and v_i representing utility from consumption and disutility from labor, respectively:

$$\max_{\{u_i, v_i\}_{i=1}^I \in \mathbb{X}} \min_{\{m_i\}_{i=1}^I \in \mathbb{Y}} \sum_{i=1}^I \pi_i m_i \left(u_i - v_i + \bar{v} z_i v^{-1}\left(v_i\right) - \bar{v} u^{-1}\left(u_i\right)\right) + \theta \sum_{i=1}^I \pi_i m_i \log m_i$$

subject to linear constraints

$$u_i - v_i \geq u_j - \left(\frac{z_j}{z_i}\right)^{1+\gamma} v_j, \qquad i, j \in \{1, \cdots, I\}, \tag{44}$$

and the constrains in (43). This problem is strictly concave in $\{u_i, v_i\}_{i=1}^I$ and strictly convex in $\{m_i\}_{i=1}^I$, on a convex set $\mathbb{X}$ imposed by the linear constraints (44), and on a convex set $\mathbb{Y}$ imposed by linear constraints on m_i in (43). It remains to bound $\mathbb{X}$ and $\mathbb{Y}$ to make them compact.

Observe that $u_i \in \mathbb{U}$ where $\mathbb{U} = \left[-\left(1-\rho\right)^{-1}, \infty\right)$ for $\rho < 1$, $u_i \in \left(-\infty, \infty\right)$ for $\rho = 1$ and

$\mathbb{U} = \left(-\infty, -(1-\rho)^{-1}\right)$ for $\rho > 1$. Further, $v_i \in \mathbb{V} = [0, \infty)$. Denote

$$U_i(u_i, v_i) = u_i - \bar{v}u^{-1}(u_i) + \bar{v}z_i v^{-1}(v_i) - v_i,$$

and notice that $U_i(u_i, v_i)$ is bounded from above, strictly concave, and, whenever u_i converges to one of the open boundaries of $\mathbb{U}$ or v_i converges to the open boundary of $\mathbb{V}$, we have $\lim U_i(u_i, v_i) = -\infty$. Therefore, for any $\underline{U} \in \mathbb{R}$, the upper countour set

$$\overline{\mathbb{W}}_i = \left\{(u_i, v_i) \in \overline{\mathbb{U}} \times \overline{\mathbb{V}} : U_i(u_i, v_i) \geq \underline{U}\right\}$$

is convex and compact. We construct the compact convex set $\mathbb{X}$ as $\mathbb{X} = \overline{\mathbb{W}}_1 \times \ldots \times \overline{\mathbb{W}}_I$, for a suitable choice of $\underline{U}$.

Fix a $\widetilde{c} > 0$ and utility allocation $u_i = u(\widetilde{c})$, $v_i = v(0) = 0$. This allocation satisfies all constraints in (44), and hence is feasible. Further denote

$$\begin{aligned}
\widetilde{U} &= \sum_{i=1}^{I} \pi_i m_i U_i(u(\widetilde{c}), v(0)) = u(\widetilde{c}) - \bar{v}\widetilde{c} \\
\overline{U} &= \max_i \max_{u_i, v_i} U_i(u_i, v_i).
\end{aligned}$$

Then set

$$\underline{U} = \left(\min_i \pi_i\right)^{-1} \left(\widetilde{U} - \varepsilon - \left(1 - \min_i \pi_i\right)\overline{U}\right), \qquad \varepsilon > 0.$$

We also note that for a given allocation $\{u_i, v_i\}_{i=1}^{I}$, the optimal distortion, the discrete-type counterpart to (13)), can be written as

$$m_i = \bar{m} \exp\left(-\frac{1}{\theta} U_i(u_i, v_i)\right), \tag{45}$$

i.e., it is decreasing in $U_i(u_i, v_i)$, and $m_i = 1$ for the feasible allocation $\{u(\widetilde{c}), v(0)\}_{i=1}^{I}$, since $U_i(u(\widetilde{c}), v(0)) = \widetilde{U}$.

Take any allocation $\{u_i, v_i\}_{i=1}^{I} \notin \mathbb{X}$ and, without loss of generality, order indices i by increasing $U_i(u_i, v_i)$. Then first $J \geq 1$ indices are such that $U_i(u_i, v_i) \leq \underline{U}$ and hence $(u_i, v_i) \notin \overline{\mathbb{W}}_i$. Further, the associated m_i given by (45) are decreasing in i, and the distribution $\{\hat{\pi}_i\}_{i=1}^{J}$ given by $\hat{\pi}_i = \pi_i m_i$ is first-order stochastically dominated by $\{\pi_i\}_{i=1}^{I}$. Then

$$\begin{aligned}
\sum_{i=1}^{I} \pi_i m_i U_i(u_i, v_i) &\leq \sum_{i=1}^{I} \pi_i U_i(u_i, v_i) = \sum_{i \leq J} \pi_i U_i(u_i, v_i) + \sum_{i > J} \pi_i U_i(u_i, v_i) \\
&\leq \sum_{i \leq J} \pi_i \underline{U} + \sum_{i > J} \pi_i \overline{U} \leq \min_i \pi_i \underline{U} + \left(1 - \min_i \pi_i\right)\overline{U} = \widetilde{U} - \varepsilon < \widetilde{U}.
\end{aligned}$$

The first inequality follows from the fact that $\{\hat{\pi}_i\}_{i=1}^{J}$ is first-order stochastically dominated by $\{\pi_i\}_{i=1}^{I}$, the second inequality from the fact that for $i \leq J$, $U_i(u_i, v_i) < \underline{U}$ and for any i, $U_i(u_i, v_i) \leq \overline{U}$, and the third inequality from the fact that $J \geq 1$ and $\underline{U} \leq \overline{U}$. The allocation $\{u_i, v_i\}_{i=1}^{I} \notin \mathbb{X}$ thus cannot be optimal, and we can exclude it from the optimization problem.

Finally, since $\mathbb{X}$ bounds all $U_i(u_i, v_i)$ from below by $\underline{U}$ and from above by $\overline{U}$, all m_i given by (45) are bounded away from zero. Since m_i are also bounded from above by $(\min_i \pi_i)^{-1}$, there exists a $\delta > 0$ such that we can restrict the set $\mathbb{Y}$ to the compact convex set

$$\mathbb{Y} = \left\{ (m_1, \ldots, m_I) : m_i \in \left[\delta, \left(\min_i \pi_i \right)^{-1} \right] \right\}.$$

This completes the verification of conditions under which the order of optimization in (42) can be exchanged.

B Derivation of optimal tax formulas

In this appendix, we derive optimal tax formulas from the optimization problem

$$\max_{c,y} \min_{\substack{m>0 \\ \mathbb{E}[m]=1}} \int_{\underline{z}}^{\bar{z}} \psi(z) U\left(c(z), \frac{y(z)}{z}\right) m(z) f(z) dz + \theta \int_{\underline{z}}^{\bar{z}} m(z) \log m(z) f(z) dz$$

subject to the IC constraint

$$\frac{dU}{dz} = -U_n\left(c(z), \frac{y(z)}{z}\right) \frac{y(z)}{z^2}. \tag{46}$$

and the budget constraint

$$G = \int_{\underline{z}}^{\bar{z}} (y(z) - c(z)) m(z) f(z) dz. \tag{47}$$

Treating indirect utility $\mathcal{U}$ as the state variable, λ as its co-state, and y and m as control variables, we can form the constrained Hamiltonian

$$\begin{aligned} H(\mathcal{U}, y, m, \lambda) =& \psi(z) \mathcal{U}(z) m(z) f(z) + \theta m(z) \log m(z) f(z) - \chi m(z) f(z) \\ & - \lambda(z) U_n\left(c(z), \frac{y(z)}{z}\right) \frac{y(z)}{z^2} + \mu [y(z) - c(z)] m(z) f(z). \end{aligned}$$

Here, χ and μ are multipliers on the constraints $\mathbb{E}[m] = 1$ and (47), respectively, and $c(z)$ is defined implicitly from the definition of the utility function

$$\mathcal{U}(z) = U\left(c(z), \frac{y(z)}{z}\right)$$

as $c(z) = C(\mathcal{U}(z), y(z))$. The optimality condition with respect to output choice $y(z)$ is

$$\begin{aligned} 0 = H_y =& \mu \left[1 - C_y(U(z), y(z))\right] m(z) f(z) - \lambda(z) U_n\left(C(\mathcal{U}(z), y(z)), \frac{y(z)}{z}\right) \frac{1}{z^2} \\ & - \lambda(z) \left[U_{nc}\left(C(\mathcal{U}(z), y(z)), \frac{y(z)}{z}\right) C_y(\mathcal{U}(z), y(z)) + U_{nn}\left(C(\mathcal{U}(z), y(z)), \frac{y(z)}{z}\right) \frac{1}{z} \right] \frac{y(z)}{z^2}, \end{aligned} \tag{48}$$

with respect to the distortion $m(z)$ is

$$0 = H_m = \psi(z) \mathcal{U}(z) f(z) + \theta [\log m(z) + 1] f(z) - \chi f(z) + \mu [y(z) - C(\mathcal{U}(z), y(z))] f(z), \tag{49}$$

and the costate dynamics restriction yields

$$\frac{d\lambda(z)}{dz} = - H_U = [\mu C_U(\mathcal{U}(z), y(z)) - \psi(z)]\, m(z) f(z) + \\ + \lambda(z) U_{nc}\left(C(\mathcal{U}(z), y(z)), \frac{y(z)}{z}\right) \frac{y(z)}{z^2} C_U(\mathcal{U}(z), y(z)). \tag{50}$$

The transversality condition is

$$\lambda(\bar{z}) \mathcal{U}(\bar{z}) = 0,$$

and since the problem is unrestricted at the left end of the type distribution (the choice $U(\underline{z})$ is unrestricted), we also have $\lambda(\underline{z}) = 0$.

Condition (49) can be used to express the distortion $m(z)$:

$$m(z) = \exp\left(\frac{\chi}{\theta} - 1\right) \exp\left(-\frac{1}{\theta}\left(\psi(z)\mathcal{U}(z) + \mu\left[y(z) - C(\mathcal{U}(z), y(z))\right]\right)\right).$$

The Lagrange multiplier χ is solved for from the restriction $\mathbb{E}[m] = 1$ as a normalization constant, which then shows up as $\bar{m} = \exp(\chi/\theta - 1)$ in (13). Integrating up condition (50) over the range of types $(\underline{z}, \bar{z})$,

$$\int_{\underline{z}}^{\bar{z}} \frac{d\lambda(z)}{dz} dz = \lambda(\bar{z}) - \lambda(\underline{z}) = 0 = \mu \int_{\underline{z}}^{\bar{z}} C_U(\mathcal{U}(z), y(z))\, m(z) f(z)\, dz - \int_{\underline{z}}^{\bar{z}} \psi(z) m(z) f(z)\, dz \\ + \int_{\underline{z}}^{\bar{z}} \lambda(z) U_{nc}\left(C(\mathcal{U}(z), y(z)), \frac{y(z)}{z}\right) \frac{y(z)}{z^2} C_U(\mathcal{U}(z), y(z))\, dz,$$

we solve for the Lagrange multiplier μ:

$$\mu = \frac{\int_{\underline{z}}^{\bar{z}} \psi(z) m(z) f(z)\, dz - \int_{\underline{z}}^{\bar{z}} \lambda(z) U_{nc}\left(C(\mathcal{U}(z), y(z)), \frac{y(z)}{z}\right) \frac{y(z)}{z^2} C_U(\mathcal{U}(z), y(z))\, dz}{\int_{\underline{z}}^{\bar{z}} C_U(\mathcal{U}(z), y(z))\, m(z) f(z)\, dz}. \tag{51}$$

This Lagrange multiplier represents the marginal social value of public funds to the planner.

With an expression for μ at hand, we solve for $\lambda(z)$ forward by integrating (50) on $(z, \bar{z})$. For that purpose, denote the terms in (50) as follows:

$$H_{U,S}(z) = \mu C_U(\mathcal{U}(z), y(z)) f(z) - \psi(z) m(z) f(z)$$
$$H_{U,N}(z) = U_{nc}\left(C(\mathcal{U}(z), y(z)), \frac{y(z)}{z}\right) \frac{y(z)}{z^2} C_U(\mathcal{U}(z), y(z)).$$

The term $H_{U,N}(z)$ is only present when preferences $U(c, n)$ are non-separable. Then rewrite (50) as

$$\lambda'(z) = H_{U,S}(z) + \lambda(z) H_{U,N}(z).$$

This equation has the solution

$$\lambda(z) = - \int_z^{\bar{z}} H_{U,S}(\zeta) \exp\left(- \int_z^{\zeta} H_{U,N}(\xi)\, d\xi\right) d\zeta.$$

To simplify notation, let us simplify the arguments of functions above, and write, for example,

$$U_{nc}(z) = U_{nc}\left(C\left(\mathcal{U}(z), y(z)\right), \frac{y(z)}{z}\right).$$

Further, we utilize the following notational simplifications:

$$\begin{aligned} \left(1 - T'(y(z))\right) z &= -\frac{U_n(z)}{U_c(z)} \\ C_U\left(\mathcal{U}(z), y(z)\right) &= \frac{1}{U_c(z)} \\ C_y\left(\mathcal{U}(z), y(z)\right) &= -\frac{1}{z}\frac{U_n(z)}{U_c(z)}. \end{aligned} \tag{52}$$

The first equation is the worker's individual optimality condition for the choice of labor supply and consumption given a particular tax schedule. The last two follow from the definition of the implicit function $C\left(\mathcal{U}(z), y(z)\right)$. With this notation and substitutions, we can reorganize the optimality condition (48) as

$$-\mu\left[T'(y(z))\right] f(z) = \frac{\lambda(z)}{z} U_c(z)\left(1 - T'(y(z))\right)\left[1 + n(z)\frac{U_{nn}(z) + w(z)U_{nc}(z)}{U_n(z)}\right]. \tag{53}$$

It is convenient to rewrite this expression in terms of labor supply elasticities. Denoting $w = \left(1 - T'(zn)\right) z$ the after-tax marginal compensation for providing labor, and rewriting the after-tax income as

$$c = zn - T(zn) = wn + I$$

to separate the marginal effect w and lump-sum effect I of the tax schedule $T(zn)$, we write the optimality condition (52) as

$$w = -\frac{U_n(wn + I, n)}{U_c(wn + I, n)}.$$

We can then implicitly differentiate to derive the uncompensated and compensated labor supply elasticities as

$$\begin{aligned} \varepsilon^u &= \frac{dn}{dw}\frac{w}{n} = \frac{-\frac{U_n}{n} + w^2 U_{cc} + wU_{nc}}{-U_{nn} - w^2 U_{cc} - 2wU_{cn}} \\ \varepsilon^c &= \left(\frac{dn}{dw} - n\frac{dn}{dI}\right)\frac{w}{n} = \varepsilon^u - \frac{dn}{dI}w = \frac{-\frac{U_n}{n}}{-U_{nn} - w^2 U_{cc} - 2wU_{cn}}. \end{aligned}$$

which then yields

$$\frac{1 + \varepsilon^u}{\varepsilon^c} = 1 + n\frac{U_{nn} + wU_{cn}}{U_n}.$$

Using this result in expression (53), we obtain

$$\frac{T'(y(z))}{1 - T'(y(z))} = -\frac{1 + \varepsilon^u(z)}{\varepsilon^c(z)}\frac{\lambda(z)}{zm(z)f(z)}\frac{U_c(z)}{\mu}. \tag{54}$$

To solve for the Lagrange multiplier μ, simplify equation (51), which yields

$$\mu = \frac{\int_{\underline{z}}^{\bar{z}} \left[\psi(\zeta) - \frac{\lambda(\zeta)}{zm(\zeta)f(\zeta)} \frac{n(\zeta)U_{nc}(\zeta)}{U_c(\zeta)}\right] m(\zeta) f(\zeta) \, d\zeta}{\int_{\underline{z}}^{\bar{z}} \frac{1}{U_c(\zeta)} m(\zeta) f(\zeta) \, d\zeta}.$$

Finally, the expression for $\lambda(z)$ can be simplified as

$$\begin{aligned} H_{U,S}(z) &= \left[\frac{\mu}{U_c(z)} - \psi(z)\right] m(z) f(z) \\ H_{U,N}(z) &= \frac{n(z) U_{nc}(z)}{z U_c(z)} \\ \lambda(z) &= -\int_z^{\bar{z}} H_{U,S}(\zeta) \exp\left(-\int_z^{\zeta} H_{U,N}(\xi) \, d\xi\right) d\zeta. \end{aligned}$$

For the case of isoelastic separable preferences (27), $U_{nc}(z) = 0$, $H_{U,N}(z) = 0$, and $(1+\varepsilon^u)/\varepsilon^c = 1+\gamma$. This yields the marginal social value of public funds in the form of inverse marginal utility formula evaluated under the worst-case distribution

$$\mu = \frac{\int_{\underline{z}}^{\bar{z}} \psi(\zeta) m(\zeta) f(\zeta) \, d\zeta}{\int_{\underline{z}}^{\bar{z}} (c(\zeta))^{\rho} m(\zeta) f(\zeta) \, d\zeta}.$$

Consequently,

$$\begin{aligned} \lambda(z) &= \int_z^{\bar{z}} \psi(\zeta) m(\zeta) f(\zeta) \, d\zeta - \mu \int_z^{\bar{z}} (c(\zeta))^{\rho} m(\zeta) f(\zeta) \, d\zeta && (55) \\ &= \int_{\underline{z}}^{\bar{z}} \psi(\zeta) m(\zeta) f(\zeta) \, d\zeta \left[\widetilde{F}_{\rho}(z) - \widetilde{\Psi}(z)\right] && (56) \end{aligned}$$

with

$$\widetilde{F}_{\rho}(z) = \int_{\underline{z}}^{z} \frac{(c(\zeta))^{\rho} m(\zeta) f(\zeta)}{\int_{\underline{z}}^{\bar{z}} (c(\xi))^{\rho} m(\xi) f(\xi) \, d\xi} d\zeta \qquad \widetilde{\Psi}(z) = \int_{\underline{z}}^{z} \frac{\psi(\zeta) m(\zeta) f(\zeta)}{\int_{\underline{z}}^{\bar{z}} \psi(\xi) m(\xi) f(\xi) \, d\xi} d\zeta,$$

where $\widetilde{F}_{\rho}(z)$ is the inverse marginal utility weighted worst-case distribution, and $\widetilde{\Psi}(z)$ is the welfare weighted worst-case distribution. The tax formula then simplifies to

$$\frac{T'(y(z))}{1 - T'(y(z))} = (1+\gamma) \frac{\widetilde{\Psi}(z) - \widetilde{F}_{\rho}(z)}{1 - \widetilde{F}_{\rho}(z)} \frac{1 - \widetilde{F}_{\rho}(z)}{z \widetilde{f}_{\rho}(z)}.$$

For the case of quasilinear utility (11), we have $\widetilde{F}_{\rho}(z) = \widetilde{F}(z)$, and obtain expression (14). For concave preferences with zero welfare concern at the top as in Section 3.4.1, we have $\widetilde{\Psi}(z) = 1$, which yields expression (28). Expression (32) from Section 3.4.2 is obtained by setting $\psi(z) \equiv 1$.

C Proofs for the baseline model

Throughout this section, we restrict our attention to the analysis of the baseline model, analyzed in Sections 3.1–3.3. In particular, we assume the quasilinear utility function

$$U(c,n) = c - \frac{n^{1+\gamma}}{1+\gamma},$$

an unbounded type space, $\bar{z} = \infty$, and restrict our attention to the characterization of the optimal marginal tax on the interval $[\hat{z}, \bar{z})$ on which the planner has no utilitarian concerns, $\Psi(\hat{z}) = \widetilde{\Psi}(\hat{z}) = 1$, and for which Assumption 2 holds. In the proof of Theorem 3.2, we additionally assume that the shape of the type distribution for $z \geq \hat{z}$ under the benchmark model is proportional to the Pareto distribution.

C.1 Proof of Theorem 3.1

Before proving Theorem 3.1, we start with three preliminary lemmas.

Lemma C.1. *Optimal output chosen by individual workers under a given tax scheme $T(y)$ satisfies:*

$$y'(z) = \frac{(1+\gamma)\frac{y(z)}{z}}{\gamma + \frac{T''(y(z))}{1-T'(y(z))}y(z)}. \tag{57}$$

Proof. Under the given tax scheme $T(y)$, the optimality condition (1) of a worker of type z for the special case of quasilinear preferences (11) implies labor supply $n(z)$ given by

$$\left(1 - T'(zn(z))\right)z = n(z)^{\gamma}$$

which can be rewritten as

$$y(z) = \left(1 - T'(y(z))\right)^{\frac{1}{\gamma}} z^{\frac{1+\gamma}{\gamma}}. \tag{58}$$

Differentiating with respect to z:

$$\begin{aligned} y'(z) &= -\frac{1}{\gamma}\left(1 - T'(y(z))\right)^{\frac{1-\gamma}{\gamma}} z^{\frac{1+\gamma}{\gamma}} T''(y(z))\, y'(z) + \frac{1+\gamma}{\gamma}\left(1 - T'(y(z))\right)^{\frac{1}{\gamma}} z^{\frac{1}{\gamma}} \\ &= -\frac{1}{\gamma}\frac{T''(y(z))}{1-T'(y(z))}y(z)\,y'(z) + \frac{1+\gamma}{\gamma}\frac{y(z)}{z} \end{aligned}$$

yields

$$\left[1 + \frac{1}{\gamma}\frac{T''(y(z))}{1-T'(y(z))}y(z)\right]y'(z) = \frac{1+\gamma}{\gamma}\frac{y(z)}{z}$$

and hence

$$y'(z) = \frac{(1+\gamma)\frac{y(z)}{z}}{\gamma + \frac{T''(y(z))}{1-T'(y(z))}y(z)}.$$

■

Lemma C.2. *Let $f(z)$ and $\widetilde{f}_i(z)$, $i = 1,2$ be density functions related by $\widetilde{f}_i(z) = m_i(z) f(z)$ where $m_i(z)$ are strictly positive functions representing changes of measure, and $m_1(z)/m_2(z)$ is strictly decreasing. Then*

$$\frac{1-\widetilde{F}_1(z)}{z\widetilde{f}_1(z)} < \frac{1-\widetilde{F}_2(z)}{z\widetilde{f}_2(z)}.$$

Proof. The expression can be rewritten as

$$\frac{1-\widetilde{F}_1(z)}{z\widetilde{f}_1(z)} = \frac{\int_z^{\overline{z}} m_1(\zeta) f(\zeta)\, d\zeta}{z m_1(z) f(z)} = \frac{\int_z^{\overline{z}} \frac{m_1(\zeta)}{m_1(z)} f(\zeta)\, d\zeta}{z f(z)} \tag{59}$$

$$< \frac{\int_z^{\overline{z}} \frac{m_2(\zeta)}{m_2(z)} f(\zeta)\, d\zeta}{z f(z)} = \frac{\int_z^{\overline{z}} m_2(\zeta) f(\zeta)\, d\zeta}{z m_2(z) f(z)} = \frac{1-\widetilde{F}_2(z)}{z\widetilde{f}_2(z)}, \tag{60}$$

where the inequality comes from the fact that for any $\zeta > z$,

$$\frac{m_1(\zeta)}{m_2(\zeta)} < \frac{m_1(z)}{m_2(z)}.$$

■

Lemma C.3. *Define*

$$\widetilde{\varphi}(z) = \frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)} = \frac{1}{1+\gamma}\frac{T'(y(z))}{1-T'(y(z))}.$$

Then

$$\widetilde{\varphi}'(z) = \frac{d}{dz}\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)} = -\frac{1}{z} - \widetilde{\varphi}(z)\left[\frac{1}{z} - \frac{\mu}{\theta} T'(y(z))\, y'(z) + \frac{f'(z)}{f(z)}\right].$$

Proof. By direct computation:

$$\frac{d}{dz}\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)} = \frac{-\widetilde{f}(z) z\widetilde{f}(z) - \left(1-\widetilde{F}(z)\right)\left(\widetilde{f}(z) + z\widetilde{f}'(z)\right)}{\left(z\widetilde{f}(z)\right)^2}$$

$$= -\frac{1}{z} - \frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)}\left[\frac{1}{z} + \frac{\widetilde{f}'(z)}{\widetilde{f}(z)}\right] = -\frac{1}{z} - \widetilde{\varphi}(z)\left[\frac{1}{z} + \frac{d}{dz}\log\widetilde{f}(z)\right].$$

Using $\widetilde{f}(z)$ from expression (17), we obtain the last line of the lemma. ■

Proof of Theorem 3.1. We restrict attention to $z \geq \hat{z}$ for which the planner's welfare weight is zero, $\psi(z) = 0$, and for which Assumption 2 holds. In this case, the tax formula (15) is given by

$$\frac{T'(y(z))}{1-T'(y(z))} = (1+\gamma)\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)} = (1+\gamma)\widetilde{\varphi}(z). \tag{61}$$

The single-crossing property (9) implies that $y(z)$ is strictly increasing in z. Since the tax rate $T'(y(z))$ is strictly positive, the tax function $T(y(z))$ strictly increases in z. This means that $m(z)$ in (17) is strictly decreasing, and, by Assumption 2, $\widetilde{f}(z)$ is also strictly decreasing for sufficiently large z. Lemma C.2 then implies that

$$\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)} < \frac{1-F(z)}{zf(z)}, \tag{62}$$

and hence for all $z > \hat{z}$, the marginal tax rate $T'(y(z))$ must be strictly lower than the marginal tax rate $T'_{rat}(y_{rat}(z))$ in the model without model misspecification.

An application of L'Hôpital's rule to the tax rate under the model without model misspecification implies

$$\lim_{z\to\infty}\frac{T'_{rat}(y(z))}{1-T'_{rat}(y(z))} = (1+\gamma)\lim_{z\to\infty}\frac{1-F(z)}{zf(z)} = \lim_{z\to\infty}\frac{1}{-\frac{d\log f(z)}{d\log z}-1} < \infty$$

by Assumption 2. This yields $\lim_{z\to\infty} T'_{rat}(y(z)) < 1$, and combined with $T'(y(z)) < T'_{rat}(y_{rat}(z))$ implied by (62), we obtain that for sufficiently large z, $T'(y(z))$ must be bounded away from one, $T'(y(z)) < 1-\bar{\varepsilon}_T$, and $\widetilde{\phi}(z)$ is bounded,

$$\widetilde{\phi}(z) = \frac{1}{1+\gamma}\frac{T'(y(z))}{1-T'(y(z))} < \frac{1}{1+\gamma}\frac{1-\bar{\varepsilon}_T}{\bar{\varepsilon}_T} = K_\phi.$$

Consequently, the optimal allocation formula (73) implies that $\lim_{z\to\infty} y(z) = \infty$.

From the optimal tax formula (61), we obtain $1-T'(y(z)) = \left(1+(1+\gamma)\widetilde{\phi}(z)\right)^{-1}$, and then we can rewrite expression (73) as

$$y(z) = \left(1+(1+\gamma)\widetilde{\phi}(z)\right)^{-\frac{1}{\gamma}} z^{\frac{1+\gamma}{\gamma}}.$$

Differentiating this expression with respect to z yields

$$\begin{aligned} y'(z) &= -\frac{1}{\gamma}\frac{(1+\gamma)\widetilde{\phi}'(z)}{\left(1+(1+\gamma)\widetilde{\phi}(z)\right)^{\frac{1+\gamma}{\gamma}}} z^{\frac{1+\gamma}{\gamma}} + \frac{1+\gamma}{\gamma}\frac{1}{\left(1+(1+\gamma)\widetilde{\phi}(z)\right)^{\frac{1}{\gamma}}} z^{\frac{1}{\gamma}} \\ &= \frac{1+\gamma}{\gamma}\frac{1}{\left(1+(1+\gamma)\widetilde{\phi}(z)\right)^{\frac{1+\gamma}{\gamma}}} z^{\frac{1}{\gamma}}\left[2+\widetilde{\phi}(z)\left(2+\gamma-\frac{\mu}{\theta}zT'(y(z))\,y'(z)+z\frac{f'(z)}{f(z)}\right)\right], \end{aligned}$$

where the second line uses Lemma C.3 to substitute in for $\widetilde{\phi}'(z)$. Since $y'(z) > 0$, the last bracket must be strictly positive.

We now show that $\lim_{z\to\infty} T'(y(z)) = 0$. Assume that it is not. Then there exists $\underline{\varepsilon}_T > 0$ such that for an arbitrarily large $\bar{z}$ there exists a $z > \bar{z}$ such that $T'(y(z)) > \underline{\varepsilon}_T$. Take such a z for which $T'(y(z)) > \underline{\varepsilon}_T$. Since $f'(z) < 0$, we have

$$0 < 2+\widetilde{\phi}(z)\left(2+\gamma-\frac{\mu}{\theta}zT'(y(z))\,y'(z)+z\frac{f'(z)}{f(z)}\right) < 2+K_\phi\left(2+\gamma-\frac{\mu}{\theta}zy'(z)\,\underline{\varepsilon}_T\right),$$

which yields a bound on $zy'(z)$:

$$zy'(z) < \frac{\theta}{\varepsilon_T \mu}\left(2+\gamma+2K_\phi^{-1}\right) = K_y. \tag{63}$$

Since, from the result in Lemma C.1, we have that

$$zy'(z) = \frac{(1+\gamma)\,y(z)}{\gamma + \frac{T''(y(z))}{1-T'(y(z))}y(z)} < K_y,$$

we can derive a restriction on $T''(y(z))$:

$$T''(y(z)) > \left(1-T'(y(z))\right)\frac{(1+\gamma)\,y(z)-K_y\gamma}{K_y y(z)} > \bar{\varepsilon}_T \frac{(1+\gamma)\,y(z)-K_y\gamma}{K_y y(z)}. \tag{64}$$

Recall that we can find an arbitrarily large z for which this inequality holds. Since $\lim_{z\to\infty} y(z) = \infty$, we can find such a z that is sufficiently large to satisfy $(1+\gamma)\,y(z) > K\gamma$, denote it $\check{z}$. Then $T''(y(\check{z})) > 0$, and, consequently, $T'(y(z))$ is increasing at $\check{z}$. Hence the inequality $T'(y(z)) > \underline{\varepsilon}_T$ holds also for z in the right neighborhood of $\check{z}$, so that the lower bound on $T''(y(z))$ given in (64) also holds for z to the right of $\check{z}$, and this argument can then be extended for any $z \geq \check{z}$. This then implies

$$\begin{aligned}
\lim_{z\to\infty} T'(y(z)) &= T'(y(\check{z})) + \int_{y(\check{z})}^{\infty} T''(\xi)\,d\xi > T'(y(\check{z})) + \int_{y(\check{z})}^{\infty} \bar{\varepsilon}_T \frac{(1+\gamma)\,\xi - K_y\gamma}{K_y\xi}d\xi \\
&> T'(y(\check{z})) + \bar{\varepsilon}_T \frac{(1+\gamma)\,y(\check{z}) - K_y\gamma}{K_y y(\check{z})}\int_{y(\check{z})}^{\infty} d\xi = \infty,
\end{aligned}$$

which contradicts the bound $T'(y(z)) < 1-\bar{\varepsilon}_T$. Therefore, the marginal tax must converge to zero,

$$\lim_{z\to\infty} T'(y(z)) = 0.$$

Is is worth noting that misspecification concerns enter the proof by way of a finite bound on $zy'(z)$ in (63). In the absence of misspecification concerns, $\theta = \infty$, so that $K_y = \infty$ in (63), and there does not exist a $y(z)$ for which the right-hand side in (64) is positive, implying we cannot guarantee a strictly positive lower bound on $T''(y(z))$. ■

C.2 Analysis of the ODE for the optimal marginal tax rate

We now derive and analyze the differential equation (25) that characterizes the behavior of the optimal marginal tax rate. This will provide intuition for the subsequent proof of Theorem 3.2 in the next subseection. We restrict our attention to the case when the benchmark type distribution is Pareto, as the typical case that leads to nonzero top marginal taxes in absence of model misspecification concerns.

Lemma C.4. *When z is Pareto distributed with shape parameter α under the benchmark model, the worst-*

case density satisfies

$$\frac{d}{dz}\log\widetilde{f}(z)=\frac{d}{dz}\log m(z)+\frac{d}{dz}\log f(z)=-\frac{\mu}{\theta}T'(y(z))\,y'(z)-(\alpha+1)\frac{1}{z}.$$

Proof. Since

$$\begin{aligned} m(z) &= \bar{m}\exp\left(-\frac{\mu}{\theta}T(y(z))\right) \\ f(z) &= \frac{\alpha}{z^{\alpha+1}} \end{aligned}$$

we have, by direct computation,

$$\frac{d}{dz}\log\widetilde{f}(z)=-\frac{\mu}{\theta}T'(y(z))\,y'(z)-(\alpha+1)\frac{1}{z}.$$

■

Proposition C.5. *When the type distribution under the benchmark model is Pareto with shape parameter α, the optimal marginal tax $T'(y)$ obeys the differential equation*

$$-\frac{T''(y)\,y}{1-T'(y)}=-\left[2-\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right]^{-1}\left[\frac{\mu}{\theta}\left[T'(y)\right]^2y-\gamma+\gamma\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right]. \tag{65}$$

Proof. We start with the optimal tax formula

$$\frac{T'(y(z))}{1-T'(y(z))}=(1+\gamma)\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)}=(1+\gamma)\,\widetilde{\phi}(z).$$

Differentiating this formula with respect to z, and using Lemma C.3,

$$\frac{T''(y(z))\,y'(z)}{(1-T'(y(z)))^2}=(1+\gamma)\left[-\frac{1}{z}-\widetilde{\phi}(z)\left[\frac{1}{z}-\frac{\mu}{\theta}T'(y(z))\,y'(z)+\frac{f'(z)}{f(z)}\right]\right].$$

Combining $y'(z)$ terms, we have

$$\left[\frac{T''(y(z))}{T'(y(z))\,(1-T'(y(z)))}-\frac{\mu}{\theta}T'(y(z))\right]y'(z)\,z=-\widetilde{\phi}(z)^{-1}-1-z\frac{f'(z)}{f(z)}.$$

We can now use Lemma C.1 to substitute out $y'(z)$ and obtain

$$\left[\frac{T''(y(z))}{T'(y(z))\,(1-T'(y(z)))}-\frac{\mu}{\theta}T'(y(z))\right]\frac{(1+\gamma)\,y(z)}{\gamma+\frac{T''(y(z))}{1-T'(y(z))}y(z)}=-\widetilde{\phi}(z)^{-1}-1-z\frac{f'(z)}{f(z)}.$$

We can now multiply by the denominator of the compound fraction on the left-hand side, use the expression for $\widetilde{\varphi}(z)$, and combine terms that contain $T''(y(z))$:

$$\left[2-\frac{\gamma-z\frac{f'(z)}{f(z)}}{1+\gamma}T'(y(z))\right]\frac{T''(y(z))}{1-T'(y(z))}y(z)=\frac{\mu}{\theta}\left[T'(y(z))\right]^2 y(z)-\gamma+\gamma\frac{\gamma-z\frac{f'(z)}{f(z)}}{1+\gamma}T'(y(z)).$$

Finally, for the case of the Pareto density,

$$-z\frac{f'(z)}{f(z)}=1+\alpha. \tag{66}$$

Substituting this expression in, we notice that the resulting differential equation does not depend explicitly on z. Since $y(z)$ is strictly monotonic, we can drop the z argument and rewrite the equation as a differential equation for $T(y)$, yielding the expression in the statement of the proposition. ■

We now study the phase diagram of the differential equation (65), which we can rewrite as

$$T''(y)=\frac{1-T'(y)}{y}\left[2-\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right]^{-1}\left[\frac{\mu}{\theta}\left[T'(y)\right]^2 y-\gamma+\gamma\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right].$$

The resulting phase diagram is depicted in Figure 1.

Define the right-hand side of the above equation as a function $h:[0,\infty)\times(-\infty,1)\to\mathbb{R}$:

$$h(y,\tau)=\frac{1-\tau}{y}\left[2-\frac{1+\gamma+\alpha}{1+\gamma}\tau\right]^{-1}\left[\frac{\mu}{\theta}\tau^2 y-\gamma+\gamma\frac{1+\gamma+\alpha}{1+\gamma}\tau\right].$$

We study the function on $(y,\tau)\in(0,\infty)\times(-\infty,1)$. For simplicity, we assume that

$$\frac{1+\gamma+\alpha}{1+\gamma}<2,$$

so that the first bracket in the definition of $h(y,\tau)$ is never zero for $\tau\in(-\infty,1)$. This does not change any conclusions about asymptotic behavior of the optimal tax.

For a given $y\in(0,\infty)$, we first find the isoclines by solving for $\bar{\tau}(y)$ such that $h(y,\bar{\tau}(y))=0$. This $\bar{\tau}(y)$ solves the cubic equation

$$(1-\bar{\tau}(y))\left[\frac{\mu}{\theta}y\bar{\tau}(y)^2+\gamma\frac{1+\gamma+\alpha}{1+\gamma}\bar{\tau}(y)-\gamma\right]=0$$

with three solutions

$$\begin{aligned}\bar{\tau}_{1,2}(y) &= \frac{-\gamma\frac{1+\gamma+\alpha}{1+\gamma}\pm\sqrt{\left(\gamma\frac{1+\gamma+\alpha}{1+\gamma}\right)^2+4\frac{\mu}{\theta}\gamma y}}{2\frac{\mu}{\theta}y}\\ \bar{\tau}_3(y) &= 1,\end{aligned}$$

where $\bar{\tau}_1(y)$ denotes the root with the minus sign. The isoclines are depicted with the black dashed lines in Figure 1. Asymptotically, the differential equation has two steady states

$$\begin{aligned}\lim_{y\to\infty}\bar{\tau}_{1,2}(y) &= 0\\ \lim_{y\to\infty}\bar{\tau}_{3}(y) &= 1.\end{aligned}$$

We can order the three isoclines as

$$\bar{\tau}_1(y) < 0 < \bar{\tau}_2(y) < \bar{\tau}_3(y),$$

and then, as depicted in the phase diagram,

$$\begin{aligned}h(y,\tau) &> 0 \qquad \tau < \bar{\tau}_1(y)\\ h(y,\tau) &< 0 \qquad \bar{\tau}_1(y) < \tau < \bar{\tau}_2(y)\\ h(y,\tau) &> 0 \qquad \bar{\tau}_2(y) < \tau < \bar{\tau}_3(y).\end{aligned}$$

The result from Theorem 3.1,

$$\lim_{z\to\infty} T'(y(z)) = \lim_{y\to\infty} T'(y) = 0,$$

is a transversality condition that pins down the unique optimal marginal tax function $T'(y)$. This optimal path is depicted in the phase diagram with the red solid line with bullet markers. It follows from the proofs of theorems 3.1 and 3.2 that any other path for $T'(y)$ that satisfies equation (65) either converges to one, or becomes negative for some sufficiently high y, both of which violate conditions that the optimal marginal tax function has to satisfy.

C.3 Proof of Theorem 3.2

Using the insights from the phase diagram, we now turn to the proof of Theorem 3.2.

Proof of Theorem 3.2. We investigate the limiting behavior of the differential equation (65). The left-hand side of this equation is the elasticity of take-home rate $1-T'(y)$ with respect to income, or

$$\frac{d\log(1-T'(y))}{d\log y} = -\left[2-\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right]^{-1}\left[\frac{\mu}{\theta}\left[T'(y)\right]^2 y-\gamma+\gamma\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right]. \qquad (67)$$

Since $\lim_{y\to\infty}(1-T'(y)) = 1$, if the limit as $y\to\infty$ of the left-hand side of the above equation exists, it has to be zero. Assume for now that this limit exists. Since $\lim_{y\to\infty}T'(y) = 0$, the first bracket on the right-hand side converges to a positive number as $y\to\infty$. For the same reason, the last term of the second bracket converges to zero as well. Hence the only way how the second bracket converges to zero is when

$$\lim_{y\to\infty}\frac{\mu}{\theta}\left[T'(y)\right]^2 y = \gamma. \qquad (68)$$

In the remainder of the proof, we prove that the limit indeed exists. Assume it does not, so that there exists an ε such that for any $\check{y}$, there exists a $y \geq \check{y}$ such that

$$\left|\frac{\mu}{\theta}\left[T'(y)\right]^2 y - \gamma\right| > \varepsilon > 0. \tag{69}$$

Further, the analysis of the phase diagram implies that along the optimal path, $T'(y)$ monotonically decreases to zero. This means that $T''(y) < 0$, and the elasticity in (67) is strictly positive. In addition, for any arbitrarily small ε_τ, there exists a $\check{y}_\tau$ such that $0 < T'(y) < \varepsilon_\tau$ for all $y \geq \check{y}_\tau$.

Assume first that inequality (69) holds as $\frac{\mu}{\theta}\left[T'(y)\right]^2 y - \gamma > \varepsilon$. Then the elasticity in (67) can be bounded as

$$\frac{d\log\left(1-T'(y)\right)}{d\log y} = \frac{-T''(y)\,y}{1-T'(y)} < -\left[2 - \frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right]^{-1}\left[\varepsilon + \gamma\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right] < 0,$$

which is a contradiction with $T''(y) < 0$, given that $T'(y) > 0$ and $\lim_{y\to\infty} T'(y) = 0$.

On the other hand, assume that inequality (69) holds as $\frac{\mu}{\theta}\left[T'(y)\right]^2 y - \gamma < -\varepsilon$. Then the elasticity in (67) can be bounded as

$$\begin{aligned}\frac{d\log\left(1-T'(y)\right)}{d\log y} &= \frac{-T''(y)\,y}{1-T'(y)} > -\frac{1}{2}\left[\frac{\mu}{\theta}\left[T'(y)\right]^2 y - \gamma + \gamma\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right] \\ &> -\frac{1}{2}\left[\frac{\mu}{\theta}\left[T'(y)\right]^2 y - \gamma\right] - \frac{\gamma}{2}\frac{1+\gamma+\alpha}{1+\gamma}\varepsilon_\tau > \frac{\varepsilon}{2} - \frac{\gamma}{2}\frac{1+\gamma+\alpha}{1+\gamma}\varepsilon_\tau > \frac{\varepsilon}{4},\end{aligned}$$

where the last inequality follows from the fact that ε_τ can be taken to be arbitrarily small when we restrict our attention to sufficiently large $y \geq \check{y} \geq \check{y}_\tau$. We therefore obtain

$$T''(y)\,y < -\frac{\varepsilon}{4}\left(1 - T'(y)\right).$$

As a consequence

$$\begin{aligned}\frac{d}{dy}\left[\frac{\mu}{\theta}\left[T'(y)\right]^2 y - \gamma\right] &= \frac{\mu}{\theta}\left[2T''(y)\,T'(y)\,y + \left[T'(y)\right]^2\right] \\ &< \frac{\mu}{\theta}\left[-\frac{\varepsilon}{2}\left(1-T'(y)\right)T'(y) + \left[T'(y)\right]^2\right] \\ &= \frac{\mu}{\theta}T'(y)\left[-\frac{\varepsilon}{2} + \left(1+\frac{\varepsilon}{2}\right)\left[T'(y)\right]\right],\end{aligned}$$

which becomes negative for sufficiently large y because $\lim_{y\to\infty} T'(y) = 0$. Denote this y as y^*. This means that inequality $\frac{\mu}{\theta}\left[T'(y)\right]^2 y - \gamma < -\varepsilon$ continues to hold as y increases above y^*. Integrating up the inequality

$$d\log\left(1 - T'(y)\right) > \frac{\varepsilon}{4} d\log y$$

for $y \geq y^*$ yields

$$\left(1 - T'(y)\right) - \log\left(1 - T'(y^*)\right) > \frac{\varepsilon}{4}\left(\log y - \log y^*\right)$$

and hence

$$T'(y) < 1 - \left(\frac{y}{y^*}\right)^{\varepsilon/4}\left(1 - T'(y^*)\right).$$

Since $y \to \infty$, the right-hand side must ultimately become negative, which violates the restriction

$T'(y) > 0$.

We have thus shown that condition (69), which is equivalent to a violation of equation (68), cannot hold simulatenously with other restrictions on the optimal marginal tax rate. Either the marginal tax rate function would have to become increasing, or become negative. From the perspective of the phase diagram in Figure 1, analyzed in Appendix C.2, if condition (69) holds for a sufficiently large y, it must be that the given $T'(y)$ is either on a trajectory above the optimal path that crosses the $\bar{\tau}_2(y)$ isocline and converges to one, or becomes negative, crossing the $\bar{\tau}_1(y)$ isocline.

Finally, equation (68) implies that the marginal tax rate has to decay as

$$\log T'(y) = -\frac{1}{2}\log y + o(y)$$

where $o(y)$ converges to a constant as $y \to \infty$. Hence, differentiating this expression with respect to $\log y$, and taking the limit as $y \to \infty$, this limit, if it exists, must be given by expression (22). ■

F Approximation of nonlinear income tax function using a cubic spline

In this Appendix, we lay out the optimization scheme of a tax function T a la Ramsey employed Section 6.

Restricted class of nonlinear income tax function

We restrict a tax function T to a flexible parametric class of functions such that marginal tax rate $T'(y)$ is given by cubic basis functions of $\ln y$ with N knots $\{(\ln y_i, \tau_i)\}_{i=1,\cdots,N}$ and $T'(y)$ is constant outside the range of the knots. Specifically, the marginal tax rate $T'(y)$ is given by

$$T'(y) = \begin{cases} \tau_1 & (y < y_1) \\ \text{CubicSpline}\left(\ln y; \{(\ln y_i, \tau_i)\}_{i=1,\cdots,N}\right) & (y_1 \leq y \leq y_N) \\ \tau_N & (y_N < y). \end{cases}$$

Specifically, we assume that the cubic spline is continuous, differentiable, and smooth at $y = y_i$ for $i = 2, \cdots, N-1$, and is continuous and differentiable at $y = y_1$ and y_N. We also assume that $T'(y) \in [0, 1]$ for any $y \in (0, \infty)$, given that the Mirrleesian solution is known to satisfy it. The tax function $T(y)$ is given by

$$T(y) = -\text{tr} + \int_0^y T'(x)\, dx$$

where tr is an intercept of the tax function.

Choice of N

Deciding the number of knots N faces a tradeoff. We find that large N introduces numerical instability in the optimization over the parameter space to maximize welfare. On the other hand, small N limits welfare gain due to the inflexibility of a cubic spline. Given the tradeoff, we choose the smallest N so that welfare is numerically close to the level from welfare from the Mirrleesian solution.

Optimization strategy

A function within this class is represented by a set of knots $\{(\ln y_i, \tau_i)\}_{i=1,\cdots,N}$ and a transfer term tr. Solving the government problem for the optimal tax function T amounts to search for a set of scalars $\left(\text{tr}, \{\ln y_i\}_{i=1,\cdots,N}, \{\tau_i\}_{i=1,\cdots,N}\right)$. The transfer term tr is set to clear the government budget constraint.

The optimization scheme is as follows. We first fix the location of the outer knots, $(\ln y_1, \ln y_N)$, and search for the marginal tax rates at the end knots and the rest of the knots, $\left(\{\ln y_i\}_{i=2,\cdots,N-1}, \{\tau_i\}_{i=1,\cdots,N}\right)$ to maximize the objective using the Nelder-Mead algorithm. The transfer term tr is set to clear the government budget constraint. This procedure is repeated over multiple combinations of location of two outer knots, $(\ln y_1, \ln y_N)$, and we pick one that achieves the highest value of the objective.

While it is natural to do so, we choose not to include $(\ln y_1, \ln y_N)$ as control variable in the Nelder-Mead algorithm. We find that the Nelder-Mead algorithm tends to put the outer knots far away from the range of the income distribution, and the marginal tax rates at the outer knots (τ_1, τ_N) can take a range of values in the optimization scheme. This is because the marginal tax rates at the outer knots have quantitatively small implication for the welfare.

Performance benchmarking in the model with single-dimensional uncertainty

The Ramsey solution is benchmarked against the true Mirrleesian solution in the one-dimensional setup. For this computation, the productivity z-space is discretized with 500 points in log-space.

Figure 13 compares Mirrleesian solutions and Ramsey solutions with three knots $(N = 3)$ for both a rational case $(\theta = \infty)$ and a robust case $(\theta = 100)$. The broken lines represent the marginal tax rates based on the Mirrlees solution (Diamond-Saez formula) and solid lines represent the rates based on the Ramsey solution (Cubic Spline). Blue and orange lines are the solution for the rational case $(\theta = \infty)$ and robust case $(\theta = 100)$, respectively. The figure shows that the Ramsey solutions approximate the Mirrlees solutions well in terms of the overall shape of the marginal tax schedule, except for low income earners. It is also noteworthy that both approach achieves quite similar level of welfare. In terms of consumption equivalence, welfare loss from choosing the Ramsey solutions are roughly 0.05% for the rational and robust cases. We conclude, the Ramsey solution with the number of knots $N = 3$ is flexible enough to approximate the Mirrleesian solution with minimal welfare loss.

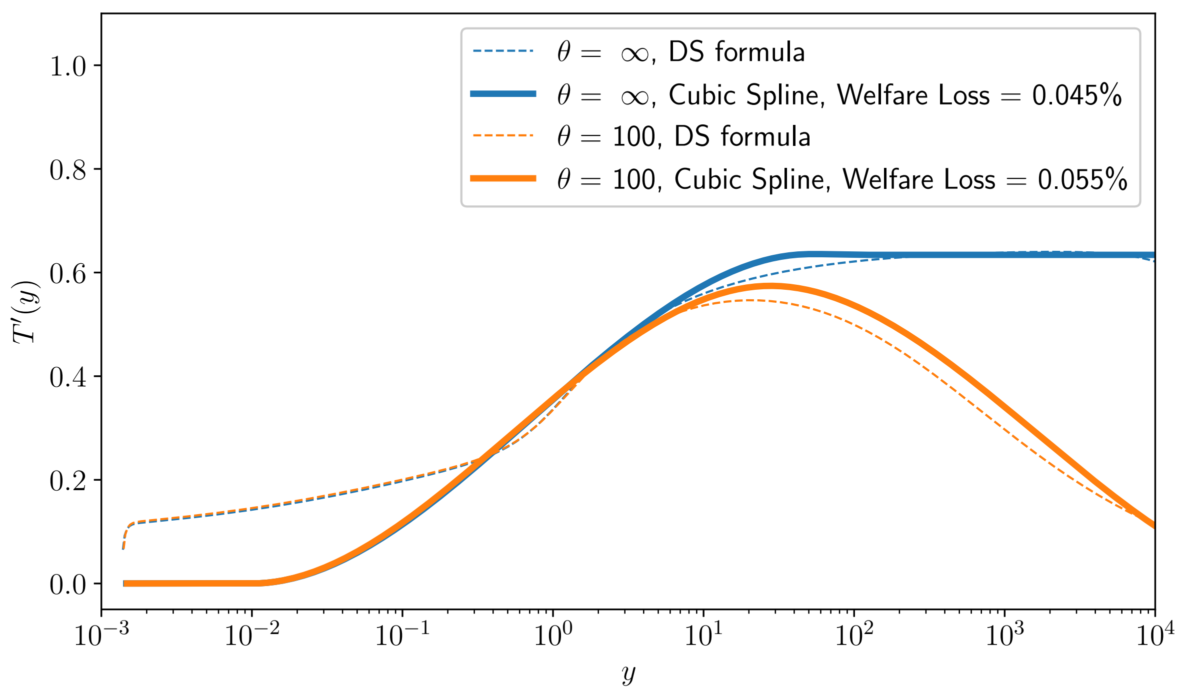

Figure 13: Optimal Ramsey tax policy in the economy with one-dimensional heterogeneity

G Optimal tax schedule under multidimensional uncertainty with the status-quo tax policy calibrated to the US

In this Appendix, we present the optimal tax schedule under multidimensional misspecification concerns with the status-quo tax policy T^0 calibrated to the one in the United States.

To approximately capture the level of marginal tax rates on income, we set the status-quo tax policy T^0 with an affine function with a constant marginal tax rate of 35%. The disutility shifter Ψ is calculated with the new status-quo tax policy.

Figure 14 shows the result.

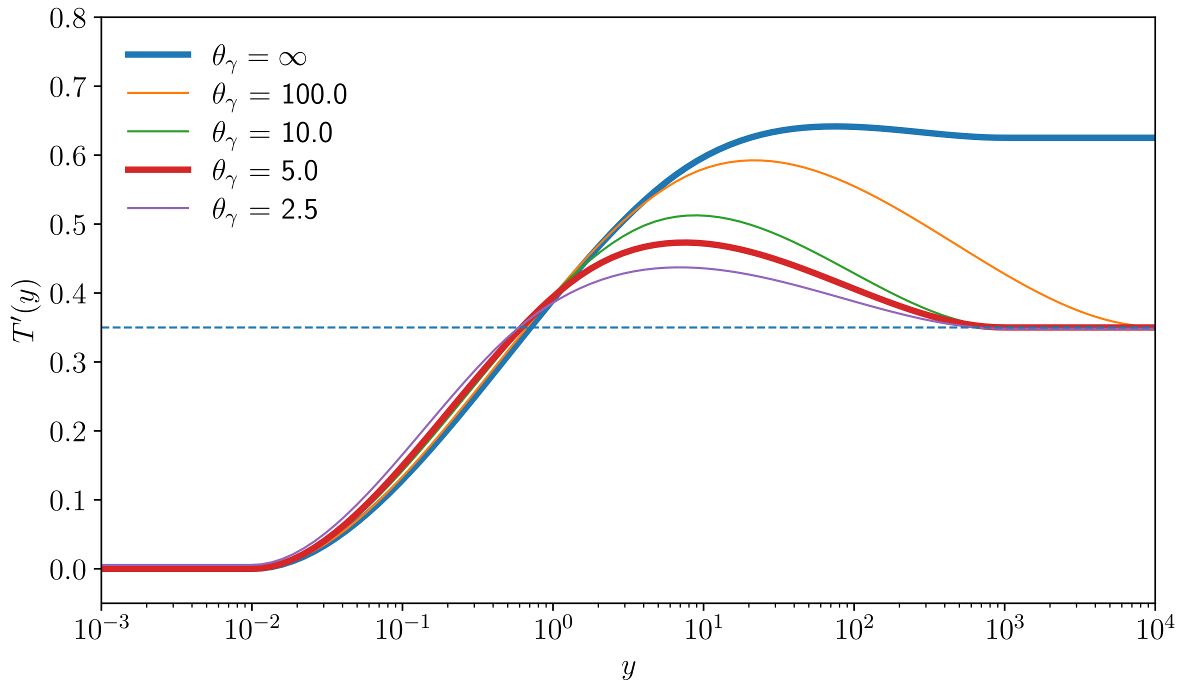

Figure 14: Optimal marginal tax schedules for alternative levels of multidimensional misspecification concerns. The status quo tax policy is an affine tax function with a constant marginal tax rate of 35%, $T^{0\prime}(y) = 0.35$.

H Comparison with Lockwood, Sial, and Weinzierl (2021)

In this Appendix, we compare the optimal tax policy in our model with the two-dimensional heterogeneity to the one in Lockwood et al. (2021).

Similar to our work, Lockwood et al. (2021) study the optimal income tax scheme when the tax planner faces uncertainty about the value of the elasticity of taxable income. They show that the uncertainty generally leads to a more progressive tax schedule with a higher marginal tax rate at the top, relative to the case without the uncertainty about the ETI. On the contrary, we find that the optimal tax policy is less progressive and the marginal tax rate at the top is lower when the planner has concerns about the misspecification of the right tail of the distribution of the ETI, relative to the case without concerns about the distributin of the ETI.

Those two findings are consistent. While Lockwood et al. (2021) maximize the expected welfare of the tax planner over the distribution of the ETI, this is equivalent to maximizing the welfare under the measure of heterogeneous population with different values of the ETI. In our experiment, Lockwood et al. (2021)'s result with the uncertainty corresponds to the case without concerns ($\theta_z = \infty$ and $\theta_\gamma = \infty$). Therefore, relative to the Lockwood et al. (2021) case without concerns, we find that the concerns about the ETI leads to lower marginal tax rate at the top. On the other hand, relative to the case without heterogeneity in the ETI, Lockwood et al. (2021) found that the heterogeneity in the ETI leads to a higher marginal tax rate at the top.

We numerically confirm this point. Figure 15 shows the optimal tax policy in three models with and without heterogeneity in the elasticity of taxable income. The optimal marginal tax schedulde is computed as a cubic spline with three knots. The orange line ("Homogeneous ETI") represents the optimal marginal tax schedule in a model without heterogeneity in the elasticity of taxable income, where the elasticty is assumed to be 0.59, the mean value of the ETI in Fig 9. The blue line ("Heterogeneous ETI without concerns") represents the optimal marginal tax schedule in a model with heterogeneity in the elasticity of taxable income, where the elasticities are distributed according to Fig 9. The read line ("Heterogeneous ETI with concerns") is the optimal marginal tax schedule with the penality parameter $\theta_\gamma = 5$. The blue line and the read line are taken from the main text. The figure confirms that the heterogeneity in the ETI leads to a higher marginal tax rate at the top (the difference between the orange line and the blue line), while the concerns about the ETI leads to a lower marginal tax rate at the top (the difference between the blue line and the red line).

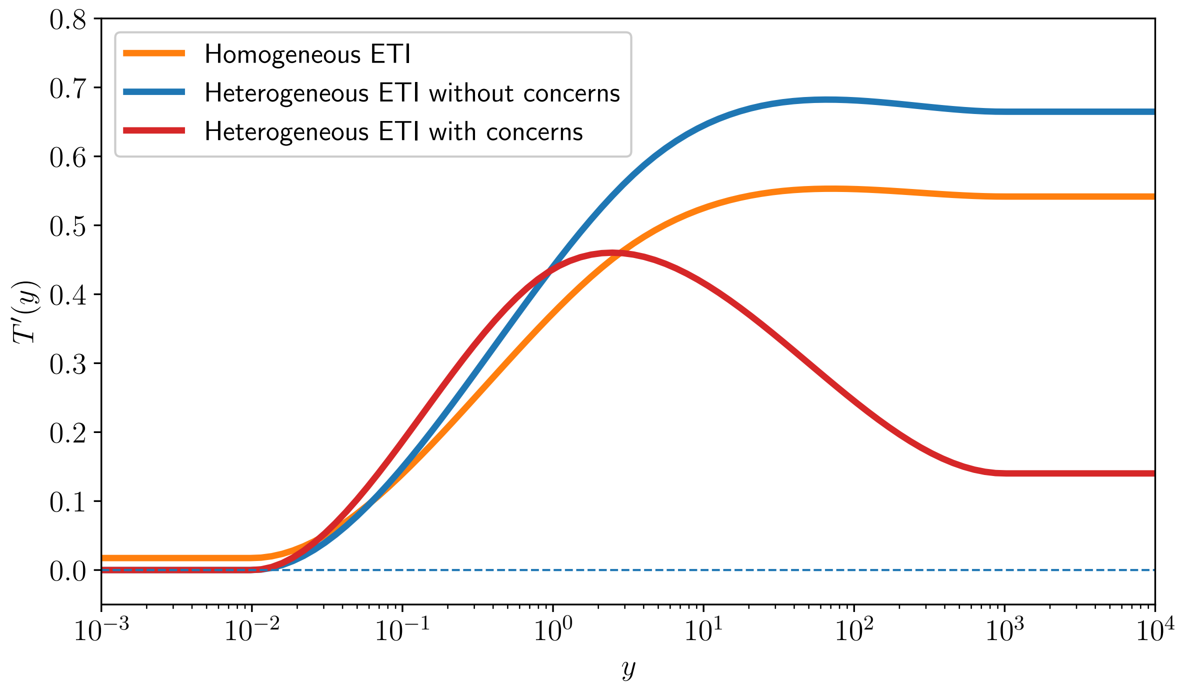

Figure 15: Commparison of optimal tax policy in models with and without heterogeneity in the elasticity of taxable income

References

Adam, Klaus and Albert Marcet. 2011. Internal Rationality, Imperfect Market Knowledge and Asset Prices. *Incompleteness and Uncertainty in Economics* 146 (3):1224–1252. 2

Almeida, Caio and René Garcia. 2017. Economic Implications of Nonlinear Pricing Kernels. *Management Science* 63 (10):3361–3380. 23

Bhandari, Anmol, Jaroslav Borovička, and Paul Ho. 2024. Survey Data and Subjective Beliefs in Business Cycle Models. *Forthcoming in the Review of Economic Studies* . 2

Borovička, Jaroslav, Lars Peter Hansen, and José A. Scheinkman. 2016. Misspecified Recovery. *Journal of Finance* 71 (6):2493–2544. 23

Carroll, Gabriel. 2015. Robustness and Linear Contracts. *American Economic Review* 105 (2):536–63. 2

Chang, Minsu and Chunzan Wu. 2023. When in Doubt, Tax More Progressively? Uncertainty and Progressive Income Taxation. *mimeo* . 2

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber. 2011. Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins. *American Economic Review* 101 (3):471–75. 36

Cressie, Noel and Timothy R. C. Read. 1984. Multinomial Goodness-of-Fit Tests. *Journal of the Royal Statistical Society, Series B (Methodological)* 46 (3):440–464. 23, 76

Diamond, Peter A. 1998. Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates. *American Economic Review* 88 (1):83–95. 1, 10, 11

Diamond, Peter A. and Emmanuel Saez. 2011. The Case for a Progressive Tax: From Basic Research to Policy Recommendations. *Journal of Economic Perspectives* 25 (4):165–190. 13

Ferriere, Axelle and Anastasios G. Karantounias. 2019. Fiscal Austerity in Ambiguous Times. *American Economic Journal: Macroeconomics* 11 (1):89–131. 35

Golosov, Mikhail and Ilia Krasikov. 2023. The Optimal Taxation of Couples. *mimeo* . 40

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski. 2016. Redistribution and Social Insurance. *American Economic Review* 106 (2):359–86. 1

Hansen, Lars Peter and Jianjun Miao. 2022. Asset Pricing under Smooth Ambiguity in Continuous Time. *Economic Theory* 74 (2):335–371. 2

Hansen, Lars Peter and Thomas J. Sargent. 2001a. Acknowledging Misspecification in Macroeconomic Theory. *Review of Economic Dynamics* 4 (3):519–535. 7

———. 2001b. Robust Control and Model Uncertainty. *American Economic Review* 91 (2):60–66. 7

———. 2008. *Robustness*. Princeton, N.J.: Princeton, N.J. 1

———. 2012. Three Types of Ambiguity. *Journal of Monetary Economics* 59 (5):422–445. 2, 35

———. 2015. Four Types of Ignorance. *Journal of Monetary Economics* 69:97–113. 35

Heathcote, Jonathan and Hitoshi Tsujiyama. 2021. Optimal Income Taxation: Mirrlees Meets Ramsey. *Journal of Political Economy* 129 (11):3141–3184. 25, 30

Karantounias, Anastasios G. 2013. Managing Pessimistic Expectations and Fiscal Policy. *Theoretical Economics* 8 (1):193–231. 2

———. 2023. Doubts about the Model and Optimal Policy. *Journal of Economic Theory* 210:105643. 35

Kwon, Hyosung and Jianjun Miao. 2017. Three Types of Robust Ramsey Problems in a Linear-Quadratic Framework. *Journal of Economic Dynamics and Control* 76:211–231. 35

Lockwood, Benjamin B., Afras Sial, and Matthew Weinzierl. 2021. Designing, Not Checking, for Policy Robustness: An Example with Optimal Taxation. *Tax Policy and the Economy* 35:1–54. 1, 36, 39, 40, 89, 90

Mirrlees, James A. 1971. An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies* 38 (2):175–208. 1, 9

Myerson, Roger B. 1979. Incentive Compatibility and the Bargaining Problem. *Econometrica* 47 (1):61–73. 9

Negishi, Takashi. 1960. Welfare Economics and Existence of an Equilibrium for a Competitive Economy. *Metroeconomica* 12 (2–3):92–97. 7

Neisser, Carina. 2021. The Elasticity of Taxable Income: A Meta-Regression Analysis. *The Economic Journal* 131 (640):3365–3391. 36

Rauh, Joshua and Ryan Shyu. 2024. Behavioral Responses to State Income Taxation of High Earners: Evidence from California. *American Economic Journal: Economic Policy* 16 (1):34–86. 40

Saez, Emmanuel. 2001. Using Elasticities to Derive Optimal Income Tax Rates. *Review of Economic Studies* 68 (1):205–229. 1, 10, 11

Saez, Emmanuel, Joel Slemrod, and Seth H. Giertz. 2012. The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review. *Journal of Economic Literature* 50 (1):3–50. 35

Sion, Maurice. 1958. On General Minimax Theorems. *Pacific Journal of Mathematics* 8 (1):171–176. 45

Vairo, Maren. 2024. Robustly Optimal Income Taxation. *mimeo* . 2