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

August 30, 2025

Abstract

This paper studies optimal tax design when the cross-sectional distribution of types may be misspecified, and the government acts cautiously vis-à-vis these misspecifications. In models without such concerns, fat-tailed distributions imply positive—often high—top marginal tax rates. We demonstrate that even vanishingly small misspecification concerns overturn this finding, driving top marginal tax rates to zero. Calibrating concerns to observed variation in income distributions shows that taxes for below-average incomes remain essentially unchanged, while progressivity for high-income earners is substantially reduced.

We thank Manuel Amador, George-Marios Angeletos, Marco Bassetto, Hal Cole, Lars Hansen, Jonathan Hartley, Jonathan Heathcote, Rishabh Kirpalani, Christos Koulovatianos, Tomoyuki Nakajima, Tom Sargent, 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, Stanford SITE 2024, MAFE Workshop on Risk Mitigation in Berlin 2024, EEA in Rotterdam 2024, the 3rd XAmsterdam Macroeconomic Workshop, USC, UPenn, NYU, Minnesota, Kent, IMF, 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 distribution of labor productivities and labor supply preferences. Despite increasingly available surveys and administrative micro-data on income, estimating a 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, these issues get even more difficult if we want to jointly estimate productivity and labor supply elasticities for such individuals.[1]

In this paper, we study optimal taxation of labor income when the government acknowledges uncertainty about the underlying type distribution. We find that these concerns generally lead to lower progressivity, especially for high-income earners. Strikingly, we find that the celebrated Diamond (1998) and Saez (2001)'s high top marginal tax rate result is overturned. Optimal tax schedules are hump-shaped, and in special cases top marginal tax rates approach zero independently of the underlying productivity distribution, the degree of uncertainty, and other 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. 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 represents 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 the literatures on public finance and decision-making under ambiguity, 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 nonlinear 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 abstracting from misspecification concerns as in the traditional Mirrlees framework, our government uses the max-min formulation

[1] Mankiw, Weinzierl, and Yagan (2009) summarize the literature and conclude that "All this leaves the policy advisor in an uncomfortable position ... [because] ... many of the key assumptions are open to debate."

discussed previously. If we set the penalty in the max-min problem to be sufficiently large, 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 "rational model", and we refer to the type distribution in this rational model as the "benchmark" distribution. We use the counterparts "robust model" 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 literature on decision-making under ambiguity, 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 of the population. 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 dimensions of the underlying distribution. We can also formally articulate when the government is more uncertain about the extreme ends ("tails") of the distribution.

We begin with a scenario where the type distribution is one-dimensional (skills). In this context, we derive a modified Diamond–Saez tax formula that incorporates misspecification concerns. The formula highlights a balance between 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. The difference lies in the replacement of the hazard rates and distributions with the worst-case distribution. This worst-case distribution is an endogenous object that relies on the shape of the tax function, and therefore emerges as a fixed-point of the robust problem along with the optimal tax schedule.

Our first theoretical results concern the analysis of the top tax rate in a familiar setting with quasilinear preferences, Rawlsian welfare weights, and a benchmark productivity distribution that has a Pareto tail. In the rational model, when the planner is certain that the benchmark distribution is the correct type distribution, the Diamond–Saez 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 conventionally used 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 productivity threshold to offset the cost of distortions due to an increase in the marginal tax rate at that threshold. For similar reasons, for bounded or thin-tailed distributions, 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, with the magnitude of this reweighting 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.

The adverse player distorts the right tail of the productivity distribution because this provides the best tradeoff between lowering the welfare objective against the cost of misspecification that is invariant to an individual's type regardless of the degree of misspecification concern. The mechanism formalizes the practical concern of policymakers in dealing with the tradeoff between gains from redistribution and distortionary costs of taxation, especially in the right tail. We demonstrate that the optimal top tax rate gradually reduces to zero, with the worst-case distribution approaching a thin-tailed one, in spite of the benchmark distribution having a Pareto tail.

Because this is an asymptotic result, we also study how fast the top rate falls. A slower decay raises more revenue from high-productivity types but also makes these types more valuable from the budgetary perspective and thus more heavily targeted by the adverse player. The logic is reversed if the decay is too fast. The balance of these two forces yields a unique asymptotic elasticity of marginal tax rates with respect to income that equals minus one half. This corresponds to a two-third reduction in the tax rate for a tenfold increase in income.

After establishing the main results in the tractable quasilinear-Rawlsian setting, we extend our analysis to broader environments. We study extensions that allow 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 above by minus one half. We then generalize the analysis to a class of power divergence penalty functions that go beyond 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 productivity distribution is calibrated as Pareto-lognormal. Although the top tax rate is not influenced by 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 observed time-series variation in income distributions over such planning horizons as a measure of the degree of uncertainty the planner faces. We construct the smallest "entropy ball" that encloses the observed distributions in the given time horizon. The radius of this set maps to the penalty parameter in our max-min problem and the center maps to the benchmark distribution. The resulting set of distributions enclosed in the entropy ball and hence viewed by the planner as plausible is rich in the sense of containing all nonparametric distributions that are statistically sufficiently similar to the benchmark.

We evaluate our findings against the rational model, in which the penalty parameter is infinite. While the presence of misspecification concerns has a negligible impact on marginal tax rates for below-average income, we observe significant effects for marginal tax rates on high-income individuals. In the rational model, households earning 20 times the average income face a marginal tax rate exceeding 68%, while under our preferred calibration, the optimal tax rate peaks at 57.5% for earnings around 7.5 times the average income. This rate falls to 40% for households with 100 times the average income, 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 productivity. We propose a generalization of the entropy penalty that allows to assign different degrees of misspecification concerns across alternative dimensions of the distribution. We calibrate the degree of misspecification to match recent estimates of elasticities of taxable incomes for high-income earners.

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. In our calibration, we again find a negligible impact of misspecification concerns for marginal tax rates on below-average incomes but a substantial effect for high-income earners. The marginal tax rate for households with 20 times and 100 times the average income reaches 37% and 24%, respectively, compared to about 68% in the case of the rational model. 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.

1.1 Related literature

Our paper contributes to the optimal taxation literature originating with Ramsey (1927) and Mirrlees (1971). Key lessons and canonical prescriptions are synthesized by Diamond and Saez (2011) and Mankiw et al. (2009), with more recent quantitative analyses by Golosov et al. (2016) and Heathcote and Tsujiyama (2021).

A growing literature studies decision-making under model misspecification (Ilut and Schneider (2023) provide an extensive review). We adopt the robust-control framework of Hansen and Sargent (2008), which delivers tractable formulations within Mirrleesian environments. Applications span linear-quadratic policy problems (Hansen and Sargent 2012, 2015, Kwon and Miao 2017), fiscal policy (Karantounias 2013, 2023, Ferriere and Karantounias 2019), sovereign default (Pouzo and Presno 2016, Roch and Roldán 2023), macro business cycles (Bidder and Smith 2012, Bhandari et al. 2025), and asset pricing (Maenhout 2004). We build on this tradition but depart by introducing misspecification concerns about the cross-sectional type distribution in an optimal income tax setting.

Within nonlinear optimal taxation specifically, a few studies incorporate statistical concerns explicitly. Lockwood et al. (2021) and Chang and Wu (2025) model a Bayesian government facing parametric uncertainty and find that such uncertainty tends to increase progressivity. By contrast, our non-Bayesian robust approach places ambiguity directly on an infinite-dimensional object—the type distribution—and, in our baseline, this force reduces progressivity relative to the standard benchmark. We discuss the mapping between these approaches and the source of the differing prescriptions in Section 5.

A related but distinct line of work studies mechanism design with moral hazard when the planner has limited knowledge of agents' action sets. Carroll (2015) develops the general framework, and Vairo (2024) applies it to optimal taxation when agents choose the riskiness of their income profiles. In that environment, uncertainty over action sets can justify uniformly progressive schedules and, in some cases, higher top rates; declining marginal tax rates would encourage socially undesirable risk-taking. Our analysis abstracts from such moral hazard considerations and focuses instead on misspecification of the type distribution.

The rest of the paper is structured as follows. Section 2 presents the model with one-dimensional types. Section 3 contains our main theoretical characterization of optimal top marginal tax rates. Section 4 uses a calibrated economy for a quantitative exploration of the full tax schedule. Section 5 extends the environment to multi-dimensional types. Section 6 concludes.

2 Model with one-dimensional types

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 labor supply choice 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 $\mathcal{U}(z;T)$. In what follows, 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. In absence of misspecification concerns, 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.[2] 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 absence of model misspecification concerns, the government chooses

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

a tax schedule to maximize the welfare objective subject to the budget constraint

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

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

$$\mathcal{E}\left(f, \widetilde{f}\right) = \mathbb{E}[m \log m] = \int_{\underline{z}}^{\bar{z}} m(z) \log m(z)\, f(z)\,dz. \tag{3}$$

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:

$$\mathcal{F}(f, \kappa) = \left\{\widetilde{f} : \mathcal{E}\left(f, \widetilde{f}\right) \leq \kappa\right\}. \tag{4}$$

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

$$\max_{T} \min_{\widetilde{f} \in \mathcal{F}(f,\kappa)} \widetilde{\mathbb{E}}[\psi \mathcal{U}] + V\left(\widetilde{\mathbb{E}}\left[T\left(\mathcal{Y}\right)\right]\right). \tag{5}$$

The first term in the objective function is equal to the government welfare $\widetilde{\mathbb{E}}[\psi\mathcal{U}]$ 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 (5) 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 max-min problem in (5) the worst-case distribution.

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

$$\max_{T} \min_{\substack{m>0 \\ \mathbb{E}[m]=1}} \mathbb{E}\left[m\psi\mathcal{U}\right] + V\left(\mathbb{E}\left[mT\left(\mathcal{Y}\right)\right]\right) + \theta\mathbb{E}\left[m\log m\right], \tag{6}$$

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 (5). 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 $\widetilde{f} = f$, 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 $\mathbb{E}[m\psi\mathcal{U}]$. On the other hand, it strives for adverse budgetary consequences by underweighing types for whom $T(\mathcal{Y}(z))$ is positive (net tax payers) while, vice versa, overweighing those for whom $T(\mathcal{Y}(z))$ is negative (net tax recipients), and thus lowering the net revenue $\mathbb{E}[mT(\mathcal{Y})]$. 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 $\theta\mathbb{E}[m\log m]$.

2.1 Mirrleesian formulation

Rather than solving for the optimal tax function (6), we follow Mirrlees (1971), and characterize the optimal allocation as a solution to a mechanism design problem, 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 (6).

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

$$U_c\left(c(z), \frac{y(z)}{z}\right) c'(z) + U_n\left(c(z), \frac{y(z)}{z}\right) \frac{y'(z)}{z} = 0. \tag{7}$$

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 (7), we obtain

$$\frac{dU}{dz} = -U_n\left(c(z), \frac{y(z)}{z}\right) \frac{y(z)}{z^2}. \tag{8}$$

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 (6).[3] The planner is thus solving

$$\min_{\substack{m>0 \\ \mathbb{E}[m]=1}} \max_{c,y} \int_{\underline{z}}^{\bar{z}} \psi(z) U\left(c(z), \frac{y(z)}{z}\right) m(z) f(z) dz + V(G) + \theta \int_{\underline{z}}^{\bar{z}} m(z) \log m(z) f(z) dz \tag{9}$$

subject to the IC constraint (8) and the budget constraint

$$G = \int_{\underline{z}}^{\bar{z}} (y(z) - c(z)) m(z) f(z) dz. \tag{10}$$

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

$$q(c, n) = -n \frac{U_n(c, n)}{U_c(c, n)} \tag{11}$$

[3] 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.

is strictly increasing in n for each fixed c implies a single-crossing property under which the local IC constraint (8) also implies global incentive compatibility, and allocations $(c(z), y(z))$ 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 (9), 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

$$\begin{aligned} H(\mathcal{U}, y, m, \lambda) \quad = \quad & \psi(z) \mathcal{U}(z) m(z) f(z) + \theta m(z) \log m(z) f(z) - \chi m(z) f(z) \\ & -\lambda(z) U_n\left(c(z), \frac{y(z)}{z}\right) \frac{y(z)}{z^2} + \mu [y(z) - c(z)] m(z) f(z). \end{aligned} \tag{12}$$

Here, χ and μ are multipliers on the constraints $\mathbb{E}[m] = 1$ and (10), respectively, and $c(z)$ is defined implicitly from the definition of the utility function as $c(z) = C(\mathcal{U}(z), y(z))$.

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*

$$U(c, n) = c - \frac{n^{1+\gamma}}{1+\gamma}, \tag{13}$$

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

$$\lim_{G \searrow \underline{G}} V'(G) = \infty \qquad \lim_{G \nearrow \infty} V'(G) = 0. \tag{14}$$

The Inada conditions (14) guarantee an interior solution to the government problem.

Since the mechanism is incentive compatible, alternative distributions can be consistently characterized by weighting functions $m(z)$ indexed by types z. The first-order condition with respect to $m(z)$ in (12) together with the restriction $\mathbb{E}[m] = 1$ yield a characterization of the worst-case distortion in the form of an exponential tilting formula

$$m(z) = \bar{m} \exp\left(-\frac{1}{\theta}[\psi(z) \mathcal{U}(z) + \mu T(y(z))]\right), \tag{15}$$

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. Since $m(z)$ is strictly positive, alternative distributions $\widetilde{f}(z) = m(z) f(z)$ preserve the support of $f(z)$ and hence the set of local IC constraints (8). The remaining optimality conditions then imply the formula for the marginal tax

$$\frac{T'(y(z))}{1-T'(y(z))} = (1+\gamma)\,\frac{\widetilde{\Psi}(z)-\widetilde{F}(z)}{1-\widetilde{F}(z)}\,\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)}, \tag{16}$$

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 (15). The first term on the right-hand side of (16) 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 (15) 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 (15) 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 (9) and hence the degree of model misspecification concerns. A small value of θ reflects more substantial concerns, which leads to stronger exponential tilting in (15). 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 (15) and the tax function $T(y(z))$ in (16) are determined jointly as an outcome of the minimax problem (9). Given the tax function, the worst-case density delivers the lowest penalized objective in (9), 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 (16) in the presence of model misspecification concerns.

When the planner is utilitarian with $\psi(z) \equiv 1$, then, due to the quasilinear form of preferences in (13), 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

$$\frac{T'(y(z))}{1 - T'(y(z))} = (1+\gamma)\frac{1 - \widetilde{F}(z)}{z\widetilde{f}(z)}. \tag{17}$$

The worst-case distortion for $z \geq \hat{z}$ then becomes

$$m(z) = \bar{m} \exp\left(-\frac{\mu}{\theta} T(y(z))\right), \tag{18}$$

corresponding to a special case of formula (15). 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 fiscal consequences of not having sufficiently many high-type workers who contribute disproportionately to tax revenues.

In Section 3.4, we provide the analysis of more general cases that relax the assumption of quasilinear preferences in (13), involve more general welfare weighting functions, and consider penalty functions other than entropy. 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 (17) 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 (17) 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.[4]

When model misspecification concerns are present, we characterize the shape of the worst-case density

$$\widetilde{f}(z) = m(z) f(z) = \bar{m} \exp\left(-\frac{\mu}{\theta} T(y(z))\right) f(z). \tag{19}$$

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 (17) to be different from zero, we need $\lim_{z\to\bar{z}} z\widetilde{f}(z) = 0$. In

[4] 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 $\alpha_y = 1.5$). Mankiw et al. (2009) emphasize the lack of robustness of these results with respect to measurement issues and modeling assumptions.

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 $-\infty$.

This assumption is a technical requirement for existence of interior limiting tax rates, and allows, for example, for arbitrary distributions with asymptotically Pareto tails.

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 (11) 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 (19) 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 $\lim_{z\to\infty} T'(y(z)) \leq \lim_{z\to\infty} T'_{rat}(y_{rat}(z))$. Misspecification concerns thus lower top marginal tax rates.

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 (7) for the case of quasilinear preferences (13) would asymptotically behave as

$$y(z) = \left(1 - T'(y(z))\right)^{\frac{1}{\gamma}} z^{\frac{1+\gamma}{\gamma}} \approx (1-\tau_\infty)^{\frac{1}{\gamma}} z^{\frac{1+\gamma}{\gamma}}, \tag{20}$$

and the worst-case distortion as

$$m(z) = \bar{m}\exp\left(-\frac{\mu}{\theta}T(y(z))\right) \approx \bar{m}\exp\left(-\frac{\mu}{\theta}\tau_\infty(1-\tau_\infty)^{\frac{1}{\gamma}} z^{\frac{1+\gamma}{\gamma}}\right). \tag{21}$$

In that case, an application of L'Hôpital's rule to the tax formula (17) yields

$$\lim_{z\to\infty}\frac{T'(y(z))}{1-T'(y(z))} = \lim_{z\to\infty}(1+\gamma)\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)} = \lim_{z\to\infty}\frac{1+\gamma}{\frac{\mu}{\theta}z\frac{d}{dz}T(y(z)) - \frac{d\log f(z)}{d\log z} - 1} = 0, \tag{22}$$

because the first term in the denominator diverges to ∞. This contradicts 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, or the shape of the underlying type distribution. 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 (18) 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 (19). 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 (15). 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 in terms of a specific differential equation that will be also helpful in the numerical characterization of the full 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 Pareto distributed right tail with shape parameter α, and $\theta < \infty$. Then the limiting tax rate satisfies*

$$\lim_{y\to\infty} \frac{\mu}{\theta} \left[T'(y)\right]^2 y = \gamma. \tag{23}$$

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

$$\lim_{y\to\infty} \frac{d\log T'(y)}{d\log y} = -\frac{1}{2}. \tag{24}$$

The proof of the theorem is provided in Appendix C.3. Expression (23) 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 $-\frac{1}{2}$. A tenfold increase in income therefore cuts the marginal tax rate by approximately two thirds.

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 (24) 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 (24) 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

$$m(\bar{z}) = \bar{m}\exp\left(-\frac{\mu}{\theta}T(\bar{y})\right) \approx \widetilde{m}\exp\left(-\frac{\mu}{\theta}\bar{y}^{1+\nu}\right) \tag{25}$$

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 (20) 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 (22), we obtain

$$\lim_{z\to\infty} \frac{T'(y(z))}{1-T'(y(z))} = (1+\gamma)\lim_{z\to\infty} \frac{1}{\frac{\mu}{\theta} z T'(y(z))\, y'(z) - \frac{d\log f(z)}{d\log z} - 1} \tag{26}$$

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 (26) 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 (26) 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 (25), 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.[5] 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 (17) with respect to the productivity z and a sequence of algebraic manipulations, we obtain the differential equation

$$T''(y) = \frac{1-T'(y)}{y}\left[2 - \frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right]^{-1}\left[\frac{\mu}{\theta}\left[T'(y)\right]^2 y - \gamma + \gamma\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right]. \tag{27}$$

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 $T''(y) = h(y, T'(y))$ 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.

[5]For example, when the decay rate higher than one ($\nu < -1$), the tax revenue converges to a finite bound. The distortion $m(z)$ is then also bounded and yields an asymptotically vanishing misspecification concern in the right tail, contradicting the zero top marginal tax result from Theorem 3.1.

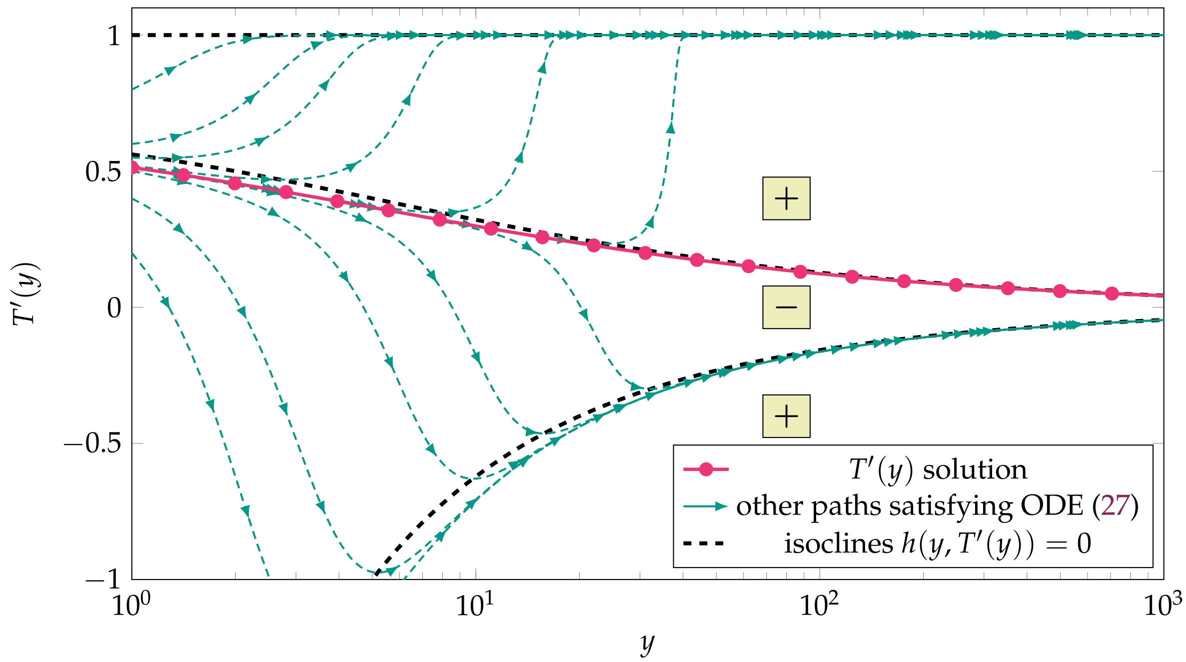

Figure 1: Phase diagram for differential equation (27). The dashed solid lines correspond to isoclines $h(y, T'(y)) = 0$, and the sign of $h(y, T'(y))$ in between these lines is depicted in the boxes. The magenta line with circles corresponds to the unique strictly positive solution that satisfies the terminal condition. Green dashed lines are other trajectories that satisfy (27). The parameters are $\mu = 1$, $\theta = 1$, $\alpha = 1.5$, $\gamma = 2$.

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 (15)–(16). 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 (27) 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.

These stability properties of the differential equation also carry over to more general versions of the model that we analyze in the quantitative application. In addition, they suggest a stable numerical algorithm for the full solution of the model that starts from a terminal condition in the right tail and iterates backward on the appropriate counterpart of equation (27).

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 distribution. 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

$$U(c,n) = \frac{c^{1-\rho}-1}{1-\rho} - v\frac{n^{1+\gamma}}{1+\gamma} \tag{28}$$

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 (12) yield the optimal tax formula in the form

$$\frac{T'(y(z))}{1-T'(y(z))} = (1+\gamma)\frac{1-\widetilde{F}_\rho(z)}{z\widetilde{f}_\rho(z)}, \tag{29}$$

where

$$\widetilde{f}_\rho(z) = \bar{u}^{-1}(c(z))^\rho m(z) f(z) \tag{30}$$

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 (30).

Theorem 3.3. *Assume that worker's preferences are given by the separable isoelastic form (28), the underlying productivity distribution is such that* $\lim_{y\to\infty} T'(y)$ *exists, and* $\theta < \infty$*. Then the marginal tax rate vanishes at the top,*

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

Moreover, if the productivity distribution has a Pareto distributed right tail, the tax rate satisfies

$$\lim_{y\to\infty} \frac{\mu}{\theta} \left[T'(y)\right]^2 y = \gamma + \rho, \tag{31}$$

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

$$\lim_{y\to\infty} \frac{d \log T'(y)}{d \log y} = -\frac{1}{2}. \tag{32}$$

This theorem generalizes Theorems 3.1 and 3.2 to the case of concave marginal utility of consumption, and the proof is provided in Appendix D.2. Setting $\rho = 0$ recovers the 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 productivity. But since the zero limiting marginal tax result does not depend on the labor supply elasticity to begin with, it is also robust to the introduction of more general separable utility form (28). The only difference is expression (31) that adjusts for the consumption elasticity.

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

$$m(z) = \bar{m} \exp\left(-\frac{1}{\theta}\left[\mathcal{U}(z) + \mu T(y(z))\right]\right), \tag{33}$$

and the marginal tax formula can be expressed as

$$\frac{T'(y(z))}{1-T'(y(z))} = (1+\gamma)\,\frac{\widetilde{\Psi}(z) - \widetilde{F}_{\rho}(z)}{z\widetilde{f}_{\rho}(z)}, \tag{34}$$

where $\widetilde{f}_{\rho}(z)$ is the inverse marginal utility weighted worst-case density (30) with cumulative distribution function $\widetilde{F}_{\rho}(z)$, 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 $\widetilde{\Psi}(z) = \widetilde{F}_{\rho}(z)$, implying zero marginal taxes on everyone.

The worst-case distortion $m(z)$ 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 (28), the underlying productivity distribution is such that $\lim_{y\to\infty} T'(y)$ exists, the planner is utilitarian with $\psi(z) \equiv 1$, the curvature of the utility function is $\rho > 0$, and $\theta < \infty$. Then*

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

Moreover, if the productivity distribution has a right tail that is Pareto distributed, the limiting elasticity of the marginal tax with respect to income is equal to

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

The proof is provided in Appendix D.3. 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 (9) of the robust planner penalizes deviations from the benchmark using an entropy penalty, also known as the Kullback–Leibler divergence. Entropy is a natural choice from a statistical perspective but 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 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 $\eta \to 0$. The limit as $\eta \to -1$ yields the reverse Kullback–Leibler divergence $\mathcal{E}_{-1}(m) = -\mathbb{E}[\log m]$, which reverses the role of the benchmark and alternative models. 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.4 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 (6) leads, in the case when utilitarian concerns are absent in the right tail of the productivity distribution, to the same optimal tax formula (17) but the worst-case distortion (18) is now given by

$$m(z) = \left[\frac{\eta}{\theta}\left(\chi - \mu T(y(z))\right)\right]^{\frac{1}{\eta}} \tag{35}$$

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. The penalty parameter θ can again be calibrated to achieve a desired size of the power divergence counterpart of the entropy ball (4) that represents the set of distributions the planner views as plausible.

Specializing to the case of quasilinear utility and Pareto-distributed productivity under the benchmark distribution, we derive the counterpart of the differential equation for the optimal marginal tax (27), leading to the following result.

Result 3.5. *Assume that worker's preferences are given by the quasilinear form (13), 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*

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

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

$$\lim_{y\to\infty} T'(y) = \tau_\eta = \frac{1+\gamma}{1+\gamma+\widetilde{\alpha}} \qquad \text{with } \widetilde{\alpha} = \alpha - \frac{1+\gamma}{\gamma}\frac{1}{\eta} > \alpha. \tag{37}$$

Expressions (36) and (37) 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.4. To provide intuition, consider first the case $\eta > 0$. In this case, the worst-case distortion formula (35) 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. In fact, we show in the appendix that the optimal marginal tax reaches zero for a finite income threshold $\bar{y}$.

When $\eta < 0$, expression (35) 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 (20) and (21) implies that the worst-case distortion (35) 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 approximate the labor income distribution in SCF 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

$$f_x\left(x; \bar{\mu}, \sigma, \alpha\right) = \alpha e^{\frac{\alpha}{2}\left(2\bar{\mu} + \alpha\sigma^2 - 2x\right)} \left[1 - \Phi\left(\frac{\bar{\mu} + \alpha\sigma^2 - x}{\sigma}\right)\right] \tag{38}$$

where $\Phi\left(x\right)$ is the standard normal cumulative distribution function. This distribution implies that productivity $z = \exp\left(x\right)$ has support $(0, \infty)$, 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.[6] We denote the class of EMG distributions as $\boldsymbol{F}_{\text{emg}}$. 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 (28), with parameters $\rho = 1$ implying logarithmic utility from consumption, $v = 1$, 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 (28), the right tail of the income distribution inherits the Pareto property, with shape parameter $\frac{\rho+\gamma}{1+\gamma}\alpha$. Combined with logarithmic utility, $\rho = 1$, income is asymptotically Pareto distributed with shape parameter α.

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 use quantiles of U.S. income distributions for years 1966–

[6] 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.

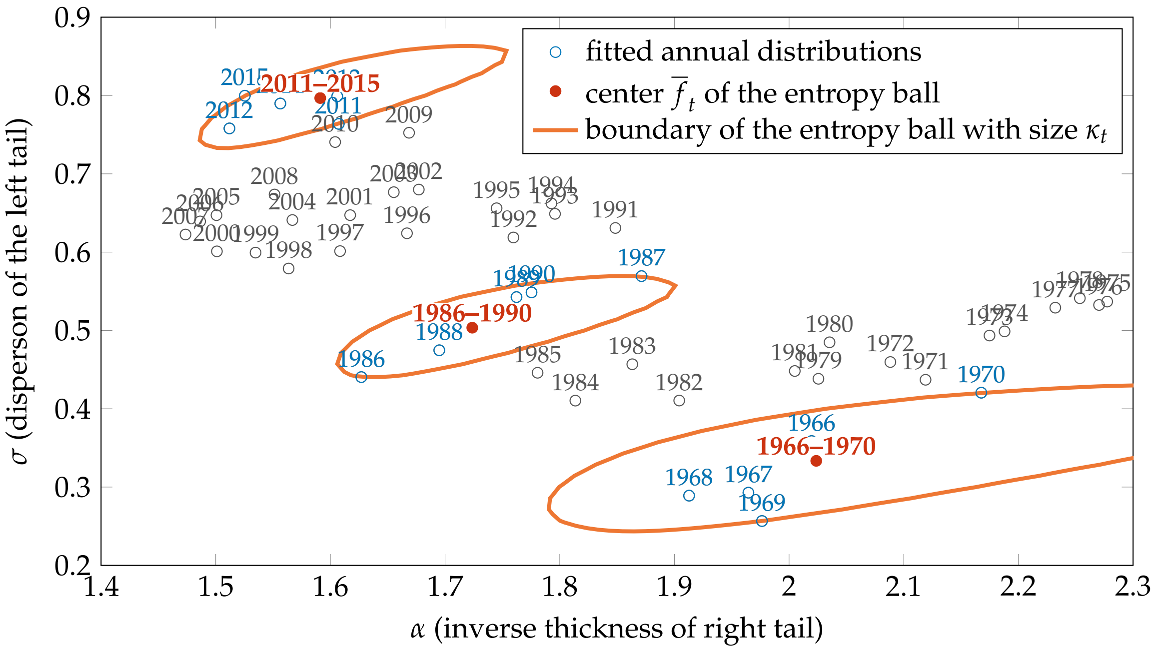

Figure 2: Estimated income distributions for U.S. data, 1966–2015. Each point shows the combination of parameters (α_t, σ_t) of the estimated EMG distribution for the given year. The orange curves represents three example entropy balls, centered at the solid red dots, for periods 1966–1970, 1986–1990,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

2015 from the World Inequality Database. We use fiscal income as our measure of income y_t, defined as the total income reported on tax returns before deductions, including labor, capital, and mixed income. 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.[7] We 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 parametric class as well as a much larger nonparametric set of distributions $\widetilde{f}$ that all satisfy $\mathcal{E}\left(\overline{f}_t, \widetilde{f}\right) \leq \kappa_t$.

[7] Estimation of the EMG distributions is described in Appendix E.

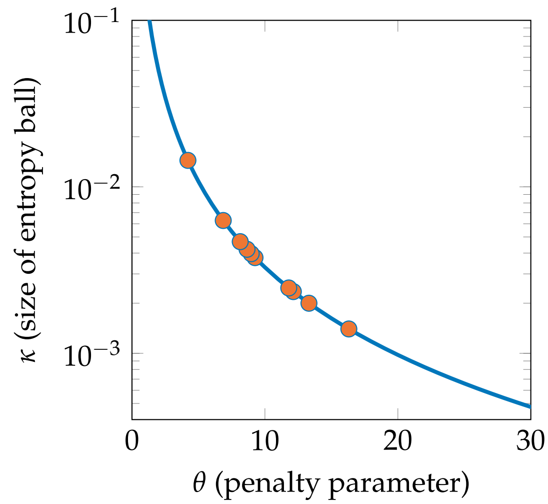

Figure 3: Relationship between the penalty parameter θ and entropy κ for the model calibrated in Section 4.1.

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 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 parametric class.

Our procedure delivers a κ_t for each 5-year window. 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 across all windows, corresponding to $\theta = 9.254$ and $\theta = 16.312$, respectively. In order to provide a conservative calibration of the misspecification concerns, we consider the smallest entropy ball size with $\theta = 16.312$ as our baseline specification. 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*

$$\lim_{y \to \infty} T'(y) = \frac{(\gamma + \rho)(1 + \gamma)}{\alpha(\gamma + \rho) + \gamma(1 + \gamma)}. \tag{39}$$

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

$$\lim_{z \to \infty} T'(y) = 0. \tag{40}$$

$$\lim_{y \to \infty} \frac{d \log T'(y)}{d \log y} = \min\left(-\frac{1}{2}, \rho - 1\right) \tag{41}$$

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.5. The first result represents the Mirrlees (1971) and Diamond (1998) analytical tax formula that also 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 (33) is bounded and bounded away from zero in the 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 by a finite $m(z)$ will not alter the zero limiting marginal tax rate in the model with misspecification concerns.

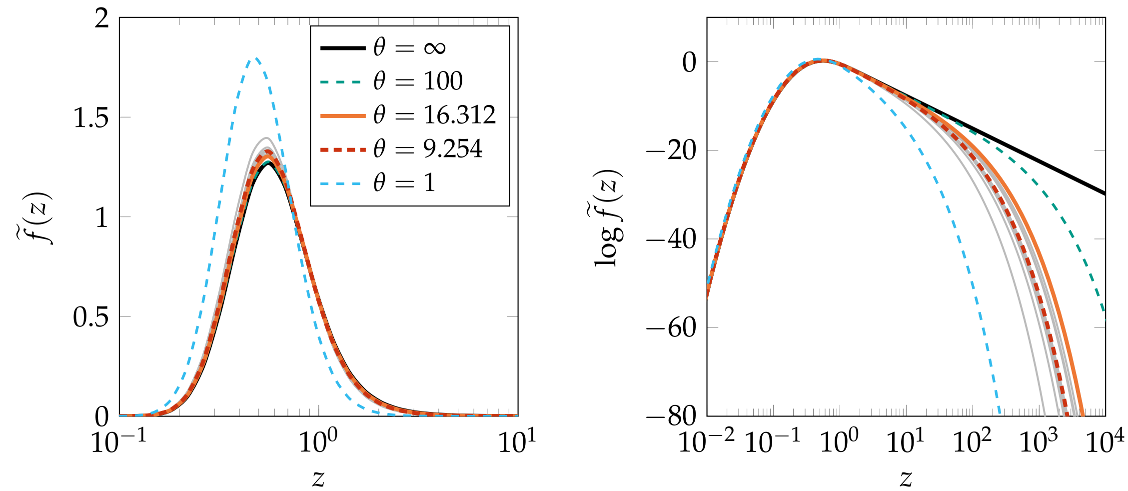

Figure 4: Worst-case distributions $\widetilde{f}(z)$ for alternative levels of misspecification concerns given by θ. The black solid line for $\theta = \infty$ corresponds to the rational benchmark for which $\widetilde{f}(z) = f(z)$. Thin grey lines represent the densities for the sizes of entropy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

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 then effectively determines the range of incomes at which the concerns start to manifest themselves in a more pronounced way.

Table 1 tabulates the quantiles of productivity z under the alternative worst-case distributions $\widetilde{f}(z)$. Misspecification concerns increase as we move through the columns from left to right. The table confirms the information from Figure 4. The pronounced effects of the misspecification concerns materialize in the right tail, while the median of the worst-

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	0.735	0.725	0.582
$\widetilde{q}(0.75)$	1.098	1.093	1.067	1.046	0.784
$\widetilde{q}(0.95)$	2.317	2.288	2.159	2.063	1.267
$\widetilde{q}(0.99)$	4.817	4.680	4.159	3.825	1.864
$\widetilde{q}(0.999)$	13.718	12.661	9.780	8.394	2.994

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.

case productivity distribution $\widetilde{q}\,(0.50)$ gets noticeably distorted only as misspecification concerns become substantial ($\theta = 1$), and the left tail remains essentially undistorted.

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 $\lim_{z\to\infty} T'\left(y\left(z\right)\right)$ is positive and quantitatively large, at 71.4%.[8] 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 looks notably different. While for incomes close to the average income, the optimal marginal tax looks similar to that under the rational case, it starts departing quickly for higher income levels. For the baseline case of $\theta = 16.312$ depicted with the solid orange line, the marginal tax peaks at 57.5% for incomes equal to 7.5 times the average income, and starts declining thereafter. At incomes corresponding to 20 times the average income, the difference between the rational and robust optimal marginal tax is 13 percentage points, and climbs to 29 percentage points for very high-income individuals with 100 times the average income.

Figure 5 also depicts as gray lines the optimal marginal tax rates for other sizes of entropy balls obtained in Section 4.2, with the median calibration of $\theta = 9.254$ displayed as the dashed red line. Since our baseline calibration is conservative, these gray lines deviate more strongly from the rational case. Nevertheless, they all remain in a plausibly tight band in the proximity of the orange baseline. Only when misspecification concerns are calibrated as counterfactually small ($\theta = 100$) or large ($\theta = 1$), the resulting optimal tax schedule deviates substantially from the baseline case.

Theoretical results from Section 3 show that not only should marginal tax rates decline

[8] 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.

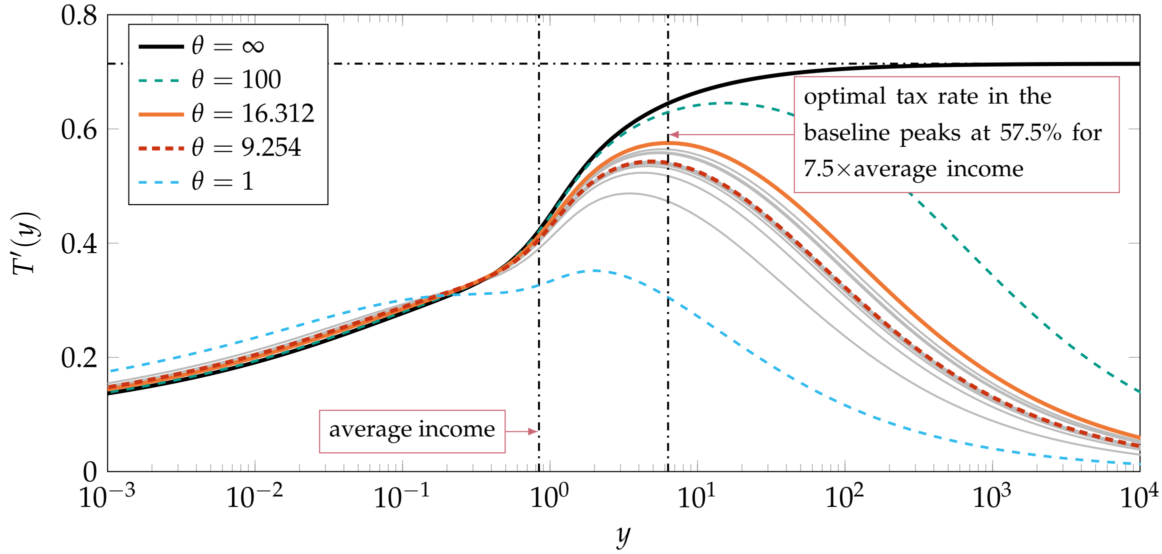

Figure 5: Optimal marginal tax schedules for alternative levels of misspecification concerns. The horizontal dashdotted line corresponds to the limiting marginal tax rate for the rational case. Vertical lines refer to income levels under the baseline model with $\theta = 16.312$.

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. Going from 20 to 100 times the average income, the marginal tax rate decreases from 55% to 40%, reflecting the average elasticity of -0.2 in this range.

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

$$m(z) = \bar{m} \exp\left(-\frac{1}{\theta}\left[\mathcal{U}(z) + \mu T(y(z))\right]\right). \tag{42}$$

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

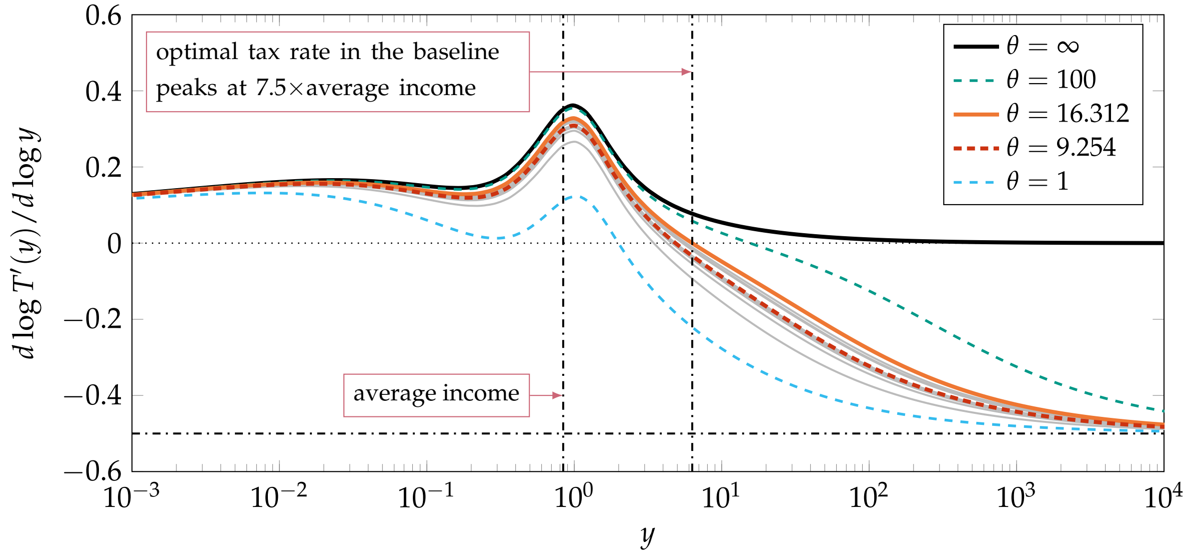

Figure 6: Growth rates of optimal marginal tax schedules, $d\log T'(y)/d\log y$, for alternative levels of misspecification concerns. The dashdotted horizontal line corresponds to the theoretical limit, equal to $-1/2$. Vertical lines refer to income levels under the baseline model with $\theta = 16.312$.

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 3.4.2, when the utility function is sufficiently concave ($\rho > \frac{1}{2}$), budgetary concerns dominate the shape of the distortion (42) in the right tail of the productivity distribution. The critical observation is that the misspecification concern given by (42) increases with z. The planner therefore fears that the attrition in the mass of agents above z who generate the revenue is stronger *relative* to the attrition of 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 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 baseline 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.

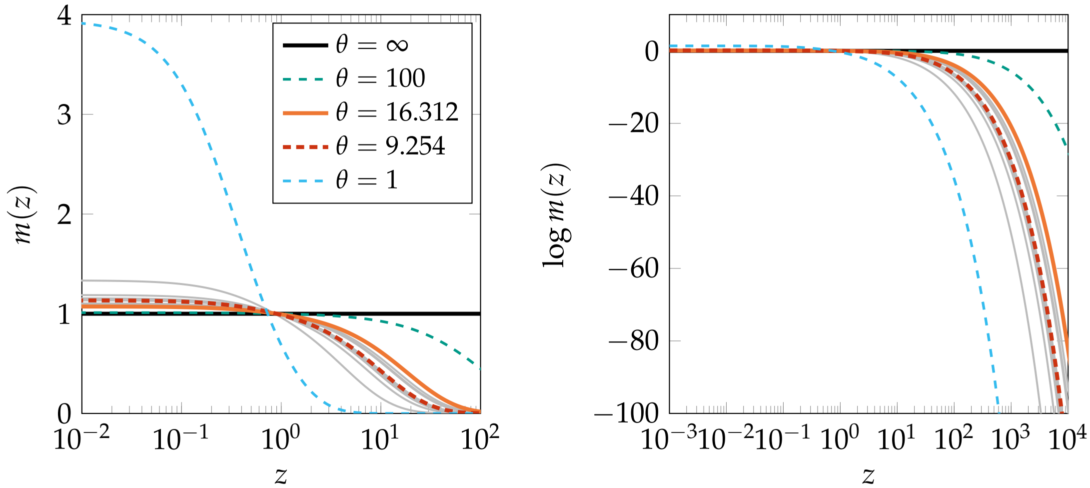

Figure 7: The likelihood ratio $m(z) = \widetilde{f}(z)/f(z)$ representing the distortion of the worst-case distribution relative to the benchmark distribution, plotted for alternative levels of misspecification concerns given by θ. The case $\theta = \infty$ corresponds to the rational benchmark for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

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 levels for high-productivity workers, as shown in the 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

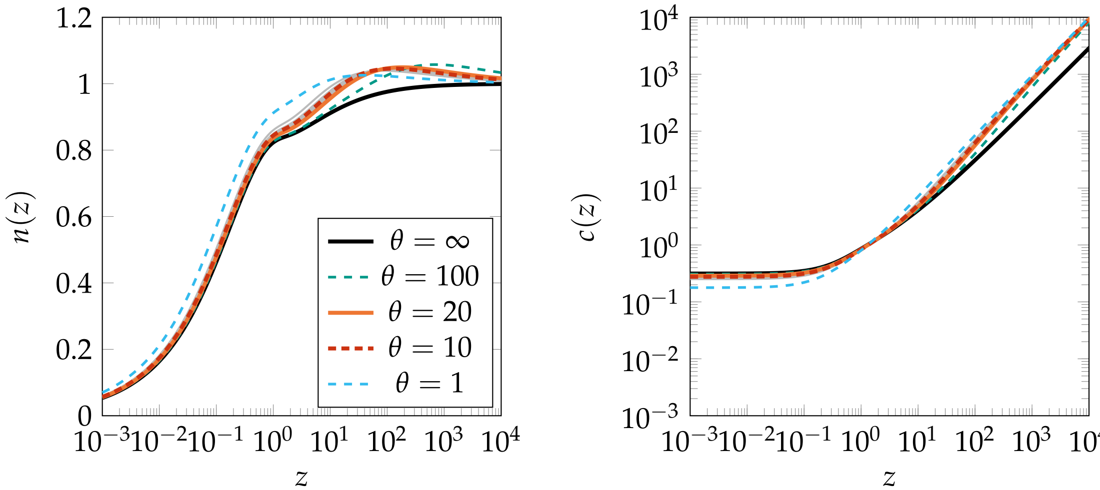

Figure 8: Labor supply under optimal tax schedules (left panel), and consumption as a function of income (right panel) under alternative levels of misspecification concerns.

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 EMG 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 taxable resources.

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 under the worst-case model $\widetilde{\mathbb{E}}[y]$, is roughly 37%, and largely stable across a range of the values of the parameter θ, reflecting the desire of the robust planner to keep low-income households well-insured. 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

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
$T_0/\widetilde{\mathbb{E}}[y]$	−0.383	−0.378	−0.367	−0.359	−0.307
$\max_y T'(y)$ (%)	71.4	64.5	57.5	54.3	35.2
$\left(\arg\max_y T'(y)\right)/\mathbb{E}[y]$	∞	18.808	7.504	5.715	2.169
$\mathbb{E}[T]/\mathbb{E}[y]$	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.

changes substantially. For the case without misspecification concerns, the top marginal tax rate asymptotes at 71.4%, while for the baseline 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.

Given the calibration of the marginal value of government spending $\bar{v}$, we obtain that the subjective tax revenue $\widetilde{\mathbb{E}}[T] = 0$ for all choices of θ by construction. 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 $\mathbb{E}[T]$ 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 aggregate income in the economy.

If we interpret the benchmark distribution as the true productivity distribution realized ex-post under the data-generating measure, then the optimal tax policy of the robust planner generates higher government spending than what the government anticipated under the worst-case model. Our model is static so it does not speak to intertemporal tradeoffs but a dynamic extension of this model can alternatively view this extra surplus as surprise tax revenue that can be carried over to the next period. The robust planner then must manage 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 consider 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.

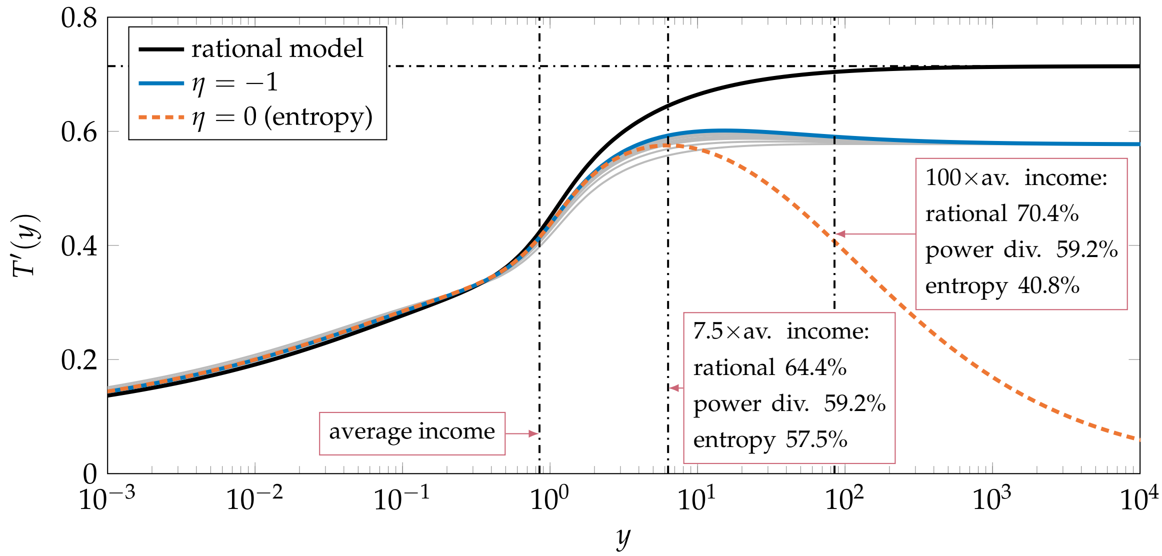

Figure 9: Optimal marginal tax schedules for the power divergence penalty with $\eta = -1$. The blue line corresponds to the tax schedule for the baseline conservative case with the smallest divergence ball (achieved for $\theta = 14.246$), and the gray lines for the other estimated ball sizes. The black line represents the tax schedule for the rational case, and the orange dashed line for the baseline case under entropy penalty from Figure 5. The horizontal dashdotted line corresponds to the limiting marginal tax rate for the rational case, and the vertical lines refer to income levels under the baseline model.

4.6 Shape restrictions on the tail of the productivity distribution

The entropy constraint used in the quantitative evaluation leads to worst-case distributions that are ultimately thin-tailed in the right tail. Here, we want to consider a case where the planner is concerned about misspecification but is confident that plausible type distributions have an asymptotic Pareto tail with a shape parameter within a given range.

We utilize the general class of power divergences analyzed in Section 3.4.3. Insights from that section carry over to the case of concave preferences we use for the quantitative evaluation. If the benchmark distribution is Pareto with tail parameter α, and the power divergence constraint has a negative exponent η, then the worst-case distribution is asymptotically Pareto with a thinner tail parameter $\widetilde{\alpha}$. Specifically, equations (37) and (39) combine to

$$\lim_{y\to\infty} T'(y) = \frac{(\gamma+\rho)(1+\gamma)}{\widetilde{\alpha}(\gamma+\rho)+\gamma(1+\gamma)} \qquad \widetilde{\alpha} = \alpha - \frac{1+\gamma}{\rho+\gamma}\frac{1}{\eta}. \tag{43}$$

We use the same parameterization of concave preferences and the EMG benchmark distribution as in Section 4.1. The power divergence exponent η determines the asymptotic tail parameter $\widetilde{\alpha}$ of the worst-case distribution. We pick $\eta = -1$, which corresponds to the case of reverse Kullback–Leibler divergence, and implies $\widetilde{\alpha} = 3.2$. The distance from the benchmark tail parameter $\alpha = 2.2$ corresponds to the range of historical values observed in our

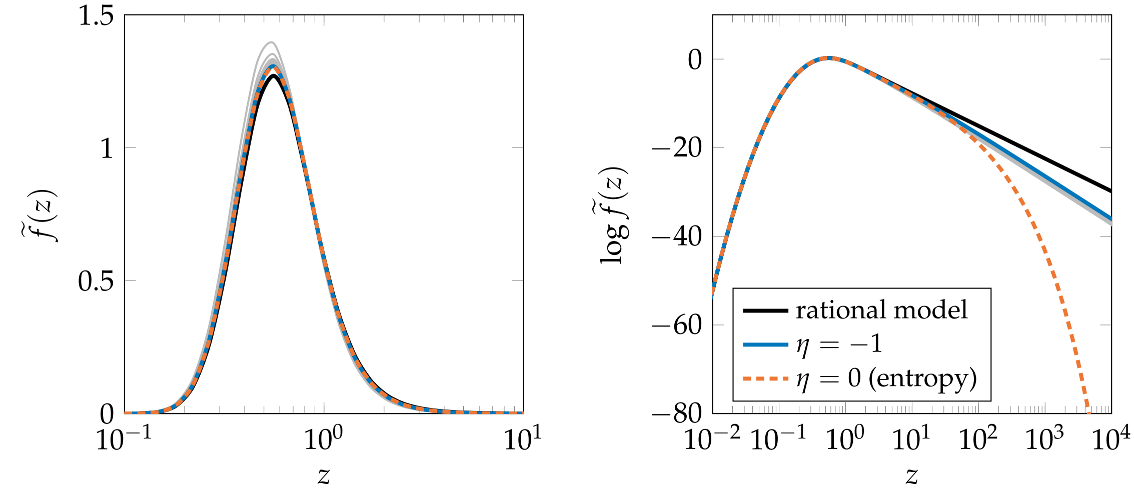

Figure 10: Worst-case distributions $\widetilde{f}(z)$ for the power divergence penalty with $\eta = -1$. The blue line corresponds to the baseline conservative case with the smallest divergence ball (achieved for $\theta = 14.246$), and the gray lines for the other estimated ball sizes. The black line represents the tax schedule for the rational case, and the orange dashed line for the baseline case under entropy penalty.

sample, and to time variation in measured tail parameters in other studies, for example, Vries and Toda (2022).

We then repeat the quantification of misspecification concerns from Section 4.2, construct the sizes κ_t of power divergence balls that encompass the time-variation of the distributions within 5-year windows, and convert them to the corresponding penalty parameters θ_t. We again pick the smallest divergence ball as our conservative baseline.

Figure 9 compares the resulting optimal marginal tax schedules with the rational case and the baseline entropy penalty from Figure 5. The baseline optimal tax schedule shown as the blue line asymptotes at 57.7%, in line with formula (43). The thin gray lines, which represent tax schedules for the alternative sizes of the estimated divergence balls, are concentrated in a tight band.

The figure shows that the tax schedule is very similar to that for the entropy constraint shown as the orange dashed line for incomes up to about 7.5 times the average income, where the entropy-constrained tax schedule peaks. At that point, the schedule already implies notably lower tax rates than the rational case. For higher income levels, the tax schedules start diverging, with the power divergence-constrained schedule essentially flattening out. At 100 times the average income, the power-divergence schedule is about half way between the rational and entropy constrained case.

Figure 10 compares the corresponding worst-case densities. The left panel shows that the alternative sizes of the power divergence balls lead to modest differences in the shape of the worst-case density in the vicinity of its mode, which also has only a modest impact on the optimal tax schedule. As expected, the main difference appears in the right panel, where the power divergence penalty constrains the worst-case distribution to be fat-tailed, compared to the thin-tailed worst-case distribution for the entropy penalty case.

These results confirm the general tendency that misspecification concerns tend to lower optimal marginal taxes for higher incomes. While using the power divergence penalty with $\eta < 0$ dictates a strictly positive limiting marginal tax as $y \to \infty$, the broad results quantitatively agree with the entropy case for a substantial share of the income distribution.

5 Multidimensional misspecification concerns

A crucial input for the design of optimal taxes, in addition to the skill distribution, are labor supply preferences which pin down the elasticity of taxable income. However, it is well-known in the public finance literature that these elasticities vary substantially across applications depending on the studied context, analyzed population, and applied econometric methods. Moreover, very little is known about the heterogeneity of these elasticities across incomes.[9]

We formalize this lack of knowledge as the planner's concern regarding the joint distribution of parameters that govern labor supply elasticity and productivity. We propose a generalization of the entropy penalty from Section 2 to allow for a flexible parametrization of the concern across the two dimensions of heterogeneity. We then design optimal nonlinear taxes that are robust to such multidimensional misspecification, and investigate the quantitative relevance of this source of misspecification in an application calibrated to the uncertainty reported in existing empirical studies.

5.1 Generalizing the entropy penalty

Household type (γ, z) now includes an additional parameter γ that determines labor supply elasticity. Types are private and distributed according to a joint density $f(\gamma, z)$. We denote $f_z(z)$ the marginal distribution of productivities, and $f_{\gamma|z}(\gamma|z)$ the conditional distribution of γ.

The government is concerned that $f(\gamma, z)$ is misspecified and contemplates a set of alternatives $\widetilde{f}(\gamma, z)$. Using two positive scalars $(\theta_\gamma, \theta_z)$, we define a penalty function for

[9] Chetty et al. (2011), Saez et al. (2012), and Neisser (2021) survey the literature on the elasticity of taxable income, documenting how estimates vary across regression techniques, sample restrictions, countries, and time. Lockwood et al. (2021) run a survey among academic economists to elicit subjective belief distributions of the elasticity.

the multidimensional case as

$$\mathcal{P}\left(f,\widetilde{f}|\theta_\gamma,\theta_z\right)=\theta_\gamma\int\mathcal{E}\left(f_{\gamma|z}\left(\cdot|z\right),\widetilde{f}_{\gamma|z}\left(\cdot|z\right)\right)\widetilde{f}_z\left(z\right)dz+\theta_z\mathcal{E}\left(f_z,\widetilde{f}_z\right)$$

where $\mathcal{E}(\cdot,\cdot)$ is the relative entropy defined in equation (3). This generalization of the penalty function flexibly captures differences in the degree of misspecification concerns about the conditional distribution of labor supply elasticity, represented by the first term, and the misspecification concerns about the productivity distribution, captured by the second term. 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. When $\theta_\gamma=\theta_z=\theta$, the penalty function reduces to relative entropy $\theta\mathcal{E}\left(f,\widetilde{f}\right)$.

The planner solves the multidimensional counterpart of problem (6):

$$\max_{T}\min_{\substack{m>0\\ \mathbb{E}[m]=1}}\mathbb{E}\left[m\psi\mathcal{U}\right]+V\left(\mathbb{E}\left[mT\left(\mathcal{Y}\right)\right]\right)+\mathcal{P}\left(f,\widetilde{f}|\theta_\gamma,\theta_z\right) \tag{44}$$

where $m\left(\gamma,z\right)=\widetilde{f}\left(\gamma,z\right)/f\left(\gamma,z\right)$ is the likelihood ratio between the benchmark and the alternative distribution of (γ,z). In the limiting case when $\theta_\gamma\to\infty$, the distortion in the conditional distribution vanishes, and $\widetilde{f}_{\gamma|z}\left(\gamma|z\right)=f_{\gamma|z}\left(\gamma|z\right)$. This can be seen from the optimality condition of the minimizing agent that yields the worst-case distortion of the conditional distribution

$$m_{\gamma|z}\left(\gamma|z\right)=\frac{\widetilde{f}_{\gamma|z}\left(\gamma|z\right)}{f_{\gamma|z}\left(\gamma|z\right)}=\bar{m}_{\gamma|z}\left(z\right)\exp\left(-\frac{1}{\theta_\gamma}\left[\psi\left(\gamma,z\right)\mathcal{U}\left(\gamma,z\right)+\mu T\left(y\left(\gamma,z\right)\right)\right]\right) \tag{45}$$

where $\bar{m}_{\gamma|z}\left(z\right)$ is a normalization constant. Similarly, as $\theta_z\to\infty$, the minimizing probability measure is not distorted in terms of the marginal productivity distribution, $\widetilde{f}_z\left(z\right)=f_z\left(z\right)$. Details of the calculations are provided in Appendix F.

5.2 Quantitative application

We calibrate the model to study the role of misspecification related to the joint distribution of productivity and labor supply elasticity. We impose additional structure to isolate the effect of uncertainty over labor supply responses to emphasize economic forces that are new relative to those studied in Section 4.

We envision a government that contemplates a policy reform from a status quo tax function T^0 to a new optimal tax function T^1, and is uncertain about the labor supply responses of individual households. Equation (45) emphasizes that misspecification concerns for a robust planner are driven by utility and budgetary concerns. To focus exclusively on the concern about the responsiveness of labor supply to tax changes, we require that at the status quo tax function T^0, utility levels and budgetary contributions are independent of distribution of labor supply preferences and thus there is no reason for the minimizing agent to distort the conditional distribution of elasticity. Formally, we obtain this by adding shifters to the labor supply preferences so that for a given productivity level, the optimal choices of household consumption, hours worked, and the resulting utility levels do not depend on differences in labor supply preferences if there is no change to the tax function relative to the status quo tax function.[10] For more details see Appendix F.

5.2.1 Calibration

We next describe the functional forms for the utility function and the benchmark type distribution, and discuss how to calibrate the penalty parameters. Most of the remaining parameters are the same as in Section 4.

We use Greenwood et al. (1988) (GHH) preferences of the form

$$\frac{1}{1-\rho}\left(c-\Psi\left(\gamma,z\right)\frac{n^{1+\gamma}}{1+\gamma}\right)^{1-\rho}+\Delta\left(\gamma,z\right), \tag{46}$$

where the shifters $\Psi\left(\gamma,z\right)$ and $\Delta\left(\gamma,z\right)$ ensure that that the allocation and utility levels at the status quo tax function T^0 do not depend on the labor supply elasticity γ. Under GHH preferences, the elasticity of taxable income (ETI) is equal to γ^{-1}, making it easy to connect to empirical studies. The utility curvature is set to $\rho = 1$.

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., $f_{\gamma|z}\left(\gamma \mid z\right) = f_{\gamma}\left(\gamma\right)$. The benchmark marginal distribution of productivities $f_z\left(z\right)$ is modeled as exponentially modified Gaussian (EMG) with the same parameters as in Section 4. The benchmark distribution of ETI is obtained from Lockwood et al. (2021) who survey economists and elicit their subjective values of the uncompensated ETI. The survey reveals a 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 11.

Two penalty parameters θ_z and θ_γ are set as follows. For the exercise in this section, we set $\theta_z \to \infty$. It follows from the discussion in Section 5.1 that the worst-case marginal distribution of productivities satisfies $\widetilde{f}_z\left(z\right) = f_z\left(z\right)$. By doing so, we focus on concerns about labor supply responsiveness, and isolate results from those in Section 4 that studied

[10] These assumptions are similar to those in Lockwood et al. (2021) who use a Bayesian approach to study the role of uncertainty about labor supply elasticities.

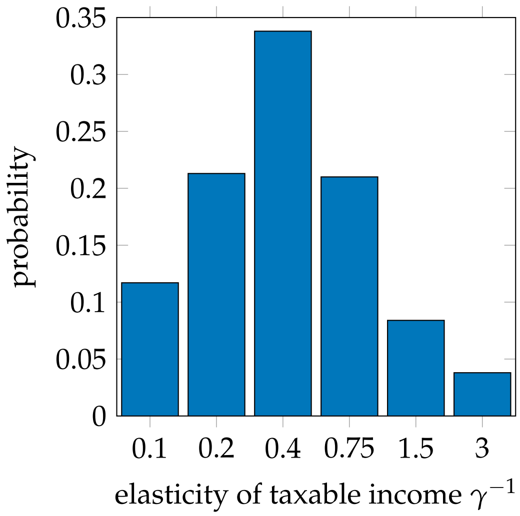

Figure 11: Benchmark distribution of elasticity of taxable income. We use a discretized distribution that is constructed using the histogram reported in Lockwood et al. (2021).

the consequences of distorting the productivity distribution.

The magnitude of the distortion of the labor supply elasticity distribution is determined by the parameter θ_γ, and varies endogenously with income level. Our calibration strategy is to pick θ_γ so that the worst-case distribution produces the elasticity for high-income earners reported in published empirical studies. For this purpose, we use Rauh and Shyu (2024), which is one of the few studies that estimate 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 points. We also discuss results for higher and lower values of θ_γ surrounding our preferred estimate.

Figure 12 plots the average ETI γ^{-1} conditional on productivity z under the planner's worst-case model against the mean of before-tax earnings y given productivity z. The black line represents the average elasticity in the rational model, and the orange line our preferred calibration. In the worst-case model, the ETI distribution is tilted toward higher values relative to the rational model, and this endogenous tilt becomes more pronounced for higher incomes. As a result, the average value of the elasticity increases with income. Rauh and Shyu (2024) estimate the ETI to be between 2.6 and 3.0 for taxpayers with an annual income of 5 million USD. Targeting those values, we pick $\theta_\gamma = 5$ for our baseline calibration, and report sensitivity to the choice of θ_γ in our results.

To compute the nonlinear 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 exam-

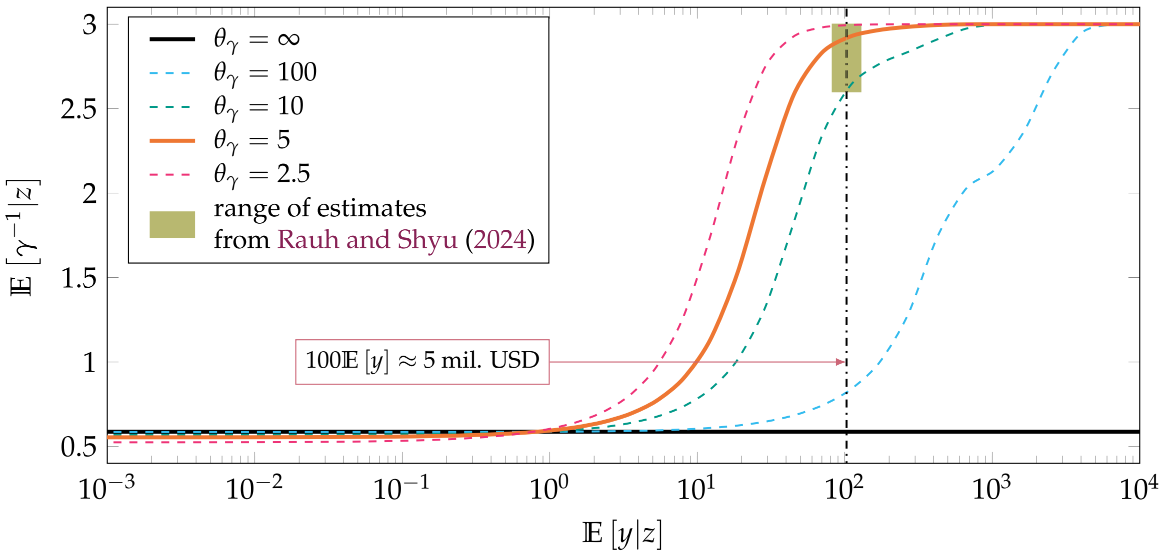

Figure 12: Mean elasticity of taxable income in worst-case models for alternative levels of multidimensional misspecification concerns. The income threshold represented by the vertical line refers to average income under the rational model.

ple, Golosov and Krasikov (2023) or Boerma et al. (2025)). Instead, we search for a tax schedule T within a flexible parametric class of functions such that the marginal tax rate $T'(y)$ is expressed through cubic basis functions of $\log y$, featuring N knots $(\log y_i, \tau_i)$, $i = 1, \ldots, N$, where $T'(y)$ remains constant outside the range of the knots. We choose $N = 3$. In Appendix G, we discuss the trade-offs involved in the parametrization, and benchmark the spline approximation against the Mirrlees solution for the one-dimensional case. We set $T^0 = 0$ as a baseline and present results for alternative choices of the status-quo tax function in Appendix H.

5.2.2 Results

Figure 13 plots the optimal marginal tax schedules for various values of θ_γ, with lower values indicating stronger concerns about the conditional distribution of γ. Consistent with the findings from the one-dimensional model, the figure shows that increased concerns result in lower marginal tax rates for high-income earners. For incomes below the average income, the difference in the optimal tax rate between the rational and robust case is negligible. At the preferred value $\theta_\gamma = 5$, the difference increases to 30.1 and 43.8 percentage points at 20 times and 100 times the average income, respectively. Asymptotically, the optimal top marginal tax rate $T'(y)$ in the baseline converges to 14.0%, compared to about 66.5% in the rational model.

The economic mechanism leading to lower top tax rates in this context is analogous to

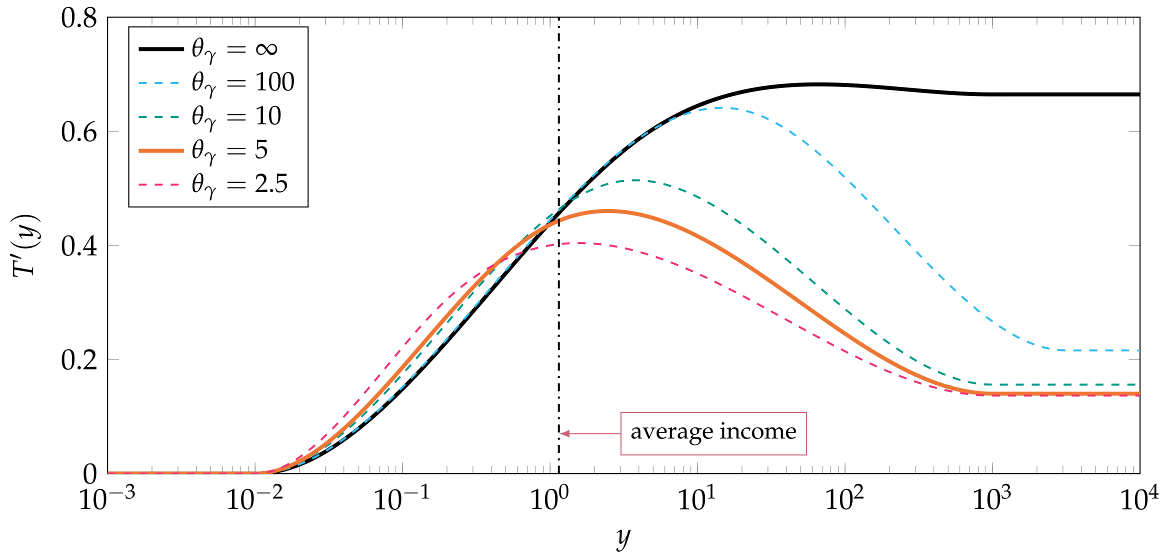

Figure 13: Optimal marginal tax schedules for alternative levels of multidimensional misspecification concerns. The black line represents the rational case, and the orange line the baseline calibration of the misspecification concerns. The dashed lines depict tax schedules for alternative choices of the penalty parameter θ_γ. The vertical dashdotted line refers to income levels under the baseline model.

those observed in the case of one-dimensional uncertainty regarding the productivity distribution that we studied in Section 4. In that scenario, the robust planner was concerned about the shape of the right tail of the productivity distribution and the ability to raise adequate revenues. As a result, the planner placed less emphasis on imposing distortionary taxes 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 imposing distortionary taxes on high-income earners in the form of reduced labor supply 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.[11]

Contrasting the worst-case conditional distribution of $\gamma^{-1}|z$ with the benchmark further clarifies this intuition. Recall that in the benchmark distribution, we assumed that γ^{-1} was distributed independently of the productivities z. In contrast, the worst-case joint

[11]The findings in Lockwood et al. (2021) that uncertainty about the elasticity of taxable income leads to more progressive taxation may appear to contradict our results. However, these differences stem from distinct thought experiments. Lockwood et al. (2021) compare optimal taxes between an economy with homogeneous elasticities and one with heterogeneous elasticities, where the planner fully trusts the elasticity distribution. Their heterogeneous elasticity economy serves as our starting point; we then introduce concerns about the shape of this distribution, which yields lower top marginal tax rates. See Appendix I for more details.

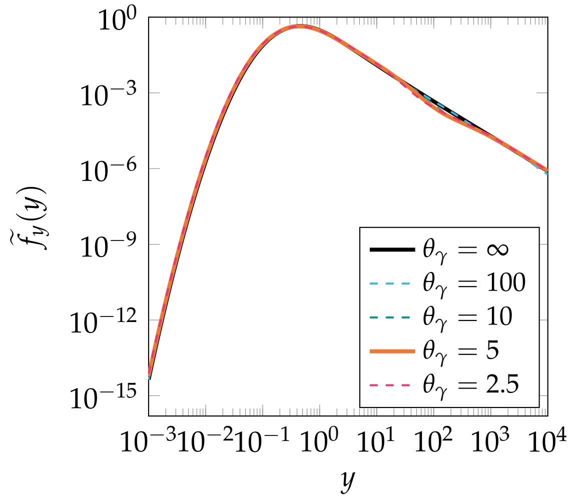

Figure 14: Worst-case density of before-tax income $\widetilde{f}_y(y)$ for alternative levels of multidimensional misspecification concerns.

density features an endogenous positive correlation between productivities and elasticities, reflecting the observation that distortionary taxation is most costly when high-income individuals have high elasticity of taxable income.

Interestingly, the resulting income distributions under the benchmark and the worst-case distributions do not differ significantly. As shown in Figure 14, 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. The worst-case and benchmark distributions are therefore essentially indistinguishable using the income data. 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 marginal taxes at the top.

In Table 3, we report additional model moments and examine how model-implied outcomes vary with the penalty parameter θ_γ. As θ_γ decreases from 100 to 1, the mean elasticity under the worst-case distribution for highly productive workers increases. We define highly productive workers as those with $z \geq \bar{z}$, where the threshold $\bar{z}$ is chosen such that, under the status quo tax function, these workers earn at least $100\mathbb{E}[y]$. The rise in worst-case elasticities for this group reflects the planner's heightened concern about behavioral responses among top earners. In contrast, other outcomes, such as average output, total transfers, and the marginal value of public goods, remain largely unchanged.

moments \ θ_γ	∞	100	10	5	2.5
$\widetilde{\mathbb{E}}[z]$	1.000	1.000	1.000	1.000	1.000
$\widetilde{\mathbb{E}}[1/\gamma]$	0.587	0.587	0.590	0.594	0.601
$\mathbb{E}[1/\gamma \mid z \geq \bar{z}]$	0.587	0.587	0.587	0.587	0.587
$\widetilde{\mathbb{E}}[1/\gamma \mid z \geq \bar{z}]$	0.587	1.119	2.746	2.956	2.997
$\mathbb{E}[y]$	1.029	1.067	1.121	1.147	1.176
$\widetilde{\mathbb{E}}[y]$	1.029	1.015	1.001	0.998	0.995
μ	1.374	1.444	1.554	1.613	1.684
$T_0/\widetilde{\mathbb{E}}[y]$	-0.412	-0.387	-0.356	-0.339	-0.321
$\mathbb{E}[T]/\mathbb{E}[y]$	0.000	0.017	0.033	0.038	0.042
$\widetilde{\mathbb{E}}[T]/\widetilde{\mathbb{E}}[y]$	0.000	0.000	0.000	0.000	0.000

Table 3: Moments under the benchmark and worst-case distributions. The productivity threshold $\bar{z}$ achieves income of $100\mathbb{E}[y]$ under the optimal tax schedule in the rational case.

6 Conclusion

The design of optimal tax schedules critically depends on assumptions about the distribution of taxpayer types. Even with abundant data, predicting this distribution over a fiscal planning horizon is challenging, particularly in the tails where sample sizes are thin and time-series variation is large.

This paper addresses that uncertainty by studying optimal taxation when the planner fears the income distribution may be misspecified. Our robust approach avoids restrictive parametric assumptions, and instead considers nonparametric misspecifications that the planner views as most consequential for welfare.

We find that misspecification concerns systematically reduce progressivity in optimal tax schedules, especially at the top of the income distribution. In exercises calibrated to observed uncertainty in the type distributions, optimal tax rates for low and middle incomes remain close to those in the benchmark without misspecification concerns. By contrast, top marginal tax rates diverge sharply, in some cases converging to zero regardless of the assumed type distribution and the details of the environment.

At root, the mechanism is budgetary arithmetic: fiscal revenues are concentrated among high-income earners, whose underlying attributes are difficult to measure. A tax system that is robust to this measurement uncertainty must weigh the risks of over-taxing the top against the need for revenue, leading to a more cautious assessment of high-income taxation.

Our analysis leaves important extensions for future work. To sharpen policy relevance, it will be valuable to embed these ideas in richer models of top incomes—such as entrepreneurial income and wealth taxation—or in multiperiod frameworks that combine income taxation with optimal debt management. Applying the robust approach to these settings is a natural next step for the theory of optimal taxation.

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

$$\max_{\{c_i,y_i\}_{i=1}^I \in \mathbb{X}} \min_{\{m_i\}_{i=1}^I \in \mathbb{Y}} \sum_{i=1}^I \pi_i m_i \left(u\left(c_i\right) - v\left(\frac{y_i}{z_i}\right)\right) + \bar{v}G + \theta \sum_{i=1}^I \pi_i m_i \log m_i, \tag{47}$$

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

$$\sum_{i=1}^I \pi_i m_i = 1, \qquad m_i \geq 0, \qquad i \in \{1,\dots,I\}, \tag{48}$$

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 (28). 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

$$u_i - v_i \geq u_j - \left(\frac{z_j}{z_i}\right)^{1+\gamma} v_j, \qquad i,j \in \{1,\cdots,I\}, \tag{49}$$

and the constrains in (48). 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 (49), and on a convex set $\mathbb{Y}$ imposed by linear constraints on m_i in (48). 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 $\rho < 1$, $u_i \in (-\infty, \infty)$ for $\rho = 1$ and

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

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

and notice that $U_i\left(u_i, v_i\right)$ 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 $\lim U_i\left(u_i, v_i\right) = -\infty$. Therefore, for any $\underline{U} \in \mathbb{R}$, the upper countour set

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

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 $u_i = u\left(\widetilde{c}\right)$, $v_i = v\left(0\right) = 0$. This allocation satisfies all constraints in (49), and hence is feasible. Further denote

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

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 (15)), can be written as

$$m_i = \bar{m}\exp\left(-\frac{1}{\theta}U_i\left(u_i, v_i\right)\right), \tag{50}$$

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

Take any allocation $\{u_i, v_i\}_{i=1}^{I} \notin \mathbb{X}$ and, without loss of generality, order indices i by increasing $U_i\left(u_i, v_i\right)$. Then first $J \geq 1$ indices are such that $U_i\left(u_i, v_i\right) \leq \underline{U}$ and hence $\left(u_i, v_i\right) \notin \overline{\mathbb{W}}_i$. Further, the associated m_i given by (50) 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

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

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 $i \leq J$, $U_i\left(u_i, v_i\right) < \underline{U}$ and for any i, $U_i\left(u_i, v_i\right) \leq \overline{U}$, 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 (50) 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 (47) 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

$$\frac{dU}{dz} = -U_n\left(c(z), \frac{y(z)}{z}\right) \frac{y(z)}{z^2}. \tag{51}$$

and the budget constraint

$$G = \int_{\underline{z}}^{\bar{z}} (y(z) - c(z))\, m(z) f(z)\, dz. \tag{52}$$

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

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

Here, χ and μ are multipliers on the constraints $\mathbb{E}[m] = 1$ and (52), 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

$$\begin{aligned} 0 = H_y =& \mu\left[1 - C_y(U(z), y(z))\right] m(z) f(z) - \lambda(z) U_n\left(C(\mathcal{U}(z), y(z)), \frac{y(z)}{z}\right) \frac{1}{z^2} \\ & - \lambda(z) \left[U_{nc}\left(C(\mathcal{U}(z), y(z)), \frac{y(z)}{z}\right) C_y(\mathcal{U}(z), y(z)) + U_{nn}\left(C(\mathcal{U}(z), y(z)), \frac{y(z)}{z}\right) \frac{1}{z} \right] \frac{y(z)}{z^2}, \end{aligned} \tag{53}$$

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

$$0 = H_m = \psi(z) \mathcal{U}(z) f(z) + \theta\left[\log m(z) + 1\right] f(z) - \chi f(z) + \mu\left[y(z) - C(\mathcal{U}(z), y(z))\right] f(z), \tag{54}$$

and the costate dynamics restriction yields

$$\frac{d\lambda(z)}{dz} = -H_U = [\mu C_U(\mathcal{U}(z), y(z)) - \psi(z)]\, m(z) f(z) + \\ + \lambda(z) U_{nc}\left(C(\mathcal{U}(z), y(z)), \frac{y(z)}{z}\right) \frac{y(z)}{z^2} C_U(\mathcal{U}(z), y(z)). \tag{55}$$

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 (54) 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 (15). Integrating up condition (55) 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:

$$\mu = \frac{\int_{\underline{z}}^{\bar{z}} \psi(z) m(z) f(z)\, dz - \int_{\underline{z}}^{\bar{z}} \lambda(z) U_{nc}\left(C(\mathcal{U}(z), y(z)), \frac{y(z)}{z}\right) \frac{y(z)}{z^2} C_U(\mathcal{U}(z), y(z))\, dz}{\int_{\underline{z}}^{\bar{z}} C_U(\mathcal{U}(z), y(z))\, m(z) f(z)\, dz}. \tag{56}$$

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 (55) on $(z, \bar{z})$. For that purpose, denote the terms in (55) 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 (55) 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,

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

Further, we utilize the following notational simplifications:

$$\begin{aligned} \left(1 - T'(y(z))\right) z &= -\frac{U_n(z)}{U_c(z)} \\ C_U(\mathcal{U}(z), y(z)) &= \frac{1}{U_c(z)} \\ C_y(\mathcal{U}(z), y(z)) &= -\frac{1}{z}\frac{U_n(z)}{U_c(z)}. \end{aligned} \tag{57}$$

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 $C(\mathcal{U}(z), y(z))$. With this notation and substitutions, we can reorganize the optimality condition (53) as

$$-\mu\left[T'(y(z))\right] f(z) = \frac{\lambda(z)}{z} U_c(z)\left(1 - T'(y(z))\right)\left[1 + n(z)\frac{U_{nn}(z) + w(z)U_{nc}(z)}{U_n(z)}\right]. \tag{58}$$

It is convenient to rewrite this expression in terms of labor supply elasticities. Denoting $w = (1 - T'(zn)) z$ 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 (57) 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 (58), we obtain

$$\frac{T'(y(z))}{1 - T'(y(z))} = -\frac{1 + \varepsilon^u(z)}{\varepsilon^c(z)}\frac{\lambda(z)}{zm(z)f(z)}\frac{U_c(z)}{\mu}. \tag{59}$$

To solve for the Lagrange multiplier μ, simplify equation (56), 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

$$H_{U,S}(z) = \left[\frac{\mu}{U_c(z)} - \psi(z) \right] m(z) f(z)$$

$$H_{U,N}(z) = \frac{n(z) U_{nc}(z)}{zU_c(z)}$$

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

For the case of isoelastic separable preferences (28), $U_{nc}(z) = 0$, $H_{U,N}(z) = 0$, and

$$\frac{1 + \varepsilon^u(z)}{\varepsilon^c(z)} = 1 + \gamma.$$

This yields the marginal social value of public funds in the form of inverse marginal utility formula evaluated under the worst-case distribution

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

Recall that throughout the derivation, $c(z) = y(z) - T(y(z))$. Consequently,

$$\lambda(z) = \int_z^{\bar{z}} \psi(\zeta) m(\zeta) f(\zeta) \, d\zeta - \mu \int_z^{\bar{z}} (c(\zeta))^{\rho} m(\zeta) f(\zeta) \, d\zeta = \bar{\psi} \left[\widetilde{F}_{\rho}(z) - \widetilde{\Psi}(z) \right],$$

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, with densities

$$\widetilde{f}_{\rho}(z) = \left(\widetilde{\mathbb{E}}[c^{\rho}] \right)^{-1} c(z)^{\rho} m(z) f(z) = \bar{u}^{-1} c(z)^{\rho} m(z) f(z) \tag{60}$$

$$\widetilde{\psi}(z) = \left(\widetilde{\mathbb{E}}[\psi] \right)^{-1} \psi(z) m(z) f(z) = \bar{\psi}^{-1} \psi(z) m(z) f(z).$$

The optimal tax formula (59) then simplifies to

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

For the case of quasilinear utility (13), we have $\widetilde{F}_{\rho}(z) = \widetilde{F}(z)$, and obtain expression (16). For concave preferences without utilitarian concerns for the right tail of the productivity distribution as in Section 3.4.1, we have $\widetilde{\Psi}(z) = 1$, which yields expression (29). Expression (34) 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:*

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

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 (13) implies labor supply $n(z)$ given by

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

which can be rewritten as

$$y(z) = \left(1 - T'(y(z))\right)^{\frac{1}{\gamma}} z^{\frac{1+\gamma}{\gamma}}. \tag{62}$$

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

$$\begin{aligned}\frac{1-\widetilde{F}_1(z)}{z\widetilde{f}_1(z)} &= \frac{\int_z^{\bar{z}} m_1(\zeta)f(\zeta)\,d\zeta}{zm_1(z)f(z)}=\frac{\int_z^{\bar{z}}\frac{m_1(\zeta)}{m_1(z)}f(\zeta)\,d\zeta}{zf(z)} && (63)\\ &< \frac{\int_z^{\bar{z}}\frac{m_2(\zeta)}{m_2(z)}f(\zeta)\,d\zeta}{zf(z)}=\frac{\int_z^{\bar{z}} m_2(\zeta)f(\zeta)\,d\zeta}{zm_2(z)f(z)}=\frac{1-\widetilde{F}_2(z)}{z\widetilde{f}_2(z)}, && (64)\end{aligned}$$

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*

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

Then

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

Proof. By direct computation:

$$\begin{aligned}\frac{d}{dz}\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)} &= \frac{-\widetilde{f}(z)\,z\widetilde{f}(z)-\left(1-\widetilde{F}(z)\right)\left(\widetilde{f}(z)+z\widetilde{f}'(z)\right)}{\left(z\widetilde{f}(z)\right)^2}\\ &= -\frac{1}{z}-\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)}\left[\frac{1}{z}+\frac{\widetilde{f}'(z)}{\widetilde{f}(z)}\right]=-\frac{1}{z}-\widetilde{\phi}(z)\left[\frac{1}{z}+\frac{d}{dz}\log\widetilde{f}(z)\right].\end{aligned}$$

Using $\widetilde{f}(z)$ from expression (19), 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 (17) is given by

$$\frac{T'(y(z))}{1-T'(y(z))}=(1+\gamma)\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)}=(1+\gamma)\,\widetilde{\phi}(z)\,. \tag{65}$$

The single-crossing property (11) 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 (19) is strictly decreasing, and, by Assumption 2, $\widetilde{f}(z)$ is also strictly decreasing for sufficiently large z. Lemma C.2 then implies that

$$\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)} < \frac{1-F(z)}{zf(z)}, \tag{66}$$

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 (66), 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 (62) implies that $\lim_{z\to\infty} y(z) = \infty$.

From the optimal tax formula (65), we obtain $1-T'(y(z)) = \left(1+(1+\gamma)\widetilde{\phi}(z)\right)^{-1}$, and then we can rewrite expression (62) 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 $zy'(z)$:

$$zy'(z) < \frac{\theta}{\varepsilon_T \mu}\left(2+\gamma+2K_\phi^{-1}\right) = K_y. \tag{67}$$

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 $T''(y(z))$:

$$T''(y(z)) > \left(1-T'(y(z))\right)\frac{(1+\gamma)\, y(z) - K_y\gamma}{K_y y(z)} > \bar{\varepsilon}_T \frac{(1+\gamma)\, y(z) - K_y\gamma}{K_y y(z)}. \tag{68}$$

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 (68) 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 (67). In the absence of misspecification concerns, $\theta = \infty$, so that $K_y = \infty$ in (67), and there does not exist a $y(z)$ for which the right-hand side in (68) 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 (27) 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*

$$-\frac{T''(y)\,y}{1-T'(y)}=-\left[2-\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right]^{-1}\left[\frac{\mu}{\theta}\left[T'(y)\right]^2y-\gamma+\gamma\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right]. \tag{69}$$

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,

$$-z\frac{f'(z)}{f(z)}=1+\alpha. \tag{70}$$

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 (69), 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 (69) 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 (69). The left-hand side of this equation is the elasticity of take-home rate $1 - T'(y)$ with respect to income, or

$$\frac{d\log(1-T'(y))}{d\log y} = -\left[2 - \frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right]^{-1}\left[\frac{\mu}{\theta}\left[T'(y)\right]^2 y - \gamma + \gamma\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right]. \qquad (71)$$

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

$$\lim_{y\to\infty}\frac{\mu}{\theta}\left[T'(y)\right]^2 y = \gamma. \qquad (72)$$

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

$$\left| \frac{\mu}{\theta} \left[T'(y) \right]^2 y - \gamma \right| > \varepsilon > 0. \tag{73}$$

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 (71) 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 (73) holds as $\frac{\mu}{\theta} [T'(y)]^2 y - \gamma > \varepsilon$. Then the elasticity in (71) can be bounded as

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

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 (73) holds as $\frac{\mu}{\theta} [T'(y)]^2 y - \gamma < -\varepsilon$. Then the elasticity in (71) can be bounded as

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

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

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

As a consequence

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

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

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

for $y \geq y^*$ yields

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

and hence

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

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 (73), which is equivalent to a violation of equation (72), 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 (73) 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 (72) 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 (24). ■

D Generalizations of the baseline model

In this appendix, we provide details on the derivation of results from Section 3.4. We rely on the optimal tax formula (61). We start with the derivation of results under concave separable preferences from Section 3.4.1, and then analyze the case with welfare concerns at the top from Section 3.4.2. Finally, we derive results for power divergence functions from Section 3.4.3.

D.1 Preliminary results

The following result generalizes Lemma C.1:

Lemma D.1. *Under the concave separable preferences (28), the optimal level of output chosen by individual workers under tax scheme $T(y)$ satisfies:*

$$y'(z) = \frac{(1+\gamma)\frac{y(z)}{z}}{\gamma + \frac{T''(y(z))}{1-T'(y(z))}y(z) + \rho\omega(y(z))}. \tag{74}$$

where $\omega(y)$ is the ratio of marginal to average take-home rate

$$\omega(y) = \frac{1-T'(y)}{1-T(y)/y}. \tag{75}$$

Proof. The first-order condition for the optimal choice of labor (1) for the concave separable preferences (28) yields

$$vy(z)^{\gamma} = z^{1+\gamma}\left(1-T'(y(z))\right)\left(y(z)-T(y(z))\right)^{-\rho}. \tag{76}$$

Writing

$$H(y,z) = z^{1+\gamma}\left(1-T'(y)\right)\left(y-T(y)\right)^{-\rho} - vy^{\gamma} = 0,$$

and using the implicit function theorem, we obtain

$$y'(z) = -\frac{\partial H/\partial z}{\partial H/\partial y} = -\frac{(1+\gamma)\, z^{\gamma}\,(1-T'(y))\,(y-T(y))^{-\rho}}{z^{1+\gamma}\left[-T''(y)\,(y-T(y))^{-\rho} - \rho\,(1-T'(y))^{2}\,(y-T(y))^{-\rho-1}\right] - v\gamma y^{\gamma-1}}.$$

This expression, after simplifications using the first-order condition (76) yields the claim of the lemma. ∎

As we show, the marginal tax $T'(y(z))$ decays to zero as $z \to \infty$, in which case the marginal take-home rate $1-T'(y(z))$ and average take-home rate $1-T(y(z))/y(z)$ converge to one and $T''(y(z))$ to zero, and optimal level of output asymptotically behaves as

$$\begin{aligned} y(z) &\approx v^{-\frac{1}{\gamma+\rho}} z^{\frac{1+\gamma}{\rho+\gamma}} \\ y'(z) &\approx \frac{1+\gamma}{\rho+\gamma} v^{-\frac{1}{\gamma+\rho}} z^{\frac{1-\rho}{\rho+\gamma}} \approx \frac{1+\gamma}{\rho+\gamma}\frac{y(z)}{z}. \end{aligned}$$

In the case of quasilinear preferences when $\rho = 0$, expression (74) reduces to the result in Lemma C.1.

In a similar way to deriving $y'(z)$, we also need a characterization of indirect utility:

Lemma D.2. *Under the concave separable preferences (28), the indirect utility function satisfies*

$$\mathcal{U}'(z)\, z = v\left(\frac{y(z)}{z}\right)^{1+\gamma} = \frac{1-T'(y(z))}{(y(z)-T(y(z)))^{\rho}}\, y(z).$$

Proof. Start with the definition of indirect utility

$$\mathcal{U}(z) = \frac{c(z)^{1-\rho}}{1-\rho} - v\frac{n(z)^{1+\gamma}}{1+\gamma} = \frac{(y(z)-T(y(z)))^{1-\rho}}{1-\rho} - \frac{v}{1+\gamma}\left(\frac{y(z)}{z}\right)^{1+\gamma}.$$

Then direct differentiation yields

$$\mathcal{U}'(z) = (y(z)-T(y(z)))^{-\rho}\left(1-T'(y(z))\right) y'(z) - v\left(\frac{y(z)}{z}\right)^{\gamma}\frac{y'(z)\,z - y(z)}{z^{2}}.$$

This can be reorganized as

$$\mathcal{U}'(z)\, z = \frac{1-T'(y(z))}{(y(z)-T(y(z)))^{\rho}}\, y'(z)\, z - v\left(\frac{y(z)}{z}\right)^{\gamma} y'(z) + v\left(\frac{y(z)}{z}\right)^{1+\gamma}$$

Now use the optimality condition (76), which can be written as

$$v\left(\frac{y(z)}{z}\right)^{\gamma} = z\frac{1-T'(y(z))}{(y(z)-T(y(z)))^{\rho}}$$

to write $\mathcal{U}'(z)\, z$ as

$$\mathcal{U}'(z)\, z = v\left(\frac{y(z)}{z}\right)^{1+\gamma} = \frac{1-T'(y(z))}{(y(z)-T(y(z)))^{\rho}}\, y(z).$$

∎

We now need to extend Lemma C.3 in order to compute the derivative of the optimal tax rate in (61).

Lemma D.3. *The right-hand side term in (61) satisfies*

$$\frac{d}{dz}\frac{\widetilde{\Psi}(z)-\widetilde{F}_{\rho}(z)}{z\widetilde{f}_{\rho}(z)}=\frac{1}{z}\frac{\bar{\psi}^{-1}\psi(z)}{\bar{u}^{-1}c(z)^{\rho}}-\frac{1}{z}-\frac{1}{1+\gamma}\frac{T'(y(z))}{1-T'(y(z))}\left[\frac{1}{z}+\frac{\widetilde{f}'(z)}{\widetilde{f}(z)}+\rho\frac{y'(z)}{y(z)}\omega(y(z))\right].$$

Proof. This result is obtained by direct differentiation, utilizing (61) to substitute out $\widetilde{\Psi}(z)-\widetilde{F}_{\rho}(z)$ in the expression for the derivative. ■

Lemma D.4. *If* $\lim_{y\to\infty}T'(y)<1$, *then* $\lim_{y\to\infty}\omega(y)=1$.

Proof. Denote $\lim_{y\to\infty}T'(y)=\tau$, and recall that $\omega(y)$ is given by (75). The lemma is equivalent to showing that

$$\lim_{y\to\infty}\frac{T(y)}{y}=\frac{1}{y}T(0)+\frac{1}{y}\int_0^y T'(\xi)\,d\xi=\tau.$$

We need to show that $\forall\varepsilon>0$, there $\exists\bar{y}$ such that for all $y\geq\bar{y}$,

$$\left|\frac{T(y)}{y}-\tau\right|\leq\delta.$$

Define y^* such that for $y\geq y^*$ we have $|T'(y)-\tau|\leq\delta/3$. Then

$$\left|\frac{T(y)}{y}-\tau\right|=\left|\frac{1}{y}T(0)+\frac{1}{y}\int_0^y\left[T'(\xi)-\tau\right]d\xi\right|\leq\frac{1}{y}|T(0)|+\frac{1}{y}\left|\int_0^{y^*}\left[T'(\xi)-\tau\right]d\xi\right|+\frac{y-y^*}{y}\frac{\delta}{3}.$$

Now find a $\bar{y}$ such that for $y\geq\bar{y}$, the first and second term in the last expression are each bounded by $\delta/3$. This establishes the result. ■

We can also extend the result in Lemma C.4. The expression for $\widetilde{f}'(z)$ that appears in Lemma D.3 can be written as

$$\begin{aligned}\frac{\widetilde{f}'(z)}{\widetilde{f}(z)}&=\frac{d}{dz}\log\widetilde{f}(z)=\frac{d}{dz}\log m(z)+\frac{d}{dz}\log f(z)\\&=-\frac{1}{\theta}\left[\psi(z)\mathcal{U}'(z)+\psi'(z)\mathcal{U}(z)\right]-\frac{\mu}{\theta}T'(y(z))\,y'(z)+\frac{d}{dz}\log f(z).\end{aligned}$$

Finally, we utilize the analog of the notation for the derivative of the Pareto density (70), and define the function $\alpha(z)$ as

$$-z\frac{d}{dz}\log f(z)=-z\frac{f'(z)}{f(z)}\doteq 1+\alpha(z). \tag{77}$$

With these preliminary results, we can derive the generalization of Proposition C.5 that characterizes the differential equation for the marginal tax rate. Differentiating the optimal tax formula (61) with respect to z yields, using the result in Lemma D.3,

$$\frac{T''(y(z))\,y'(z)}{(1-T'(y(z)))^2} = \bar{u}(1+\gamma)\frac{\psi(z)}{\bar{\psi}z(y(z)-T(y(z)))^{\rho}} - \frac{1+\gamma}{z}$$
$$-\frac{T'(y(z))}{1-T'(y(z))}\left[\frac{1}{z}-\frac{1}{\theta}[\psi(z)\mathcal{U}(z)]' - \frac{\mu}{\theta}T'(y(z))\,y'(z) + \frac{d}{dz}\log f(z) + \rho\frac{y'(z)}{y(z)}\omega(y(z))\right].$$

Using Lemmas D.1, D.2, and D.3, collecting terms containing $T''(y(z))$, and utilizing expressions (75) and (77), yields, after some algebra,

$$\frac{T''(y(z))\,y(z)}{1-T'(y(z))} = \left[2-\frac{1+\gamma+\alpha(z)}{1+\gamma}T'(y(z)) - \Upsilon(z)\right]^{-1}\left[\frac{\mu}{\theta}[T'(y(z))]^2 y(z) - \gamma \right. \tag{78}$$
$$\left. +\gamma\frac{1+\gamma+\alpha(z)}{1+\gamma}T'(y(z)) + \rho\omega(y(z))\left(\frac{\alpha(z)}{1+\gamma}T'(y(z))-1\right) + (\gamma+\rho\omega(y(z)))\Upsilon(z)\right]$$

where, for $\rho \neq 1$,

$$\Upsilon(z) = \psi(z)(y(z)-T(y(z)))^{1-\rho}\cdot$$
$$\cdot\left[\frac{\bar{u}}{\bar{\psi}}\frac{\omega(y(z))}{y(z)} + \frac{T'(y(z))}{1+\gamma}\frac{1}{\theta}\left[z\frac{\psi'(z)}{\psi(z)}\left(\frac{1}{1-\rho}-\frac{\omega(y(z))}{1+\gamma}\right)+\omega(y(z))\right]\right].$$

The differential equation (78) characterizes the optimal tax rate for the case of concave separable preferences (28), general utilitarian Pareto weight function $\psi(z)$, a general benchmark distribution represented by the function $\alpha(z)$, and degree of robustness captured by θ. The function $\Upsilon(z)$ summarizes the contribution of the utilitarian concern.

When $\rho = 0$, $\psi(z) = 0$, and $\alpha(z) = \alpha$, the differential equation reduces to the quasilinear case under a Pareto distribution without a utilitarian concern in the right tail given by (27). We now consider the more general cases from Section 3.4.

D.2 Concave separable preferences

We first analyze the behavior of top marginal taxes in the case of concave separable preferences in absence of utilitarian concerns at the top with a benchmark Pareto distribution in the right tail as in Section 3.4.1. In this case, $\psi(z) = 0$, $\widetilde{\Psi}(z) = 1$, and $\alpha(z) = \alpha$ for sufficiently large z, and the differential equation (78) in the right tail can be written as explicitly depending on y instead of z:

$$\frac{T''(y)\,y}{1-T'(y)} = \left[2-\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right]^{-1}\cdot \tag{79}$$
$$\cdot\left[\frac{\mu}{\theta}[T'(y)]^2 y - \gamma + \gamma\frac{1+\gamma+\alpha}{1+\gamma}T'(y) + \rho\omega(y)\left(\frac{\alpha}{1+\gamma}T'(y)-1\right)\right].$$

The differential equation differs from the quasilinear case only by the presence of the last term on the second line.

Proof of Theorem 3.3. We first prove that $\lim_{y\to\infty} T'(y) \doteq \tau = 0$ must equal to zero. Assume, on the contrary, that $\tau > 0$. Then

$$m_\rho(z) = \bar{u}^{-1} c(z)^\rho m(z) = \bar{u}^{-1}(y(z) - T(y(z)))^\rho \bar{m} \exp\left(-\frac{\mu}{\theta} T(y(z))\right)$$

from equation (60) is decreasing in the right tail. To see this, notice that we have

$$m_\rho'(z) = \bar{u}^{-1} \bar{m} y'(z) \exp\left(-\frac{\mu}{\theta} T(y(z))\right) c(z)^{\rho-1} \left[\rho\left(1 - T'(y(z))\right) - \frac{\mu}{\theta} T'(y(z)) c(z)\right]. \quad (80)$$

The single-crossing property implies that optimal allocations $(c(z), y(z))$ must be increasing in z. As a consequence, the term in the bracket in the above expression is negative as $z \to \infty$. Specifically, if $\tau \in (0, 1)$, then the first-order condition (76) implies that

$$y(z) = z^{\frac{1+\gamma}{\gamma+\rho}} v^{-\frac{1}{\gamma+\rho}} \left(1 - T'(y(z))\right)^{\frac{1-\rho}{\gamma+\rho}} \omega(y(z))^{\frac{\rho}{\gamma+\rho}} \overset{z\to\infty}{\approx} z^{\frac{1+\gamma}{\gamma+\rho}} v^{-\frac{1}{\gamma+\rho}} (1-\tau)^{\frac{1-\rho}{\gamma+\rho}},$$

where the last comparison uses Lemma D.4. Hence $\lim_{z\to\infty} y(z) = \lim_{z\to\infty} c(z) = \infty$, and the bracket in (80) is dominated by the second term. On the other hand, if we had $\tau = 1$, then $c(z)$ is at least bounded from below by some positive constant, while the first term in the bracket converges to zero. In conclusion, if we had $\tau > 0$, then $m_\rho(z)$ is strictly decreasing in the right tail, and we can invoke Lemma C.2, concluding that the asymptotic tax rate τ is bounded above by the corresponding asymptotic tax rate in the rational case, and hence $\tau < 1$.

In this case, we can compute

$$\begin{aligned}\lim_{z\to\infty} \frac{T'(y(z))}{1 - T'(y(z))} = \frac{\tau}{1-\tau} &= \lim_{z\to\infty} (1+\gamma) \frac{1 - \widetilde{F}_\rho(z)}{z \widetilde{f}_\rho(z)} \\ &= \lim_{z\to\infty} \frac{1+\gamma}{\frac{\mu}{\theta} T'(y(z)) y'(z) z - \rho\omega(y(z)) z \frac{d}{dz} \log y(z) - z \frac{d}{dz} \log f(z) - 1}.\end{aligned}$$

Under the initial assumption that $\tau > 0$, and using the result that $\tau < 1$, the first term in the denominator diverges to infinity, which determines the behavior of the whole denominator. Then the formula yields $\tau = 0$, which is a contradiction. As a result, we must have $\tau = 0$.

Now assume in addition that the tail of the benchmark distribution is Pareto distributed, $\alpha(z) = \alpha$. Then the marginal tax rate obeys the differential equation (79). We can now follow the same arguments as in the proof of Theorem 3.2. Given that $\lim_{y\to\infty} T'(y) = 0$, the left-hand side of (79) converges to zero. The right-hand side must therefore converge to zero as well, which, given Lemma D.4, is only consistent with

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

We therefore obtain (31), and the elasticity (32) immediately follows. ■

D.3 Welfare concerns at the top

We now focus on the case from Section 3.4.2 where the planner imposes a uniform welfare weight function $\psi(z) \equiv 1$ on all households. In this case, $\bar{\psi} = 1$, and the welfare contribution $\Upsilon(z)$ in the differential equation (78) simplifies to

$$\Upsilon(z) = (y(z) - T(y(z)))^{1-\rho}\left[\bar{u}\frac{\omega(y(z))}{y(z)} + \frac{T'(y(z))}{1+\gamma}\frac{1}{\theta}\omega(y(z))\right]. \tag{81}$$

Proof of Theorem 3.4. Proving that $\lim_{y\to\infty} T'(y) = 0$ follows the same steps as in the proof of Theorem 3.3. The worst-case distribution is now given by (33), and the limiting tax can be computed by applying L'Hôpital's rule as

$$\lim_{z\to\infty}\frac{T'(y(z))}{1-T'(y(z))} = \frac{\tau}{1-\tau} = \lim_{z\to\infty}(1+\gamma)\frac{\widetilde{\Psi}(z) - \widetilde{F}_\rho(z)}{z\widetilde{f}_\rho(z)} \tag{82}$$
$$= \lim_{z\to\infty}\frac{(1+\gamma)\left(1-\bar{u}(y(z)-T(y(z)))^{-\rho}\right)}{\frac{1}{\theta}\mathcal{U}'(z)z + \frac{\mu}{\theta}T'(y(z))y'(z)z - \rho\omega(y(z))z\frac{d}{dz}\log y(z) - z\frac{d}{dz}\log f(z) - 1}.$$

As in the proof of Theorem 3.3, this expression is only consistent with $\lim_{z\to\infty} T'(y(z)) = 0$, as the numerator converges to $1+\gamma$, and the first two terms in the denominator diverge to infinity.

We now focus on the elasticity result. When $\psi(z) \equiv 1$, the term $\Upsilon(z)$ in differential equation (78) is given by (81). Under the Pareto distributed productivity, $\alpha(z) = \alpha$, and the differential equation does not explicitly depend on z, so we can view it as a function of y only:

$$\frac{T''(y)y}{1-T'(y)} = \left[2 - \frac{1+\gamma+\alpha}{1+\gamma}T'(y) - \Upsilon(y)\right]^{-1}\left[\frac{\mu}{\theta}\left[T'(y)\right]^2 y - \gamma \right. \tag{83}$$
$$\left. + \gamma\frac{1+\gamma+\alpha}{1+\gamma}T'(y) + \rho\omega(y)\left(\frac{\alpha}{1+\gamma}T'(y) - 1\right) + (\gamma + \rho\omega(y))\Upsilon(y)\right]$$

where

$$\Upsilon(y) = (y - T(y))^{1-\rho}\left[\bar{u}\frac{\omega(y)}{y} + \frac{T'(y)}{1+\gamma}\frac{1}{\theta}\omega(y)\right].$$

Again, given that $\lim_{y\to\infty} T'(y) = 0$, the left-hand side of (83) converges to zero. The terms in the second bracket on the right-hand side therefore have to balance so that the bracket converges to zero as well. Since $\lim_{y\to\infty}\omega(y) = 1$ by Lemma D.4, then after dropping dominated terms, the bracket asymptotically behaves as

$$\frac{\mu}{\theta}\left[T'(y)\right]^2 y - (\gamma+\rho) + (\gamma+\rho)\Upsilon(y) \approx \frac{\gamma+\rho}{1+\gamma}\frac{1}{\theta}T'(y)y^{1-\rho} + \frac{\mu}{\theta}\left[T'(y)\right]^2 y - (\gamma+\rho)$$

The first and second terms must have nonpositive asymptotic growth rates, and one of these growth rates has to be zero to offset the last constant term. Denoting

$$\varepsilon_\tau = \lim_{y\to\infty}\frac{d\log T'(y)}{d\log y},$$

we have

$$\lim_{y\to\infty}\frac{d}{d\log y}\log\left(\frac{\gamma+\rho}{1+\gamma}\frac{1}{\theta}T'(y)\,y^{1-\rho}\right)=\varepsilon_\tau+1-\rho\leq 0$$
$$\lim_{y\to\infty}\frac{d}{d\log y}\log\left(\frac{\mu}{\theta}\left[T'(y)\right]^2 y\right)=2\varepsilon_\tau+1\leq 0$$

and with one of these inequalities binding, we obtain

$$\varepsilon_\tau=\min\left(-\frac{1}{2},\rho-1\right).$$

■

D.4 Power divergence functions

Here we study the characterization of optimal tax rates under the class of penalty functions from Section 3.4.3. Cressie and Read (1984) study a class of power divergence functions

$$\mathcal{E}_\eta(m)=\mathbb{E}\left[\phi_\eta(m)\right]=\mathbb{E}\left[\frac{m^{1+\eta}-1}{\eta(1+\eta)}\right].$$

The function $\phi_\eta(m)$ satisfies $\phi''_\eta(m)=m^{\eta-1}$, so it is convex for $m>0$. The functions for $\eta\in\{-1,0\}$ can be constructed by taking appropriate limits using L'Hôpital's rule. For $\eta=0$, we obtain the entropy

$$\lim_{\eta\to 0}\mathcal{E}_\eta(m)=\lim_{\eta\to 0}\frac{\mathbb{E}\left[e^{(1+\eta)\log m}\right]-1}{\eta(1+\eta)}=\lim_{\eta\to 0}\frac{\mathbb{E}\left[e^{(1+\eta)\log m}\log m\right]}{1+2\eta}=\mathbb{E}\left[m\log m\right]\doteq\mathcal{E}_0(m).$$

For $\eta=-1$, the same calculation yields the negative of the empirical likelihood

$$\lim_{\eta\to -1}\mathcal{E}_\eta(m)=\mathbb{E}\left[-\log m\right]\doteq\mathcal{E}_{-1}(m).$$

The divergence $\mathcal{E}_\eta(m)$ is positive for all positive m, and uniquely minimized at $m\equiv 1$. To see this, observe that the minimization problem

$$\min_m \mathbb{E}\left[\phi_\eta(m)\right]\quad \text{s.t. } \mathbb{E}\left[m\right]=1$$

leads to first-order conditions

$$\phi'_\eta(m)=\frac{m^\eta}{\eta}=\chi$$

where χ is the Lagrange multiplier on the constraint. Hence m must be constant, and therefore $m\equiv 1$.

D.4.1 Optimal taxes under power divergence

The characterization of the planner's problem follows Appendix B. The planner solves

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

subject to the incentive constraint (51) and the budget constraint (52). Specializing to the case of quasilinear utility (13), and restricting attention to the part of the space $z \geq \hat{z}$ for which $\psi(z) = 0$, the associated Hamiltonian becomes

$$\begin{aligned} H(\mathcal{U}, y, m, \lambda) \quad &= \quad \theta \frac{m(z)^{1+\eta} - 1}{\eta(1+\eta)} f(z) - \chi m(z) f(z) + \frac{\lambda(z)}{z} \left(\frac{y(z)}{z}\right)^{1+\gamma} \\ &\quad + \mu \left[y(z) - C(\mathcal{U}(z), y(z))\right] m(z) f(z). \end{aligned}$$

The optimality condition with respect to the probability distortion $m(z)$ yields

$$0 = H_m = \frac{\theta}{\eta} m(z)^{\eta} f(z) - \chi f(z) + \mu \left[y(z) - C(\mathcal{U}(z), y(z))\right] f(z).$$

This yields the worst-case distortion

$$m(z) = \left[\frac{\eta}{\theta}\left(\chi - \mu T(y(z))\right)\right]^{\frac{1}{\eta}}. \tag{84}$$

The distortion is a decreasing function of the tax $T(y(z))$. The remaining optimality conditions remain as in Appendix B, yielding the usual optimal tax formula

$$\frac{T'(y(z))}{1 - T'(y(z))} = (1+\gamma) \frac{1 - \widetilde{F}(z)}{z \widetilde{f}(z)}. \tag{85}$$

We can now follow the same steps as in the proof of Proposition C.5. The modification is that

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

where the last term uses the Pareto density specification for the productivity distribution. The derivation of the differential equation now follows Proposition C.5, with the newly specified distortion $m(z)$ given by (84).

Proposition D.5. *On the interval where the marginal tax $T'(y)$ is strictly positive, it obeys the differential equation*

$$-\frac{T''(y)\, y}{1 - T'(y)} = -\left[2 - \frac{1+\gamma+\alpha}{1+\gamma} T'(y)\right]^{-1} \left[\frac{\mu}{\eta} \frac{\left[T'(y)\right]^2 y}{\chi - \mu T(y)} - \gamma + \gamma \frac{1+\gamma+\alpha}{1+\gamma} T'(y)\right]. \tag{86}$$

Proof. Using Lemma C.3, we differentiate the optimal marginal tax formula (85) to obtain

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

We collect terms containing $y'(z)$ and use Lemma C.1 to substitute out $y'(z)$:

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

This expression does not explicitly depend on z, so it can be dropped from the arguments. Reorganizing to isolate the expression for $T''(y)$ then yields

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

obtaining the claim. ■

We now provide an informal characterization of the tail behavior of the marginal tax rate.

First consider the case $\eta < 0$. This requires $\chi - \mu T(y(z))$ in (84) to be negative, so $\mu T(y)$ is bounded from below. Denote $\lim_{z\to\infty} T'(y(z)) = \tau$, and assume $\tau > 0$. Recall that the optimal choice of output asymptotically behaves as

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

and hence

$$T(y(z)) \approx \tau y(z) \approx \tau(1-\tau)^{\frac{1}{\gamma}} z^{\frac{1+\gamma}{\gamma}}.$$

This implies that the worst-case distortion can be approximated for large z as

$$m(z) = \left[\frac{\eta}{\theta}(\chi-\mu T(y(z)))\right]^{\frac{1}{\eta}} \approx z^{\frac{1+\gamma}{\gamma}\frac{1}{\eta}}.$$

If the tail of the benchmark distribution $f(z)$ is Pareto, then

$$\widetilde{f}(z) = m(z)\,f(z) = m(z)\frac{\alpha}{z^{\alpha+1}} \approx z^{\frac{1+\gamma}{\gamma}\frac{1}{\eta}-(\alpha+1)} = z^{-\left(-\frac{1+\gamma}{\gamma}\frac{1}{\eta}+\alpha+1\right)} = \frac{1}{z^{\widetilde{\alpha}+1}}$$

which is approximately Pareto distributed again, with tail coefficient

$$\widetilde{\alpha} = \alpha - \frac{1+\gamma}{\gamma}\frac{1}{\eta} > \alpha$$

As a consequence, the limiting marginal tax rate is

$$\tau = \frac{1+\gamma}{1+\gamma+\widetilde{\alpha}} < \frac{1+\gamma}{1+\gamma+\alpha} = \tau^{rat}.$$

Misspecification concerns therefore still lower marginal taxes at the top, even though the form of the penalty that puts less weight on the right tail of the distribution weakens the effect. As $\eta \to 0$, we approach the entropy penalty case, which implies $\widetilde{\alpha} \to \infty$, and $\tau \to 0$. On the other hand, choosing $\eta \to -\infty$ makes the misspecification concerns in the right tail vanish, and $\widetilde{\alpha} \to \alpha$, approaching the rational case $\tau \to \tau_{rat}$.

Now consider the case $\eta > 0$. This requires $\chi - \mu T(y(z))$ in (84) to be positive, so $\mu T(y)$ is bounded from above. We show that there exists a $\bar{z} < \infty$ such that $m(z) = 0$ for all $z \geq \bar{z}$, and hence the marginal tax rate reaches zero for a finite $\bar{y} = y(\bar{z})$.

Consider to the contrary that the given $\bar{z} = \infty$. We can exclude $\tau = 1$ based on Lemma C.2 and the fact that $m(z)$ is strictly decreasing. Hence $\bar{y} = y(\bar{z}) = \infty$. We now show that this is not possible. First, if $\tau \in (0, 1)$, then $\lim_{z\to\infty} y(z) = \lim_{z\to\infty} T(y(z)) = \infty$, and $T(y(z))$ cannot be bounded from above. Hence $\tau = 0$. Second, assuming that $T'(y)$ converges to zero monotonically in the tail, the left-hand side of (86) must converge to zero, and we must have

$$\lim_{y\to\infty} \frac{\mu}{\eta} \frac{[T'(y)]^2 y}{\chi - \mu T(y)} = \gamma. \tag{87}$$

First assume that $\lim_{y\to\infty} \chi - \mu T(y) > 0$. In this case, $T'(y)$ in the numerator has to asymptotically behave as $y^{-1/2}$ so that the left-hand side of (87) converges to a finite strictly positive value. But in that case, $\lim_{y\to\infty} T(y) = \infty$. As an alternative, consider $\lim_{y\to\infty} \chi - \mu T(y) = 0$. Then an application of L'Hôpital's rule implies

$$\lim_{y\to\infty} \frac{\mu}{\eta} \frac{[T'(y)]^2 y}{\chi - \mu T(y)} = \lim_{y\to\bar{y}} -\frac{1}{\eta} \left[2T''(y)\, y + T'(y)\right] = 0.$$

In conclusion, $\bar{z} = \infty$ is not consistent with any possible behavior of optimal marginal taxes in the right tail. As a consequence, optimal marginal taxes reach zero for a finite $\bar{z}$ and $\bar{y} = y(\bar{z})$. The form of the power divergence puts a sufficiently strong emphasis on the misspecification concern in the right tail that the planner fears that there are no workers with productivities above a certain productivity threshold $\bar{z}$, leading to optimal marginal taxes converging to zero at this $\bar{z}$.

D.5 Tail behavior of marginal taxes in the quantitative model

Proof of Lemma 4.1. The exponentially modified Gaussian density (38) implies that that productivity z has density

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

Utilizing formula (82), we first notice that

$$\lim_{z\to\infty} z \frac{d}{dz} \log f(z) = -(1+\alpha) + \lim_{z\to\infty} \frac{\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\bar{\mu}+\alpha\sigma^2-\log z}{\sigma}\right)^2\right)}{1 - \Phi\left(\frac{\bar{\mu}+\alpha\sigma^2-\log z}{\sigma}\right)} = -(1+\alpha) \tag{88}$$

as in the case of Pareto distribution. Further, denoting $\tau = \lim_{z\to\infty} T'(y(z))$, optimality condition (76) and the fact that $\lim_{z\to\infty} \omega(y(z)) = 1$ imply that asymptotically as $z \to \infty$,

$$y(z) \approx v^{-\frac{1}{\gamma+\rho}} (1-\tau)^{\frac{1-\rho}{\gamma+\rho}} z^{\frac{1+\gamma}{\gamma+\rho}}$$

so that

$$\lim_{z\to\infty} z\frac{d}{dz} \log y(z) = \frac{1+\gamma}{\gamma+\rho}.$$

When $\theta = \infty$, formula (82) then yields

$$\lim_{z\to\infty} \frac{T'(y(z))}{1-T'(y(z))} = \lim_{z\to\infty} \frac{(1+\gamma)\left(1-\bar{u}\,(y(z)-T(y(z)))^{-\rho}\right)}{-\rho\omega(y(z))\,z\frac{d}{dz}\log y(z) - z\frac{d}{dz}\log f(z) - 1} = \frac{1+\gamma}{-\rho\frac{1+\gamma}{\gamma+\rho}+\alpha}.$$

Reorganizing this expression yields (39).

The second result of the lemma follows directly from Theorem 3.4, noticing from equation (88) that the EMG distribution asymptotically behaves as the Pareto distribution.

Finally, for the limit as $y \to 0$ ($z \to 0$), the Inada condition implies that $\mathcal{C}(0) > 0$, and $\mathcal{U}'(0) < \infty$. Consequently, $\mathcal{U}(0)$ and $T(y(z))$ are finite, and the distortion $m(0)$ from expression (33) is finite as well. We then obtain

$$\lim_{z\to 0} z\widetilde{f}_\rho(z) = \lim_{z\to 0} z\,(\mathcal{C}(z))^\rho\, m(z)\, f(z) = 0.$$

Applying L'Hôpital's rule as in (82) but for $z \to 0$ yields

$$\begin{aligned}\lim_{z\to 0} \frac{T'(y(z))}{1-T'(y(z))} &= \lim_{z\to 0} (1+\gamma)\,\frac{\widetilde{\Psi}(z) - \widetilde{F}_\rho(z)}{z\widetilde{f}_\rho(z)} \qquad (89)\\ &= \lim_{z\to 0} \frac{(1+\gamma)\left(1-\bar{u}\,(y(z)-T(y(z)))^{-\rho}\right)}{\frac{1}{\theta}\mathcal{U}'(z)\,z + \frac{\mu}{\theta}T'(y(z))\,y'(z)\,z - \rho\omega(y(z))\,z\frac{d}{dz}\log y(z) - z\frac{d}{dz}\log f(z) - 1}.\end{aligned}$$

Since $\widetilde{\Psi}(z) > \widetilde{F}_\rho(z)$ in the neighborhood of zero, the limiting marginal tax in (89) must be nonnegative and smaller than one. The numerator on the second line converges to a strictly negative finite number. The first two terms in the denominator converge to zero (or are equal to zero in the case $\theta = \infty$). The third term is negative because both $\omega(y(z))$ and $z\frac{d}{dz}\log y(z)$ are positive. Finally, an application of L'Hôpital's rule to the analogue of equation (88) reveals that

$$\lim_{z\to 0} -z\frac{d}{dz}\log f(z) = -\infty.$$

As a result, the limit in (89) is zero, establishing the last claim of the lemma. ■

E Estimation of the exponentially modified Gaussian distribution from data

This appendix outlines the procedure for estimating the parameters of the exponentially modified Gaussian (EMG) distribution using income data for the United States, as referenced in Section 4.

Data source

The income data is sourced from the World Inequality Database (WID)[12], which provides detailed information on country-level income and wealth distributions across the world. In the U.S., fiscal income is reported at the tax unit level (individuals or households). Observations are annual.

For each year $t \in \{1966, 1967, \ldots, 2015\}$, the dataset contains mean income $\bar{y}_t$ and income levels $y_{q,t}$ at quantiles $q \in \mathcal{Q} = \{0.5, 0.9, 0.95, 0.99, 0.995, 0.999, 0.9999\}$. To standardize the distributions across years and take out trend income growth, we compute mean-adjusted quantiles

$$\tilde{y}_{q,t} = \frac{y_{q,t}}{\bar{y}_t},$$

ensuring that the mean of the adjusted distribution is normalized to one for each year.

Parameter estimation

For each year t, we fit an EMG distribution f_t with parameters $\Theta_t = (\bar{\mu}_t, \sigma_t, \alpha_t)$, where $\bar{\mu}_t$ is the mean of the Gaussian component, σ_t is the standard deviation of the Gaussian component, and α_t is the shape parameter of the Pareto component. To ensure consistency with the mean-adjusted data, we normalize $\bar{\mu}_t$ to

$$\bar{\mu}_t = \ln\left(\frac{\alpha_t - 1}{\alpha_t}\right) - \frac{\sigma_t^2}{2}.$$

We estimate (σ_t, α_t) by minimizing the unweighted root mean squared error between the empirical and theoretical quantiles:

$$\mathcal{D}\left(\Theta; \{\tilde{y}_{q,t}\}_{q\in\mathcal{Q}}\right) = \left(\sum_{q\in\mathcal{Q}} \left(\ln \tilde{y}_{q,t} - Q\left(q;\Theta\right)\right)^2\right)^{1/2}$$

where $Q(q;\Theta)$ is the quantile function of the EMG distribution with parameters Θ. The Nelder–Mead algorithm is used for optimization.

F Proofs for the model with multidimensional uncertainty

In this section, we derive the expressions for the worst-case density for the multidimensional case in Section 5.1, and then derive explicit formulas for shifters Ψ and Δ in equation (46) that guarantee conditional densities of $\gamma|z$ are not distorted at the status quo tax function T^0.

[12] The database is available at https://wid.world.

F.1 Worst-case densities for the planner's problem

It is useful to factorize the worst-case distoriton as

$$m(\gamma,z)=\frac{\widetilde{f}(\gamma,z)}{f(\gamma,z)}=\frac{\widetilde{f}_{\gamma|z}(\gamma|z)}{f_{\gamma|z}(\gamma|z)}\frac{\widetilde{f}_{z}(z)}{f_{z}(z)}=m_{\gamma|z}(\gamma|z)\,m_{z}(z).$$

Problem (44) can then be equivalently written as

$$\max_{T}\min_{m_z,m_{\gamma|z}>0}\int\int\psi(\gamma,z)\,\mathcal{U}(\gamma,z)\,m_{\gamma|z}(\gamma|z)\,m_z(z)\,f(\gamma,z)\,d\gamma dz+V(G)$$
$$+\theta_\gamma\int\int m_{\gamma|z}(\gamma|z)\log m_{\gamma|z}(\gamma|z)\,m_z(z)\,f(\gamma,z)\,d\gamma dz+\theta_z\int\int m_z(z)\log m_z(z)\,f(\gamma,z)\,d\gamma dz$$

subject to the budget constraint

$$G=\int\int T(y(\gamma,z))\,m_{\gamma|z}(\gamma|z)\,m_z(z)\,f(\gamma,z)\,d\gamma dz$$

and the normalization constraints $1=\mathbb{E}[m_z]$ and $1=\mathbb{E}\left[m_{\gamma|z}|z\right]$. Imposing a Lagrange multiplier μ on the budget constraint and multipliers χ_z and $\chi_{\gamma|z}(z)\,f_z(z)$ on the normalization constraints yields the first-order conditions with respect to the choice of $m_{\gamma|z}(\gamma|z)$ and $m_z(z)$

$$0=\psi(\gamma,z)\,\mathcal{U}(\gamma,z)+\theta_\gamma\left[1+\log m_{\gamma|z}(\gamma|z)\right]+\mu T((y(\gamma,z)))+\frac{\chi_{\gamma|z}(z)}{m_z(z)}$$

and

$$\begin{aligned}0 &= \int\psi(\gamma,z)\,\mathcal{U}(\gamma,z)\,m_{\gamma|z}(\gamma|z)\,f_{\gamma|z}(\gamma|z)\,d\gamma+\theta_\gamma\int m_{\gamma|z}(\gamma|z)\log m_{\gamma|z}(\gamma|z)\,f_{\gamma|z}(\gamma|z)\,d\gamma+\\ &\quad+\theta_z\left[1+\log m_z(z)\right]+\mu\int T((y(\gamma,z)))\,m_{\gamma|z}(\gamma|z)\,f_{\gamma|z}(\gamma|z)\,d\gamma+\chi_z.\end{aligned}$$

Using the normalization constraints to solve for the Lagrange multipliers, we obtain the distortions

$$m_{\gamma|z}(\gamma|z)=\bar{m}_{\gamma|z}(z)\exp\left(-\frac{1}{\theta_\gamma}\left[\psi(\gamma,z)\,\mathcal{U}(\gamma,z)+\mu T((y(\gamma,z)))\right]\right)$$

and

$$m_z(z)=\bar{m}_z\exp\left(-\frac{1}{\theta_z}\left[\widetilde{\mathbb{E}}\left[\psi\mathcal{U}+\mu T(y)\,|z\right]+\theta_\gamma\mathcal{E}\left(f_{\gamma|z}(\cdot|z),\widetilde{f}_{\gamma|z}(\cdot|z)\right)\right]\right),$$

where $\bar{m}_{\gamma|z}(z)$ and $\bar{m}_z$ are normalization constants. The conditional distortion $m_{\gamma|z}(\gamma|z)$ reflects differences in utility and budgetary contributions across labor supply elasticities. The marginal distortion $m_z(z)$, on the other hand, reflects average differences in these contributions across different levels of productivity.

F.2 Formulas for preference shifters Ψ and Δ

We summarize the computation of the shifters Ψ and Δ in the following lemma.

Lemma F.1. *Suppose the utilitarian planner is endowed with a concave GHH utility function (46), the status quo tax policy T^0, and let $\bar{c}\left(z;T\right), \bar{n}\left(z;T\right)$ be the optimal choices of workers with utility*

$$\bar{u}\left(z;T\right) = \max_{c \leq zn - T(zn), n} \frac{1}{1-\rho}\left(c - \bar{\psi}\frac{n^{1+\bar{\gamma}}}{1+\bar{\gamma}}\right)^{1-\rho} \tag{90}$$

for some positive scalars $\bar{\psi}$ and $\bar{\gamma}$. Then there exist labor disutility shifters $\Psi\left(\gamma, z; T^0\right)$ and utility shifters $\Delta\left(\gamma, z; T^0\right)$ that satisfy the following properties:

1. *The optimal choices of workers given the tax policy T, $\mathcal{C}\left(\gamma, z; T\right)$, $\mathcal{N}\left(\gamma, z; T\right)$ and the optimal level of utility $\mathcal{U}\left(\gamma, z; T\right)$ do not depend on γ when $T = T^0$, and for all γ', z,*

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

2. *When $T = T^0$, the minimizing probability measure given productivity level z is not distorted, and for all γ, z we have $\tilde{f}_{\gamma|z}\left(\gamma|z\right) = f_{\gamma|z}\left(\gamma|z\right)$.*

Proof. To show the first part, we construct $\Psi\left(\gamma, z; T^0\right)$ and $\Delta\left(\gamma, z; T^0\right)$ so that the optimal choices and the indirect utility under (46) coincide with those for (90). Optimality with respect to n for a worker with preferences (46) yields the first order condition

$$\Psi\left(\gamma, z; T^0\right) n^{\gamma} = z\left(1 - T^{0\prime}\left(zn\right)\right). \tag{91}$$

For $\bar{n}\left(z; T^0\right)$ to satisfy the condition, $\Psi\left(\gamma, z; T^0\right)$ needs to satisfy

$$\Psi\left(\gamma, z; T^0\right) = \frac{z\left(1 - T^{0\prime}\left(z\bar{n}\left(z; T^0\right)\right)\right)}{\bar{n}\left(z; T^0\right)^{\gamma}}.$$

Since $\mathcal{N}\left(\gamma, z; T^0\right)$ is the optimal choice for a household with preferences (46), it satisfies the first-order condition (91), and we have $\mathcal{N}\left(\gamma, z; T^0\right) = \bar{n}\left(z; T^0\right)$ for any z. Then $\mathcal{C}\left(\gamma, z; T^0\right) = \bar{c}\left(z; T^0\right)$ follows from the budget constraint. Finally, construct $\Delta\left(\gamma, z; T^0\right)$ as a difference in the utility level to ensure that $\mathcal{U}\left(\gamma, z; T^0\right) = \bar{u}\left(z; T^0\right)$,

$$\Delta\left(\gamma, z; T^0\right) = \bar{u}\left(z; T^0\right) - \frac{1}{1-\rho}\left(\mathcal{C}\left(\gamma, z; T^0\right) - \Psi\left(\gamma, z; T^0\right)\frac{\mathcal{N}\left(\gamma, z; T^0\right)^{1+\gamma}}{1+\gamma}\right)^{1-\rho}.$$

The second part follows from the minimizing formula of the distortion in equation (45) and first part of the lemma. ■

G Approximation of nonlinear income tax function using a cubic spline

In this section, we describe the choice of the nonlinear tax function used in Section 5.2.

Restricted class of nonlinear income tax function

We restrict the tax function T to a flexible parametric class of functions such that the 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 is constant outside the range of the knots. Specifically,

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

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 $y \in (0, \infty)$. The tax function $T(y)$ is given by

$$T(y) = T_0 + \int_0^y T'(x)\, dx$$

where T_0 is a separately determined intercept of the tax function.

Choice of N

Deciding the number of knots N introduces 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 the cubic spline. Given the tradeoff, we choose the smallest N such that welfare is numerically close to the level of welfare in cases where the fully nonlinear Mirrleesian allocation can be feasibly computed.

Optimization strategy

The tax function is represented by a set of knots $\{(\ln y_i, \tau_i)\}_{i=1,\cdots,N}$ and the intercept term T_0. Solving the government problem for the optimal tax function T amounts to searching over a set of scalars $\left(T_0, \{\ln y_i\}_{i=1,\cdots,N}, \{\tau_i\}_{i=1,\cdots,N}\right)$.

The optimization scheme is as follows. We first fix the location of the outer knots, $(\ln y_1, \ln y_N)$, and then search for the location of remaining knots $\{\ln y_i\}_{i=2,\cdots,N-1}$ and the marginal tax rates at all knots $\{\tau_i\}_{i=1,\cdots,N}$ to maximize the objective using the Nelder–Mead algorithm. The intercept term T_0 is set to clear the government budget constraint. This procedure is repeated over multiple candidates for the location of two outer knots, $(\ln y_1, \ln y_N)$, and we pick one that achieves the highest value of the objective.

We choose not to fully optimize over the location of the outer knots $(\ln y_1, \ln y_N)$. Instead, we

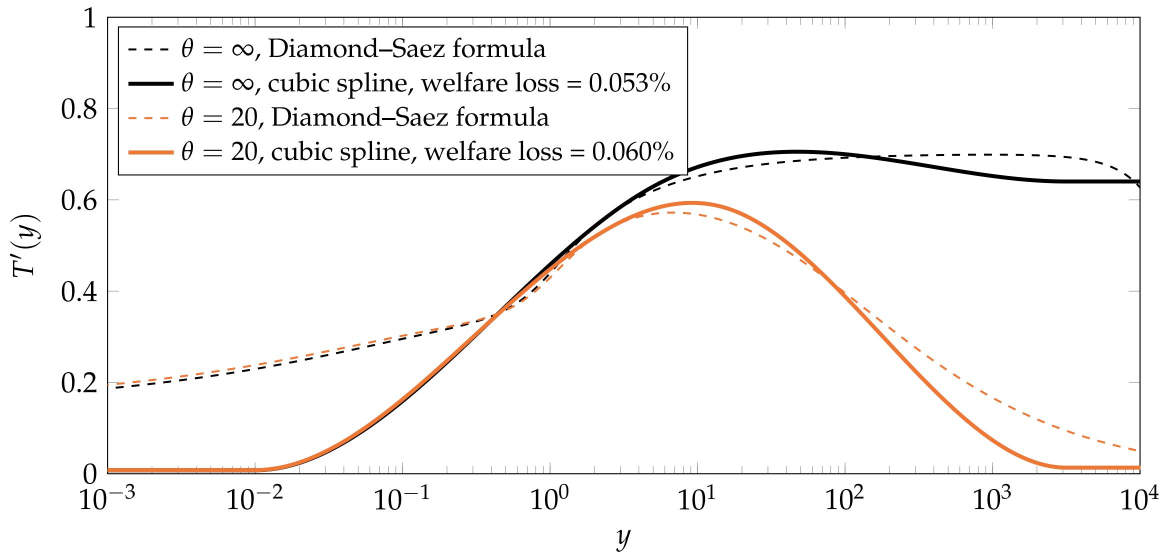

Figure 15: Comparison of the fully nonlinear Mirrleesian optimal marginal tax with the $N = 3$ cubic spline approximation in the economy with one-dimensional heterogeneity.

restrict their range to ensure that the marginal tax rates are variable in a wide range of the income distribution and the optimization scheme is numerically stable. While optimizing over the location achieves marginally higher welfare, it often leads to a narrow range of the income distribution where the marginal tax rates are variable, making the resulting tax function less interpretable. At the same time, putting the outer knots too far away from the range of the income distribution can lead to numerical instability in the optimization scheme. This is because the marginal tax rates in the distant tails of the income distribution have negligible implications for welfare.

Performance benchmarking in the model with single-dimensional uncertainty

Our solution with $N = 3$ cubic splines is benchmarked against the fully nonlinear Mirrleesian solution in the one-dimensional setup. For this computation, the productivity z-space is discretized with 500 equidistant points in log-space.

Figure 15 compares Mirrleesian solutions and our solutions with three knots $(N = 3)$ for both the rational case $(\theta = \infty)$ and a robust case $(\theta = 20)$. The dashed lines represent the marginal tax rates based on the Mirrlees solution (Diamond–Saez formula), and solid lines represent the tax rates based on our spline-based solution. Black and orange lines are the solutions for the rational case and robust case, respectively. The figure shows that our solutions approximate the Mirrlees solutions well in terms of the overall shape of the marginal tax schedule, except for very low income earners. It is also noteworthy that both approaches achieve very similar levels of welfare. In terms of consumption equivalence, the welfare loss from choosing our solution method is less than 0.1% for the rational and robust cases. We conclude that our solution with the number of knots $N = 3$ is sufficiently flexible to approximate the Mirrleesian solution with minimal welfare loss.

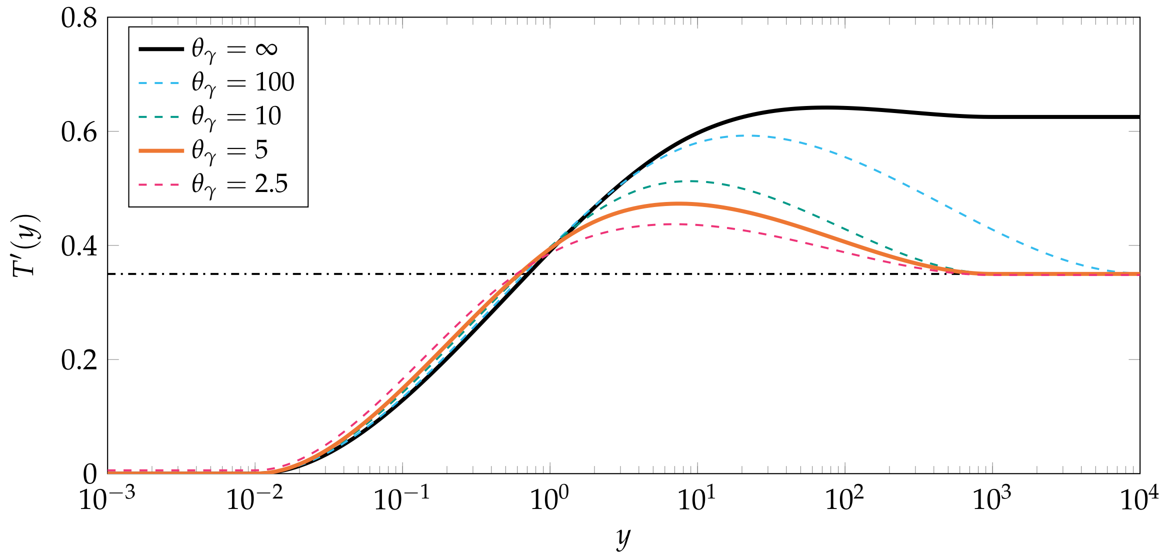

Figure 16: Optimal marginal tax schedules for alternative levels of multidimensional misspecification concerns. The status quo tax policy is an affine tax function with a constant marginal tax rate of 35%, shown by the horizontal dashdotted line.

H Optimal tax schedule with status-quo tax policy calibrated to the U.S.

In this section, we present the optimal tax schedule with multidimensional misspecification concerns and 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 to an affine function with a constant marginal tax rate of 35%. While this does not correspond to marginal tax rates of low income earners, it turns out that the choice of the status-quo for low-income earners is essentially inconsequential. The preference shifters Ψ and Δ are recalculated with the new status-quo tax policy.

The results are shown in Figure 16. As in the baseline with $T^0 = 0$, concerns about the labor supply responses lower tax rate for high incomes. Differently from the baseline, the top tax rate now asymptotes to 35%.

The reasoning is as follows. Typically, the optimal top tax rate decreases with greater concern about model misspecification, as discussed in the main text. In our formulation, concerns about misspecification in the elasticity parameter γ are linked to the size of the proposed reform (see Lemma F.1). The worst-case elasticities for top earners depend on whether the proposed tax rate is above or below the status quo. When the proposed rate exceeds the status-quo rate, the government fears that top earners may be highly elastic. For sufficiently low values of θ_γ, this concern leads to a preference for smaller tax increases. Conversely, when the proposed top tax rate is below the status quo, the worst-case scenario at low θ_γ is that top earners are extremely inelastic. In this case, it becomes optimal for the government to raise the tax rate substantially, since revenue can be raised with little efficiency cost, rendering the initial proposal suboptimal. This logic implies that optimal top tax rates will not fall below the status-quo level.

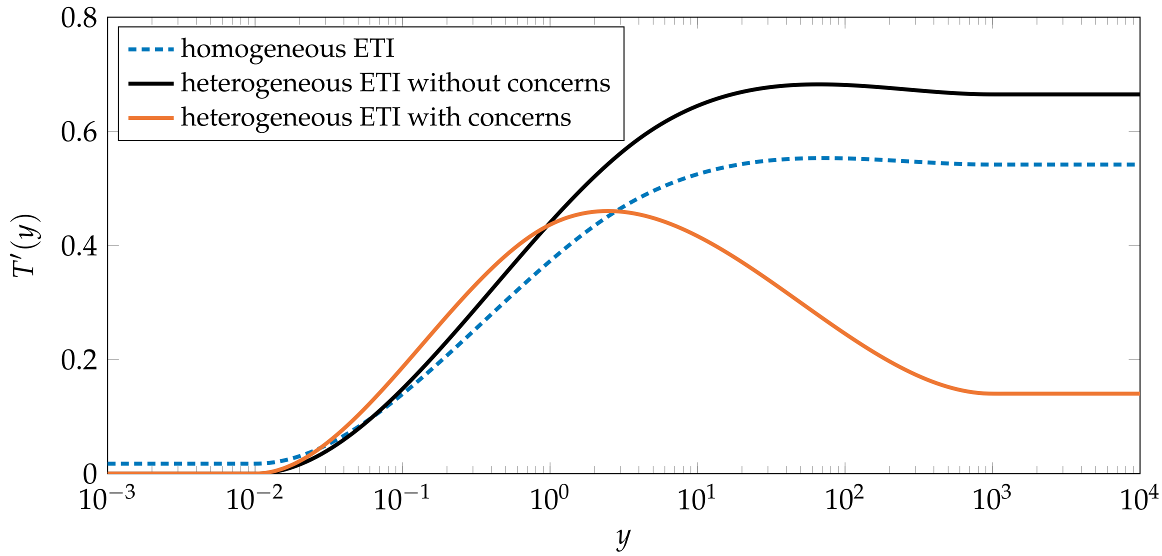

Figure 17: Comparison of optimal tax policy in models with and without heterogeneity in the elasticity of taxable income.

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

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

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 (ETI). 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 uncertainty about the ETI. In contrast, 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 ETI distribution, relative to the case without concerns about the ETI distribution.

Those two findings are consistent. Their approach maximizes the expected welfare of the tax planner over the distribution of ETI, which is equivalent to maximizing the welfare under a heterogeneous population with a known distribution of ETI. In our experiment, their analysis with uncertainty corresponds to the case without misspecification concerns ($\theta_z = \infty$ and $\theta_\gamma = \infty$). Relative to this starting point, we find that misspecification concerns about the ETI distribution lead to lower marginal tax rates at the top. On the other hand, relative to the case without heterogeneity in ETI, they find that introducing known heterogeneity in the ETI leads to a higher marginal tax rate at the top.

We numerically confirm these findings. Figure 17 shows the optimal tax policy in three models with and without heterogeneity in the elasticity of taxable income. The optimal marginal tax schedule is computed as a cubic spline with three knots. The black line (homogeneous ETI) represents the optimal marginal tax schedule in a model without heterogeneity in the elasticity of taxable income, where the elasticity is assumed to be 0.59, the mean value of the ETI in Figure 11. The blue dashed 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 Figure 11. The orange line (heterogeneous ETI with concerns) is the optimal marginal tax schedule with the penalty parameter $\theta_\gamma = 5$. The black and orange lines are taken from Figure 13 in the main text. The figure confirms that introducing heterogeneity in ETI leads to a higher marginal tax rate at the top (the difference between black and blue lines), while concerns about the ETI lead to a lower marginal tax rate at the top (the difference between the black and orange lines).

References

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

Bhandari, Anmol, Jaroslav Borovička, and Paul Ho. 2025. Survey Data and Subjective Beliefs in Business Cycle Models. *Review of Economic Studies* 92 (3):1375–1437.

Bidder, Rhys and Matthew E. Smith. 2012. Robust Animal Spirits. *Journal of Monetary Economics* 59 (8):738–750.

Boerma, Job, Aleh Tsyvinski, and Alexander P. Zimin. 2025. Bunching and Taxing Multidimensional Skills. Unpublished working paper, revise and resubmit at Econometrica.

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

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

Chang, Minsu and Chunzan Wu. 2025. When in Doubt, Tax More Progressively? Uncertainty and Progressive Income Taxation. Forthcoming in *International Economic Review*.

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.

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.

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.

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.

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

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

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

Greenwood, Jeremy, Zvi Hercowitz, and Gregory W. Huffman. 1988. Investment, Capacity Utilization, and the Real Business Cycle. *American Economic Review* 78 (3):402–417.

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

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

———. 2008. *Robustness*. Princeton University Press, Princeton, New Jersey.

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

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

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

Ilut, Cosmin and Martin Schneider. 2023. Ambiguity. In *Handbook of Economic Expectations*, edited by Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw, chap. 24, 749–777. Academic Press.

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

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

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.

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.

Maenhout, Pascal J. 2004. Robust Portfolio Rules and Asset Pricing. *The Review of Financial Studies* 17 (4):951–983.

Mankiw, N. Gregory, Matthew Weinzierl, and Danny Yagan. 2009. Optimal Taxation in Theory and Practice. *Journal of Economic Perspectives* 23 (4):147–174.

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

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

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

Pouzo, Demian and Ignacio Presno. 2016. Sovereign Default Risk and Uncertainty Premia. *American Economic Journal: Macroeconomics* 8 (3):230–266.

Ramsey, Frank P. 1927. A Contribution to the Theory of Taxation. *Economic Journal* 37 (145):47–61.

Rauh, Joshua D. 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.

Roch, Francisco and Francisco Roldán. 2023. Uncertainty Premia, Sovereign Default Risk, and State-Contingent Debt. *Journal of Political Economy: Macroeconomics* 1 (2):334–370.

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

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.

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

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

Vries, Tjeerd and Alexis Akira Toda. 2022. Capital and Labor Income Pareto Exponents Across Time and Space. *Review of Income and Wealth* 68 (4):1058–1078.