Optimal Labor Income Taxation: A Flexible Moral Hazard Approach

Narayana R. Kocherlakota*

Abstract

This paper reconsiders the question of optimal labor income taxes for the very rich in the context of a flexible moral hazard (FMH) model. In this setting, risk is not exogenous. Rather, each agent can affect the probabilities of all possible income outcomes by allocating a fixed time endowment across a variety of distinct tasks. I prove that the optimal income tax rates on high-end earners and the optimal Pareto tail index of the pre-tax labor income distribution are both endogenously determined by agent preferences. In particular, a society with less risk-averse agents will find it optimal to impose a lower tax rate on the rich, even though its members' choices give rise to a smaller Pareto right tail index. In contrast, this kind of negative co-movement between inequality and optimal tax rates is a suboptimal response in the classical Mirrlees (1971)-Diamond (1998)-Saez (2001) setup to changes in the exogenous distribution of skills.

*Simon Business School (University of Rochester) and NBER. I thank three referees, the editor, Ludwig Straub, and especially Doron Ravid for helpful comments.

1 Introduction

The standard Mirrlees (1971) approach to optimal taxation treats individual-level ex-ante risk (in terms of the determination of skills) as being exogenous. In doing so, the paradigm ignores the many actions that individuals can and do take to influence their income risks. In this paper, I instead assume that individuals' non-contractible or unobserved choice is the allocation of a fixed amount of time across a number of tasks. Each task's time allocation translates directly into the probability of occurrence of a given amount of income. In this way, agents have a distinct influence on the probability of each labor income outcome, and so individual risk is completely controllable.

I characterize optimal income tax schedules within an aggregative version of this *flexible moral hazard* (FMH) model.[1] The main result concerns the taxation of the very rich. In the FMH model studied in this paper, the optimal cross-sectional distribution of labor income and the optimal income tax rate are both determined as endogenous responses to agent preferences. This endogeneity means that the model admits different cross-economy normative implications for tax rates and inequality from those in the classic Mirrlees (1971)-Diamond (1998)-Saez (2001) (MDS) model. In particular, in the FMH model, a society with lower risk aversion will find it optimal to use lower tax rates on the very rich even though the distribution of labor income has more inequality.[2] In the MDS model, this negative covariation between inequality and tax rates is instead a suboptimal response to changes in the exogenous distribution of skills.

I achieve analytical results by using a parameterized model in which the typical agent's non-contractible time allocation problem is equivalent to choosing a probability mass function (pmf) p over an exogenous grid $\{y_i\}_{i=1}^{\infty}$ of possible incomes.[3] The agent's utility loss from choosing a pmf p is given by:

$$\phi \sum_{i=1}^{\infty} p_i^{1+\psi} y_i^{\chi} \tag{1}$$

where all parameters are positive. Because $\chi > 0$, this formulation captures the intuition that it is more costly for an individual to undertake activities that lead to higher incomes. Given a tax schedule $\{T_i\}_{i=1}^{\infty}$ (where T_i is the tax imposed on income y_i), the agent's benefit

[1]Here, I use the language of Georgiadis, Ravid, and Szentes (2024). I discuss the relationship between my model and theirs later in the introduction. Mirrlees (1996) describes the basic principles and results of the moral hazard approach to optimal income taxation when agents make a single scalar decision, as originally studied in Mirrlees (1974) and Varian (1980).

[2]Throughout, I measure inequality using the Pareto tail index of the cross-sectional distribution of income. Higher tail indices mean that the society is more equal. Lower tail indices mean that the society is less equal.

[3]The infinite grid is necessary to address the taxation of very rich. As discussed more fully in the text, the discreteness of the grid sidesteps the problem of optimal tax schedules being indeterminate over sets of Lebesgue measure zero.

from choosing a pmf p is the expectation of a power utility function over after-tax incomes:

$$\sum_{i=1}^{\infty} p_i (y_i - T_i)^{\alpha}, 0 < \alpha < 1. \tag{2}$$

In this parametric setup, the paper obtains two main results. The first is positive. Consider any exogenous tax schedule that is asymptotically linear (for very high incomes). Then the agents' chosen probabilities are consistent with a density that has a Pareto right tail, with a tail index given by:

$$\frac{\chi - \alpha}{\psi} - 1$$

Thus, the empirical observation that the right tail of the cross-sectional income distribution is well-approximated as being Pareto emerges endogenously in the FMH model. The thinness of the right tail is an increasing function of the parameter χ, as higher values of χ imply that agents find it relatively more costly to expend time on high-income outcomes. The right tail is heavier for higher values of α, as agents derive relatively more utility from high-income outcomes. Higher values of ψ also lead to heavier right tails, as agents have a stronger preference for choosing pmfs with "smoother" probabilities.

The second main result is a characterization of the tax schedule that maximizes ex-ante utility, subject to the non-contractibility of the agents' infinite-dimensional probability choices. I prove that the optimal tax schedule is linear for very high incomes, which means that it induces the kind of right tail behavior described above in the cross-sectional distribution of income. The optimal high-income tax rate t^* is given by:

$$\frac{1}{1 + \alpha/\psi}.$$

The formula is, in a sense, standard. Consider the "probability supply" problem of an agent who is choosing what probability p to assign to an income level y given an income-independent tax rate t. If the agent's costs and benefits are defined by (1)-(2), then their decision problem is:

$$\max_p p(1-t)^{\alpha} y^{\alpha} - p^{1+\psi} y^{\chi}.$$

The "elasticity of probability supply" with respect to the after-tax rate in this problem (that is, $\frac{d\ln(p)}{d\ln(1-t)}$) is easily seen to be equal to α/ψ. As a result, it is not surprising that this parameter is the key determinant of the optimal tax rate, or that the parameter χ does not show up in the optimal tax formula.[4]

[4]This intuition is overly simplistic, as it ignores the constraint that probabilities have to sum to one. A key part of the proof of the formula is to show that the multiplier on this constraint in the optimal tax

The FMH model predicts that in an aggregative economy, the optimal tax rate on the very rich and the Pareto tail index of the optimal cross-sectional distribution of income are both increasing functions of the coefficient of relative risk aversion $(1-\alpha)$. In contrast, in the MDS setup, marginal tax rates on the very rich are a decreasing function of the (exogenous) Pareto right tail index of the cross-sectional skill distribution.[5] Note too that the Pareto right tail index derived above is an increasing function of (χ/ψ), which does not appear in the formula for optimal tax rates. This characterization means it may not be optimal to adjust the optimal high-end tax rate in response to changes in right tail inequality. Intuitively, unlike in the MDS model, the shape of the cross-sectional distribution of income is not necessarily reflective of agents' behavioral elasticity with respect to taxes.

The theoretical findings stressed above imply that the coefficient of relative risk aversion (CRRA) $(1-\alpha)$ is a joint determinant of optimal high-income tax rates and right tail inequality. The paper provides some suggestive evidence for this mechanism being at work in the US. In particular, it uses a relatively novel data set from the Federal Reserve to document that, over the past thirty years, high-income earners in the US have increased the fraction of their portfolios invested in relatively risky assets. This shift is consistent with those individuals having experienced a decline in risk aversion. As discussed above, the FMH model implies that it is optimal for societies to respond to such a change in preferences with lower high-end tax rates and higher right-tail inequality. The paper documents that (as is well-known) we have indeed seen these changes in the US since 1980.

The baseline analysis in the paper assumes that agents are ex-ante identical, but ex-post heterogeneous (with their differences determined by the realizations of their income shocks). I consider the robustness of the results to allowing for agents being privately informed about their value of χ (which is restricted to lie in a known finite set). This preference parameter measures the agents' distaste for activities that increase the probability of higher income outcomes, and so $(1/\chi)$ can be seen as being similar to the Mirrleesian notion of skill. This optimal tax problem displays a combination of adverse selection and *infinite-dimensional* moral hazard.

I provide two characterizations of optimal tax systems in this setting in which agents are privately informed about their initial types.[6] The first is that there is no distortion at the top of the top, in the sense that the high-income tax schedule for the high-skill (low χ) types is the same as in the baseline type-identical case (that is, linear with a tax rate of $(1+\alpha/\psi)^{-1}$).

problem has a vanishingly small effect on the solution at asymptotically high income levels.

[5]Scheuer and Werning (2017) obtain a zero covariance result in the MDS framework, in terms of the optimal response to the introduction of superstar technologies.

[6]A justification for these characterizations is in Online Appendix B.

The second is that if the differences in skills are sufficiently small[7], the optimal high-income tax schedules for *all* types are linear with the (common) tax rate of $(1+\alpha/\psi)^{-1}$ (again, the same as in the baseline type-identical case). Thus, the results about optimal taxes in the baseline model with ex-ante homogeneity are robust to adding (a small amount of) ex-ante heterogeneity.

From a theoretical perspective, the paper builds on the recent work of Georgiadis, Ravid, and Szentes (GRS) (2024) on FMH models. However, the papers differ in at least a couple of key respects. The first is technical. GRS study an abstract principal-agent wage contracting problem in which the agent can privately choose an arbitrary probability measure over a *compact* set of output outcomes. In contrast, the current paper assumes that agents confront an unbounded (and hence non-compact) set of incomes. This feature of the model is necessary to treat the optimal taxation of high-income agents, but creates a number of distinct technical challenges.

There is another important (and more economic) distinction. As in this paper, the agent's choices of probabilities in the GRS setup result in a loss of utility. However, GRS model this loss as being described by a monotone mapping from chosen probability distributions into disutility. This approach is distinct from that in this paper which, as sketched above, assumes that individuals are varying probabilities on a "realization-by-realization" basis. This modeling difference means that the analysis in this paper delivers quite different characterizations of optimal arrangements from those in GRS. For example, GRS stress that in their setting, optimal wage contracts may exhibit output distributions with only one or two points in their support.[8] In contrast, in the framework in this paper, if an agent has allocated zero time to a task, there is no marginal loss associated with increasing that time allocation - and the associated probability - slightly. Hence, it proves to be socially optimal to assign positive probability to all income levels above a cutoff. This kind of difference in results makes the model in this paper more appropriate for the study of optimal income taxation.[9]

All proofs are in the appendix.

[7]Specifically, if the maximal and minimal values for χ do not differ by more than $\psi(1-\alpha)$.

[8]See Corollary 3 and Example 2 on page 400 of GRS.

[9]In earlier work, Mattsson and Weibull (MW) (2023) study an optimal contracting problem in which the agents can independently vary the probabilities of a *finite* number of outcomes. But, unlike the current paper, MW model the agent's loss function in terms of deviations from an exogenous baseline positive probability vector. The loss function is such that the agent has infinite marginal benefit from increasing the probability of any outcome above zero.

2 Model

This section describes the basic choice model and establishes the validity of the first-order approach. The baseline model features a unit measure of agents who are ex-ante identical but who are ex-post heterogeneous. The analysis assumes that a Law of Large Numbers (as in Sun (2006)) applies, so that if the ex-ante identical agents receive independent income draws from a common density f, then the ex-post cross-sectional density of income is also given by f. This approach to modeling heterogeneity is the basis of the large macroeconomic literature on inequality that uses the incomplete markets models of Aiyagari (1994) and Huggett (1993). It is also utilized in dynamic public finance (see, for example, Kocherlakota (2010) and Stantcheva (2017)).

Mathematically, agents in this paper's model allocate time so as to choose a probability mass function (pmf) over a fixed grid of a *countably* infinite set of income outcomes. This framework differs from that in the standard MDS model, which features a continuum of income realizations. At least arguably, the discrete approach is more realistic. But there is also an important technical challenge associated with using a continuum in conjunction with the kind of flexible probability choice at the heart of this paper. In such contexts, optimization gives rise to restrictions that are known to be valid only almost everywhere, and so need not be applicable in (arbitrary) sets of Lebesgue measure zero. Of course, these null sets can include a lot of income levels that are salient for taxation (like all of the rationals). This "almost everywhere" versus "everywhere" issue does not arise with the discrete model used in this paper.

2.1 Utility

There is a unit measure of agents with stochastically independent incomes. The agents' incomes have common positive support $S = \{y_i\}_{i=1}^{\infty}$, where:

$$0 < y_i < y_{i+1}, i \geq 1$$
$$lim_{i \to \infty} y_i = \infty.$$

The countable grid of possible incomes will be kept fixed throughout. I assume that the grid is such that for any $\epsilon > 0$, $\sum_{i=1}^{\infty} y_i^{-1-\epsilon}$ is finite. This condition is satisfied if the income grid takes the form $y_i = \Psi i$ for some positive constant Ψ.

There is a government that imposes a tax schedule $T \in \mathrm{T}$, where:

$$\mathrm{T} = \{T \in \mathbb{R}^{\infty} | (y_i - T_i) \geq 0 \text{ for all } i \in \mathbb{N}\}$$

A schedule T implicitly defines tax payments as a function of income. Taxes cannot be larger than income but may be negative (that is, transfers).

Each agent is endowed with one unit of time, which they can split across a countably infinite number of tasks indexed by the natural numbers. (Note that time is not freely disposable - it must be spent on these activities.) Their time choice can thus be represented as an element of the set of probability mass functions $\mathcal{F} \equiv \{f \in [0,1]^\infty | \sum_{i=1}^{\infty} f_i = 1\}$. Choosing a time f_i for task i translates one-for-one into the probability of occurrence of income y_i. Consequently, an agent's choice of a time allocation f in F is equivalent to their choice of a probability mass function (pmf) f in F.

If an agent chooses a pmf f, and faces a tax schedule T, they derive ex-ante expected utility:

$$U(f;T) = \sum_{i=1}^{\infty} (y_i - T_i)^\alpha f_i - \phi \sum_{i=1}^{\infty} f_i^{\psi+1} y_i^\chi \tag{3}$$

$$0 < \alpha < 1$$
$$\psi > 0$$
$$\chi > 0$$

The first summand of the utility function represents the expectation of a concave power utility over random consumption (here, equal to after-tax income). The second component of the utility function is the agent's disutility from their time allocation f. As is standard in many applications, the time disutility takes a convex power form and is additively separable across tasks.

The agent's objective function is governed by only four parameters: ϕ, α, χ, and ψ. The first parameter ϕ determines the level of the agent's disutility, and plays little role in the main results. The parameter α determines the curvature of the agent's utility function over consumption. The parameter χ captures the agent's relative disutility from using their time on tasks that increase the probabilities of higher levels of output. In a sense, χ^{-1} can be viewed as being similar to skill in the Mirrleesian setup (and will indeed play exactly this role in the analysis of ex-ante type heterogeneity later in the paper). The parameter ψ determines the agent's willingness to substitute their time between different kinds of tasks. As a result, when ψ is large, the agent has a strong preference for flatter income pmfs that put more weight on high income realizations.

In Section 3, we shall see that agents with these preferences respond to realistic tax schedules by choosing pmfs with empirically plausible right tails. This result means that this class of utility functions can be seen as well-suited to address questions about optimal taxation of high-income earners.

2.2 Implementation

Let T be any tax schedule in T. In response to this tax schedule, agents are free to choose any pmf $f \in \mathcal{F}$. They make this choice by solving the problem:

$$f \in \arg\max_{g \in \mathcal{F}} \sum_{i=1}^{\infty} (y_i - T_i)^{\alpha} g_i - \phi \sum_{i=1}^{\infty} g_i^{\psi+1} y_i^{\chi}$$

Notice that the agent's problem has a strictly concave (in g) objective and linear constraints. As a result, we can fully characterize the unique solution using the relevant first-order condition. Let κ be the multiplier (of indeterminate sign) on the constraint that the pmf sums to one. Then, we can differentiate with respect to g to arrive at the following first-order condition that applies for all $i \in \mathbb{N}$:

$$\begin{aligned} &(y_i - T_i)^{\alpha} = \phi(\psi + 1) f_i^{\psi} y_i^{\chi} + \kappa, f_i > 0 \\ &(y_i - T_i)^{\alpha} \leq \kappa, f_i = 0 \end{aligned} \tag{4}$$

The following proposition[10] uses this first-order condition to provide necessary and sufficient conditions for a pmf f to be implemented by T.

Proposition 1. *Consider any tax schedule $T \in$ T. Suppose $f \in \mathcal{F}$ is a pmf such that:*

$$\sum_{i=1}^{\infty} f_i (y_i - T_i)^{\alpha} < \infty$$
$$\sum_{i=1}^{\infty} f_i^{1+\psi} y_i^{\chi} < \infty$$

Then f is the unique pmf implemented by T if and only if there exists $\kappa \in \mathbb{R}$ such that for all $i \in \mathbb{N}$:

$$f_i = \left(\frac{max((y_i - T_i)^{\alpha} - \kappa, 0)}{y_i^{\chi} \phi(1 + \psi)}\right)^{1/\psi}$$

The key part of this proposition is that the two summability conditions imply that the agent's objective is finite when evaluated at f. It is then straightforward to show that (4) is a necessary and sufficient condition for the optimality of f.

[10] It is possible to derive a similar result from GRS's Lemma 1 (which does not rely on their maintained monotonicity assumption). I thank Doron Ravid for communicating this point to me.

Proposition 1 does not establish the existence of a solution to the agent's problem. The next proposition provides conditions on T and the parameters (χ, α, ψ) that ensure that T does indeed implement a unique pmf.

Proposition 2. *Consider any tax schedule $T \in \mathrm{T}$ such that for some positive finite b, $b \geq (1 - T_i/y_i)$ for all $i \in \mathbb{N}$. Suppose too that:*

$$\frac{\chi - \alpha}{\psi} > 2$$

For any real κ, define $f(\kappa) \in \mathbb{R}^\infty$ as:

$$f_i(\kappa) = (\frac{max((y_i - T_i)^\alpha - \kappa, 0)}{y_i^\chi \phi(\psi + 1)})^{1/\psi}, i \in \mathbb{N}.$$

There exists a unique κ such that $f(\kappa)$ is in $\mathcal{F}$ and it is the unique pmf implemented by T.

The restriction on T rules out the possibility that the agent's after-tax income rises rapidly as income converges to infinity. The proof works by showing that the sum of the components of $f(\kappa)$ is finite for all κ, and then establishing that there is a unique κ such that the sum is one (meaning that f is a well-defined pmf). It is then straightforward to show that, given the assumption about (χ, α, ψ), the pmf f satisfies the summability conditions in Proposition 1.

Propositions 1 and 2 represent a large contrast with standard principal-agent problems, in which the agent's choice variable is one-dimensional. In those settings, the agent's first-order necessary condition is only sufficient under relatively strong assumptions about the relevant *exogenous* uncertainty (Rogerson (1985)). As a result, it can be difficult to translate the requirement of implementability into a mathematically manageable characterization. In the current environment, the first-order condition is necessary and sufficient under the simple and natural requirement that the agent's loss function is convex over probabilities.

3 Positive Results for Exogenous Tax Schedules

In the standard Mirrleesian model, the (assumed) right tail of the skill distribution is the key driver in the determination of the right tail of the income distribution. This section shows that, in the FMH model, the shape of the right tail of the pmf of income is an endogenous response to individual preferences and the tax system.

3.1 An Endogenous Right Tail

The following proposition shows how the right tail of the agent's chosen pmf is shaped by the asymptotic progressivity of the tax schedule and the agent's utility function. It considers a class of tax schedules that asymptote to the widely used approximation to US tax schedules (Feldstein (1969) or Heathcote, Storesletten, and Violante (2017)):

$$T_i = y_i - \Gamma y_i^{\eta}, 0 \leq \eta \leq 1$$

Here, η is a measure of regressivity (in the sense that T is more regressive for larger values of η).

The proposition, and the remainder of the paper, uses the following notation and terminology. Given a positive $z \in \mathbb{R}_+^{\infty}$, I write that:

$$z \rightarrow_{\infty} y^{\gamma}, \gamma \neq 0$$

if there exists finite and positive L such that:

$$lim_{i \rightarrow \infty} z_i y_i^{-\gamma} = L.$$

If $\gamma > 0$, this notation means that z_i converges to infinity at the same rate as $y_i^{-\gamma}$. If $\gamma < 0$, then this notation means that z_i converges to zero at the same rate as y_i^{γ}. Similarly, a pmf f is said to be asymptotically Pareto with a tail index $(\eta - 1), \eta > 2$, if:

$$f \sim_{\infty} y^{-\eta}.$$

The idea behind this terminology is that for large y, the implied values of f over the (discrete) grid S are well-described by the relevant (continuous) Pareto density.

Proposition 3. *Suppose $T \in \mathrm{T}$ is a tax schedule such that:*

$$(y - T) \sim_{\infty} y^{\eta}$$

where $1 \geq \eta > 0$. Suppose too that:

$$\frac{\chi - \alpha}{\psi} > 2.$$

For any $\kappa \in \mathbb{R}$, define:

$$\hat{f}_i(\kappa) \equiv (\frac{max((y_i - T_i)^{\alpha} - \kappa, 0)}{\phi(1 + \psi) y_i^{\chi}})^{\frac{1}{\psi}}.$$

There exists a unique κ^* such that:

$$\sum_{i=1}^{\infty} \hat{f}_i(\kappa^*) = 1$$

and $\hat{f}(\kappa^*)$ is the unique pmf implemented by T. The implemented pmf $\hat{f}$ satisfies:

$$\hat{f}(\kappa^*) \sim_{\infty} y^{\frac{\alpha\eta-\chi}{\psi}}$$

and so is asymptotically Pareto with tail index $(\frac{\chi-\alpha\eta}{\psi} - 1)$.

Proposition 3 considers a wide range of tax schedules and shows that, given the parametric class of utility functions, agents respond to these tax schedules by choosing a pmf which is consistent with the cross-sectional data on income in the sense of being asymptotically Pareto. The size of the tail index is lower (reflecting more inequality) when the tax schedule is more regressive (higher η). Similarly, the Pareto tail index is lower when the agent is less averse to high-income tasks (lower χ) or to risk (higher α). The chosen pmf is flatter (and so less equal) when the agent views the various tasks as less substitutable (that is, for larger values of ψ). The restriction on $(\alpha, \eta, \chi, \psi)$ in the proposition ensures that the agent's aversion to high-income tasks is sufficiently strong that income has a well-defined mean.

3.2 Asymptotically Linear Tax Schedules

Many countries use tax schedules that are asymptotically linear. The following corollary specializes Proposition 3 to this class of tax schedules by setting $\eta = 1$.

Corollary 1. *Suppose the tax schedule $T \in \mathrm{T}$ and is asymptotically linear, so that for some constant $h \in (0, 1)$:*

$$lim_{i\to\infty} T_i/y_i = h.$$

Suppose that:

$$\frac{\chi - \alpha}{\psi} > 2$$

and T implements f (as described in Proposition 3). Then:

$$f \sim_{\infty} y^{\frac{\alpha-\chi}{\psi}}$$

so that the implemented pmf is asymptotically Pareto with a tail index $(\frac{\chi-\alpha}{\psi} - 1)$.

We shall see in the next section that the optimal tax schedules are asymptotically linear. Hence, the right tail of the optimal pmf of income is as described in Corollary 1.

4 Normative Results

In this section, we study the behavior of optimal tax rates on high-income earners in the FMH model. The main result is that the asymptotically optimal tax rate is a function of the preference parameter (α/ψ). However, it is independent of the preference parameter χ, which (as we have seen) is nonetheless a key determinant of the right tail of the income distribution.

4.1 Optimal Taxation: Setting Up the Problem

This subsection describes the optimal tax problem and then converts it into a manageable mathematical formulation.

4.1.1 The Tax Problem

The optimal tax problem is to choose a tax schedule $T \in \mathrm{T}$ that maximizes ex-ante utility subject to three constraints. The first constraint restricts the net tax collections to be non-negative:

$$\sum_{i=1}^{\infty} T_i f_i \geq 0$$

The second constraint is that the pmf f is implemented by the tax schedule T. The final constraint is an upper bound on f of the following form:

$$f_i \leq M y_i^{-\bar{\theta}}, i \in \mathbb{N}$$

where M is sufficiently large to admit a non-empty set of pmfs:

$$M \sum_{i=1}^{\infty} y_i^{-\bar{\theta}} > 1.$$

The upper bound will serve a couple of related purposes. If $\bar{\theta} > 2$, then the upper bound ensures that the pmf f has a first moment. Also, it will facilitate a proof of existence of an optimum. But I will always set the upper bound so that it does not bind. Accordingly, throughout the remainder of this section, I impose the following two restrictions on the parameters:

$$\bar{\theta} > 2$$
$$\frac{\chi - \alpha}{\psi} > \bar{\theta}.$$

Overall, given M and $\bar{\theta}$, the optimal tax problem $P(M,\bar{\theta})$ is:

$$\max_{(T,f)\in \mathrm{T}\times\mathcal{F}} \sum_{i=1}^{\infty}(y_i - T_i)^{\alpha} f_i - \phi\sum_{i=1}^{\infty} f_i^{1+\psi} y_i^{\chi}$$

$$s.t.\ T \text{ implements } f \tag{5}$$

$$\sum_{i=1}^{\infty} T_i f_i \geq 0 \tag{6}$$

$$M y_i^{-\bar{\theta}} \geq f_i \geq 0, i\in\mathbb{N} \tag{7}$$

$$y_i \geq T_i, i\in\mathbb{N} \tag{8}$$

The upper bound on f is incorporated via the constraint (7).

4.1.2 The Implementability Constraint

This subsection uses Proposition 1 to derive a simple formulation of the implementability constraint in the optimal tax problem $P(M,\bar{\theta})$.

Proposition 1 showed that, subject to two summability conditions, a tax schedule T implements f if and only if there exists κ such that for all $i\in\mathbb{N}$:

$$T_i = y_i - (\phi(1+\psi) f_i^{\psi} y_i^{\chi} + \kappa)^{1/\alpha},\ \text{ if } f_i > 0 \tag{9}$$

$$T_i \geq y_i - \kappa^{1/\alpha},\ \text{ if } f_i = 0 \tag{10}$$

Note that, since $(y_i - T_i) \geq 0$ for all $i\in\mathbb{N}$, these conditions imply that:

$$\kappa \geq -\inf_{i\in\mathbb{N}} \phi(1+\psi) f_i^{\psi} y_i^{\chi}.$$

If we substitute (9)-(10) into the objective of problem $P(M,\bar{\theta})$, we obtain:

$$\phi\psi\sum_{i=1}^{\infty} f_i^{\psi+1} y_i^{\chi} + \kappa$$

Similarly, if we substitute (9)--(10) into the resource constraint, we obtain:

$$\sum_{i=1}^{\infty}(y_i - (\phi(1+\psi) f_i^{\psi} y_i^{\chi} + \kappa)^{1/\alpha}) f_i \geq 0$$

We can then reformulate the optimal taxation problem $P(M,\bar{\theta})$ by substituting out for the tax schedule to obtain the problem $P^*(M, \bar{\theta})$:

$$\max_{\kappa \in \mathbb{R}, f \in \mathcal{F}} \phi\psi \sum_{i=1}^{\infty} f_i^{1+\psi} y_i^{\chi} + \kappa$$

$$s.t. \sum_{i=1}^{\infty} y_i f_i \geq \sum_{i=1}^{\infty} (\phi(1+\psi) f_i^{\psi} y_i^{\chi} + \kappa)^{1/\alpha} f_i \tag{11}$$

$$M y_i^{-\bar{\theta}} \geq f_i \geq 0, i \in \mathbb{N} \tag{12}$$

$$\kappa \geq -\inf_{i \in \mathbb{N}} \phi(1+\psi) f_i^{\psi} y_i^{\chi} \tag{13}$$

The restriction (13) captures the requirement that κ is sufficiently large to keep after-tax income non-negative for all i. The restrictions (11)-(12) ensure that f satisfies the resource constraint and the upper bound restriction.

The following two lemmas summarize the equivalence between the two problems $P(M, \bar{\theta})$ and $P^*(M, \bar{\theta})$. Both require that $\bar{\theta} > 2$, so that the pmf f has a finite mean. The first lemma shows that the constraint set of $P(M, \bar{\theta})$ is a superset of the constraint set of $P^*(M, \bar{\theta})$. The proof is largely a check that, given $\bar{\theta} > 2$, the summability conditions in Proposition 1 are satisfied because of the upper bound on f.

Lemma 1. *Let $\bar{\theta} > 2$. Suppose (f, κ) is an element of the constraint set to $P^*(M, \bar{\theta})$. Then (T, f) is an element of the constraint set to $P(M, \bar{\theta})$, where T is defined as:*

$$T_i = y_i - (\phi(1+\psi) f_i^{\psi} y_i^{\chi} + \kappa)^{1/\alpha}, i \in \mathbb{N}$$

As well, (T, f) has the same value for the objective:

$$U(f; T) = \phi\psi \sum_{i=1}^{\infty} f_i^{1+\psi} y_i^{\chi} + \kappa.$$

The second lemma proves the converse.

Lemma 2. *Let $\bar{\theta} > 2$. Suppose (T, f) is an element of the constraint set to $P(M, \bar{\theta})$. Then, (f, κ) is in the constraint set to $P^*(M, \bar{\theta})$, where κ satisfies:*

$$\kappa = (y_i - T_i)^{\alpha} - \phi(1+\psi) f_i^{\psi} y_i^{\chi}$$

for any $i \in \mathbb{N}$ such that $f_i > 0$. As well, (f, κ) has the same value for the objective:

$$U(f; T) = \phi\psi \sum_{i=1}^{\infty} f_i^{1+\psi} y_i^{\chi} + \kappa.$$

4.2 Main Theorem

The two lemmas imply that we can characterize the solutions to $P(M, \bar{\theta})$ by examining the properties of the solutions to the more mathematically tractable $P^*(M, \bar{\theta})$ and then translating those into properties of the corresponding tax schedules. The main theorem uses this approach to characterize the tail behavior of optimal tax schedules.

Theorem 1. *Let $\bar{\theta} > 2$, and suppose:*

$$\frac{\chi - \alpha}{\psi} > \bar{\theta}.$$

Suppose (T^*, f^*) is a solution to $P(M, \bar{\theta})$. Then:

$$lim_{i \to \infty} y_i^{-1} T_i^* = (1 + \alpha/\psi)^{-1}$$

and:

$$f^* \sim_\infty y^{\frac{\alpha - \chi}{\psi}}.$$

Hence, the implied optimal pmf of income is asymptotically Pareto with a tail index of

$$\frac{\chi - \alpha}{\psi} - 1.$$

Theorem 1 shows that optimal tax schedules are asymptotically linear, with a limiting tax rate that depends only on the preference parameters α and ψ. The optimal tax rate is increasing in risk aversion and in the agent's desire to equate the probabilities of the various income outcomes. Consistent with Corollary 1, the asymptotically linear tax schedule induces agents to choose a pmf that is asymptotically Pareto with a tail index given by $(\frac{\chi - \alpha}{\psi} - 1)$. The next subsection provides a sketch of the proof of Theorem 1 (with the complete details reserved for the Appendix).

Theorem 1 is a necessary condition that any optimal tax schedule and its implied pmf of income must satisfy. The following proposition provides sufficient conditions for the existence of an optimum.

Proposition 4. *Suppose that as in the hypothesis of Theorem 1:*

$$\frac{\chi - \alpha}{\psi} > \bar{\theta} > 2$$

Suppose too that $\bar{\theta}$ is sufficiently close to $\frac{\chi-\alpha}{\psi}$ so that:

$$(\bar{\theta}+\frac{\alpha}{\psi}(\bar{\theta}-2))>\frac{\chi-\alpha}{\psi}.$$

Then there exists a solution to $P(M,\bar{\theta})$ for any value of M.

The proof establishes the compactness of the constraint set and the continuity of the objective function in the product topology. The compactness result relies on the f_i's being required to converge to zero sufficiently rapidly, as implied by the auxiliary condition on $\bar{\theta}$.

4.3 A Proof Sketch

In this subsection, I sketch the proof of Theorem 1. I abstract from the bounds on f and κ in $P^*(M,\bar{\theta})$, and assume that the optimal tax schedule T^* is such that the limit $\lim_{i\to\infty} T_i^*/y_i$ exists. However, the proof of Theorem 1 (in the appendix) is rigorous in its treatment of these issues.

4.3.1 First-Order Condition

Consider a solution (f,κ) to the maximization problem $P^*(M,\bar{\theta})$. Let λ be the (positive multiplier) on the resource constraint and μ be the multiplier on the "probabilities sum to one" constraint. Define the Lagrangian to the problem as:

$$L((f_i)_{i=1}^{\infty})=\phi\psi\sum_{i=1}^{\infty}f_i^{1+\psi}y_i^{\chi}+\lambda\sum_{i=1}^{\infty}(y_i-(\phi(1+\psi)f_i^{\psi}y_i^{\chi}+\kappa)^{1/\alpha})f_i+\mu(1-\sum_{i=1}^{\infty}f_i)$$

Let $c_i=(\phi(1+\psi)f_i^{\psi}y_i^{\chi}+\kappa)^{1/\alpha}$ represent the after-tax income associated with income y_i. Then (assuming interiority as noted above), each f_i must satisfy the first-order condition:

$$\mu=\psi(c_i^{\alpha}-\kappa)+\lambda(y_i-c_i)-\lambda\psi\alpha^{-1}\phi(1+\psi)f_i^{\psi}y_i^{\chi}(\phi(1+\psi)f_i^{\psi}y_i^{\chi}+\kappa)^{1/\alpha-1} \quad (14)$$

$$=(\psi c_i^{\alpha}-\psi\kappa)+\lambda(y_i-c_i)-\lambda\psi\alpha^{-1}(c_i^{\alpha}-\kappa)c_i^{1-\alpha} \quad (15)$$

The first term on the right hand side is the derivative of the objective with respect to f_i. The second term is the incremental revenue raised from those agents who earn y_i. The final term is the derivative of $(-c_i)$ in the resource constraint with respect to f_i.

4.3.2 An Upper Bound on the Optimal Asymptotic Tax Rate

We can rewrite the above first-order condition as:

$$\mu = y_i(\psi \hat{c}_i^{\alpha} y_i^{\alpha-1} - \psi\kappa y_i^{-1} + \lambda - \lambda\hat{c}_i - \lambda\frac{\psi}{\alpha}\hat{c}_i^{1-\alpha}(\hat{c}_i^{\alpha} - \kappa y_i^{-\alpha})) \tag{16}$$

where $\hat{c}_i = c_i/y_i, i \in \mathbb{N}$. Suppose that:

$$\lim_{i\to\infty} \hat{c}_i < (1+\psi/\alpha)^{-1}.$$

Then (16) implies:

$$\begin{aligned}\mu &\geq lim_{i\to\infty} y_i(-0 + \lambda - \lambda(1+\psi/\alpha)\hat{c}_i). \\ &= \lambda(\lim_{i\to\infty} y_i)(1 - (1+\psi/\alpha)\lim_{i\to\infty}\hat{c}_i) \\ &= \infty.\end{aligned}$$

which is a contradiction. This kind of tax schedule is suboptimal because it engenders a pmf which puts too little mass on high income realizations.

4.3.3 A Lower Bound On the Optimal Asymptotic Tax Rate

Suppose instead that:

$$\lim_{i\to\infty} \hat{c}_i^{-1} < (1+\psi/\alpha)$$

Then $\lim_{i\to\infty} c_i = \lim_{i\to\infty} \hat{c}_i y_i = \infty$, and (14) implies that:

$$\begin{aligned}\mu &= \lim_{i\to\infty} c_i(\psi c_i^{\alpha-1} - \psi\kappa c_i^{-1} + \lambda(1/\hat{c}_i - 1) - \lambda\frac{\psi}{\alpha}(1 - \kappa c_i^{-\alpha})) \\ &= \lim_{i\to\infty} c_i \times \lim_{i\to\infty} \lambda(1/\hat{c}_i - 1 - \psi/\alpha) \\ &< -\infty.\end{aligned}$$

which is a contradiction. This kind of asymptotic after-tax schedule is suboptimal because it induces a pmf which puts too much mass on high income realizations.

It follows that the optimal after-tax schedule is asymptotically linear, and the optimal asymptotic tax rate is $1 - (1+\psi/\alpha)^{-1} = (1+\alpha/\psi)^{-1}$. Note that the above analysis implies that, for large incomes, an optimal tax schedule is fully determined by the Lagrangian term based on the resource constraint:

$$\lambda(y_i - c_i) - \lambda\psi\alpha^{-1}(c_i^{\alpha} - \kappa)c_i^{1-\alpha}$$

5 Other Considerations

This section discusses the connections between the above normative analysis and empirical evidence on taxes and inequality. It also assesses the robustness of the optimal taxation results to allowing for a small amount of ex-ante heterogeneity in a characteristic about which agents are privately informed. Finally, it compares the optimal income pmf in the FMH setup to that in a version of the model without any informational constraints.

5.1 Suggestive Evidence About Risk Aversion and Taxes

This subsection discusses how the FMH model's normative implications can be seen as providing a potential interpretation for the empirically observed joint evolution of inequality and tax rates. In a recent paper, de Vries and Toda (2022) document that the labor income Pareto exponent has fallen sharply since 1980 in the US (among other countries). At the same time, it is well-known that the top marginal tax rate has fallen.[11] In the MDS model, this kind of co-movement between inequality and the top marginal tax rate is typically viewed as suboptimal. In that framework, the change in the distribution of labor income is seen as being driven by a fall in the tail index in the (exogenous) distribution of skills. The optimal tax rate on very high incomes should rise in response.

The FMH model implies instead that a simultaneous decline in tax rates on high-income earners and the Pareto index of the right tail may be optimal if the latter is triggered by a decline in the exogenous CRRA $(1-\alpha)$ over time. The following calculation is indicative of the relevant magnitudes involved. Piketty (2014) documents that the high-end marginal tax rate fell from around 70% in the early 1980s to about 35% in the early 2010s. Suppose that the tax rate was optimal in both periods. Then, the FMH model's optimal tax formula (Theorem 1) implies that:

$$CRRA_{old} = 1 - (1/0.7 - 1)\psi \approx 1 - 0.43\psi$$
$$CRRA_{new} = 1 - (1/0.35 - 1)\psi \approx 1 - 1.86\psi.$$

where $CRRA_{old}$ and $CRRA_{new}$ represent the coefficients of relative risk aversion at the two dates. The difference between $CRRA_{old}$ and $CRRA_{new}$ is 1.43ψ (which may be small if the parameter ψ is, meaning that agents view the time spent on the various activities as near-substitutable). Given this fall in risk aversion, the model implies that, for any fixed values

[11] See Piketty (2014).

of ψ and χ, the Pareto tail index associated with asymptotically linear taxes would fall by:

$$\frac{CRRA_{new} - CRRA_{old}}{\psi} = 1.43$$

Critically, this implied change in tail inequality depends only on the observed tax rates 0.7 and 0.35, and so is independent of ψ or χ. It is roughly consistent with the empirical decline in the Pareto tail index over the same period (de Vries and Toda (2022), p. 1066).

Is there evidence in support of the requisite decline in risk aversion (which, as noted above, may be small)? In 2019, the Federal Reserve released a new data set, the Distributional Financial Accounts (DFA), that builds on the pre-existing triennial Survey of Consumer Finances and the quarterly flow of funds report. Among many other useful features, the DFA allows researchers to track the portfolio allocations of high-income Americans over time. Table 1 documents that the top income percentile has increased the share of their portfolios invested in corporate equities (as well as the share invested in either corporate equities or private businesses). At the same time, there is little evidence of an increase in the equity risk premium (Smith and Timmerman (2022), Figure 2a, p. 559) or a reduction in asset return volatility.[12] Hence, this increase in the share of risky assets (viewed, for example, through the lens of the textbook Merton portfolio allocation formula) is consistent with a decline in risk aversion.[13]

In keeping with the focus of the paper on high-income earners, Table 1 contains evidence about the top income percentile. However, in the model, α is a common preference parameter for all agents and so an increase in α should lead to an increase in the shares of risky assets in the portfolios of all investors. Tables 2-3 in Online Appendix A use data from the DFA to document that lower income investors have also shifted their portfolios toward more risky assets.

5.2 Consequences of Other Inequality Drivers

The prior subsection explored the consequences for right tail inequality and optimal asymptotic tax rates of an increase in α, keeping the other preference parameters unchanged. But the formula in Corollary 1 implies that (assuming asymptotically linear taxes) the decline in the Pareto exponent documented by de Vries and Toda (2022) could also be due to a decline

[12] Using CBOE data, my calculations are that the average value of the VIX from 1990-2007 was 19, while the average value from 2007-2024 was 20. (Chicago Board Options Exchange, CBOE Volatility Index: VIX [VIXCLS], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/VIXCLS, November 12, 2024.)

[13] The evidence is only suggestive. It bears other interpretations, including a possible increase in the relative wealths of those agents who are less risk averse.

Table 1: Portfolio Shares of the Top 1% of the Income Distribution

Year	Corp. Equity + Mutual Fund Shares	Private Business	Defined Contribution Pensions	Real Estate	Defined Benefit Pensions	Consumer Durables	Other Assets
1989	0.21	0.24	0.03	0.15	0.03	0.07	0.27
2001	0.28	0.20	0.04	0.18	0.02	0.04	0.23
2013	0.37	0.18	0.05	0.14	0.01	0.03	0.21
2024	0.47	0.15	0.04	0.15	0.02	0.02	0.16

The data are from the Distributional Financial Accounts (DFA). The entries are the fraction of the value of assets held in the relevant form by the top income percentile. The yearly data is based on the fourth quarter. Note that 1989:Q3 is the earliest observation in the data set. The largest of the (many) components of the residual category "other assets" are deposits, money market mutual fund shares, and US government and municipal securities.

in χ or an increase in ψ. In this subsection, I discuss the impact of these kinds of parameter changes on the optimal high-income tax rate.

Recall that Theorem 1 shows that the optimal asymptotic tax rate is given by:

$$\frac{1}{1+\alpha/\psi} \tag{17}$$

Hence, a decrease in χ that increases right tail labor income inequality has *no* effect on the optimal tax rate on high-income earners. Intuitively, while χ does impact the agents' willingness to engage in high-income tasks, it does not affect their willingness to *substitute* between the various activities in response to tax changes. Hence, it is suboptimal for a society to adjust the high-income tax rate in response to changes in χ.

On the other hand, the tax formula (17) implies that if ψ rises, then right tail labor income inequality and optimal asymptotic tax rates should both rise. The intuition behind this implication is standard. An increase in ψ means that agents are less elastic in terms of their willingness to substitute across the various tasks. It is socially optimal to respond to this reduced elasticity of substitution with a higher tax rate.

To sum up: the FMH model implies that the optimal response of tax rates on high-income earners to an increase in labor income inequality depends on the source of that increase. The observed fall in the tax rate on high-income earners in the US data may be socially optimal, if the increase in inequality is due to an increase in α (a fall in risk aversion). But a reduction in tax rates on high-income earners is a necessarily suboptimal response to an inequality increase if its underlying driver is a fall in χ or a rise in ψ. It would be useful in future research to seek additional sources of evidence that shed light on the evolution of ψ and χ (much as Section 5.1 used the DFA to impute the evolution of α).

5.3 Privately Observed Ex-Ante Types

In the baseline analysis, agents are ex-ante identical. This subsection extends the analysis to allow for ex-ante heterogeneity with respect to a characteristic that is privately known to the agents. The resulting optimal tax problem is a complex one, as it features a combination of adverse selection and an infinite-dimensional moral hazard problem. The main finding is in the nature of a robustness check: the above characterization of optimal high-income tax schedules remains valid as long as the amount of ex-ante heterogeneity is sufficiently small.

Suppose there are N types of agents, where the types are indexed by $n \in \{1, 2, 3, .., N\}$. For each n, there is a fraction π_n of type n agents in the population. The type n agents have preference parameters given by $(\alpha, \psi, \phi, \chi_n)$, where $\chi_n < \chi_{n+1}, n \in \{1, 2, 3, .., N-1\}$.

This specification means that type n agents experience less relative disutility from assigning higher probabilities to high-income states than type $(n+1)$ agents. Hence, the type n agents can be viewed as more skilled in the Mirrleesian sense than the type $(n+1)$ agents. As in Section 2, we can define a type n agent's utility from a tax schedule $T \in \mathrm{T}$ and a pmf choice f in $\mathcal{F}$ as:

$$U_n(f;T) = \sum_{i=1}^{\infty}(y_i - T_i)^{\alpha} f_i - \phi \sum_{i=1}^{\infty} f_i^{1+\psi} y_i^{\chi_n}, n \in \{1, 2, ..., N\}$$

Agents' types are private information and, as in the baseline model, their choices of pmf's are unobserved (non-contractible). From the Revelation Principle, we can identify the full range of achievable outcomes by allowing agents to self-select between type-specific tax schedules $(T^n)_{n=1}^N \in \mathrm{T}^N$ and simultaneously choosing their optimal pmf. However, one issue that emerges in the definition of incentive-compatibility is how a tax schedule should respond to the occurrence of zero-probability events. Thus, suppose a type m agent selects a type n tax schedule, and chooses $f_i > 0$ even though a type n agent would choose $f_i^n = 0$. In principle, T^n could be structured so as to punish type m agent severely for this *detectable deviation*, given that the punishment would not affect a type n agent.

Instead, I limit these punishments for detectable deviations in a fashion that is both technically convenient and (more arguably) realistic. Specifically, I define a vector $(f^n)_{n=1}^N \in \mathcal{F}^{\mathcal{N}}$ of type-dependent pmf choices to be implementable by a vector of tax schedules $(T^n)_{n=1}^N \in \mathrm{T}^N$ if there exists a vector $(\kappa^n)_{n=1}^N \in \mathbb{R}^N$ such that for all $n \in \{1, 2, 3, ..., N\}$:

$$T_i^n = y_i - (\phi(1+\psi) y_i^{\chi_n} (f_i^n)^{\psi} + \kappa^n)^{1/\alpha} \text{ for all } i \in \mathbb{N} \tag{18}$$

$$U_n(f^n;T^n) \geq \max_{f \in \mathcal{F}} U_n(f;T^m) \text{ for all } m \neq n. \tag{19}$$

Given Proposition 1, (18) implies that a type n agent chooses the pmf f^n when confronted with tax schedule T^n. Note though that, relative to what is allowed in Proposition 1, (18) embeds a punishment limit for detectable deviations, because it restricts taxes in zero probability states to be close to taxes in near-zero-probability states. The second requirement (19) says that type n agents prefer their type-specific tax schedule and pmf to choosing any other tax schedule and pmf.

As in Section 4, we consider a social planner who seeks to choose the implementable outcome that maximizes a weighted average of the types' utilities, where the positive weights are given by $(\omega_1, \omega_2, ..., \omega_N)$. Thus, the heterogeneous agent planner's problem (HAPP) is defined as:

$$\max_{(T^n,f^n)_{n=1}^N} \sum_{n=1}^{N} \omega_n U_n(f^n;T^n) \tag{20}$$
$$\sum_{n=1}^{N}\sum_{i=1}^{\infty} \pi_n f_i^n T_i^n \geq 0$$
$$(T^n)_{n=1}^N \text{ implements } (f^n)_{n=1}^N$$
$$My_i^{-\bar{\theta}} \geq f_i^n \geq 0, i \in \mathbb{N}, 1 \leq n \leq N$$
$$y_i \geq T_i^n, i \in \mathbb{N}, 1 \leq n \leq N$$

This is a direct generalization of the social planner's problem in Section 4.1.1 to allow for multiple types and to allow for the full set of deviations that are possible in this adverse selection-moral hazard problem. As in Theorem 1, we ensure summability by assuming $\bar{\theta}$ satisfies:

$$\frac{\chi_1 - \alpha}{\psi} > \bar{\theta} > 2.$$

This abstract optimal taxation problem is easiest to interpret in the case in which $N = 2$, so that agents are either "high-skilled" or "low-skilled". Then, the planner's problem is to specify a tax schedule for those who have opted for at least some amount of social support ("low-skilled") and a separate tax schedule for those who have not ("high-skilled"). The incentive problem then allows any person to choose to receive social support or not, and then respond to the resulting tax schedule by choosing an allocation of time (f) from the infinite-dimensional set $\mathcal{F}$.

The following remarks[14] provide two characterizations of a solution $(T^n, f^n)_{n=1}^N$ to the problem HAPP. Given the complexity of potential deviations, the usual logic behind the no-distortion-at-the-top principle does not immediately apply. However, the first characterization of optimal taxes is that there is *no-distortion-at-the-top-of-the-top*: that is, it is suboptimal to distort the *high-income* part of (the high-skill) type 1's tax schedule relative to the case in which there is no ex-ante heterogeneity.

Remark 1. Suppose $(T^n, f^n)_{n=1}^N$ is a solution to HAPP. Then:

$$\lim_{i\to\infty} T_i^1/y_i = \frac{1}{1+\alpha/\psi}$$
$$f^1 \sim_\infty y^{\frac{\alpha-\chi_1}{\psi}}.$$

The second characterization shows that when the types are sufficiently close together, it

[14] See Online Appendix B for a heuristic justification based on that in Section 4.3.

is suboptimal to distort *any* type's high-income tax schedule.

Remark 2. Suppose $(\chi_N - \chi_1) < \psi(1-\alpha)$ and that $(T^n, f^n)_{n=1}^N$ is a solution to HAPP. Then for any $n \in \{1, 2, ..., N\}$:

$$\lim_{i\to\infty} T_i^n / y_i = \frac{1}{1+\alpha/\psi}$$
$$f^n \sim_\infty y^{\frac{\alpha-\chi_n}{\psi}}.$$

In both characterizations, the relevant incentive-compatibility constraints may be binding. But, even if binding, the incentive-compatibility constraints have a vanishingly small effect on the tax schedule at *asymptotically* high income levels.

To sum up, even in the presence of Mirrleesian ex-ante private information about skill types (disutility of generating high incomes), the optimal tax results in Theorem 1 always apply to the highest-skill type. Theorem 1 applies to *all* skill types as long they are not too different from each other.

5.4 First Best

Online Appendix C discusses a version of the baseline FMH model, but with no informational frictions. It shows that the optimal labor income pmf in this alternative environment has a right tail Pareto index given by:

$$\frac{\chi - 1}{\psi} - 1.$$

Since $\alpha < 1$, this right tail index is smaller than that implied by the FMH model. This is an unsurprising but too often overlooked point: better allocations of risk through the tax system (or other vehicles) should be expected to lead to *more* pre-tax income inequality.

6 Conclusion

The paper addresses the question of optimal taxation of high-end earners in an FMH model, which allows individuals to influence their income risk profile. The model implies that in less risk-averse societies, it is optimal for high-end tax rates to be lower and right tail income inequality to be higher. In contrast, because it treats risk as exogenous, the Mirrlees model implies that this kind of negative co-movement between tax rates and inequality (which is a feature of post-1980 US data) is suboptimal.

In the FMH model, agents make costly non-contractible decisions to affect the probabilistic structure of their future labor incomes. This basic mechanism is similar to that in optimal taxation models in which agents can influence their future wages through human capital accumulation.[15] Of course, in the FMH model, agents have considerably more control over their labor income risks. It would seem fruitful in future research to meld the two kinds of paradigms together.

This paper has endogenized income risk in the context of the optimal tax paradigm. It would be useful in future research to explore the consequences of endogenizing individual-level risk in other kinds of models, like in the vast quantitative incomplete markets literature in macroeconomics.[16] To be clear, there is considerable research along these lines in which agents are allowed to choose between different occupations (for example, to be entrepreneurs rather than workers). But this discrete approach is almost certainly too coarse relative to a reality in which similarly placed agents are able to choose any of a multi-dimensional set of career strategies with a correspondingly wide range of risk exposures. The current paper illustrates that incorporating these possibilities into models could have important effects on their positive and normative conclusions.

[15] See Badel and Huggett (2017), Badel, Huggett, and Luo (2020), and Stantcheva (2017), among others.

[16] Kopytov, Taschereau-Dumouchel, and Xu (2024) is an example of the kind of work that I have in mind.

Appendix

The appendix contains the proofs of Propositions 1-4 and Lemmas 1-2.

Proof of Proposition 1

The agent's utility is defined as:

$$U(f;T) = \sum_{i=1}^{\infty}(y_i - T_i)^{\alpha} f_i - \phi \sum_{i=1}^{\infty} f_i^{1+\psi} y_i^{\chi}.$$

The hypothesized summability conditions ensure that $|U(f;T)| < \infty$.

We can then prove the sufficiency of the condition in the Proposition. Suppose that there exists a real κ such that:

$$f_i = (\frac{\max((y_i - T_i)^{\alpha} - \kappa, 0)}{\phi(1+\psi)y_i^{\chi}})^{1/\psi}.$$

for all $i \in \mathbb{N}$. Then I claim that T implements f. Suppose not, and there exists $g \in \mathcal{F}$ such that:

$$\begin{aligned} 0 < \sum_{i=1}^{\infty} c_i^{\alpha} g_i - \sum_{i=1}^{\infty} c_i^{\alpha} f_i \\ -\phi \sum_{i=1}^{\infty} g_i^{1+\psi} y_i^{\chi} + \phi \sum_{i=1}^{\infty} f_i^{1+\psi} y_i^{\chi} \end{aligned}$$

where:

$$c_i \equiv y_i - T_i.$$

Then the subgradient inequality implies that:

$$\begin{aligned} 0 &< \sum_{i=1}^{\infty}(c_i^{\alpha} - \phi(\psi+1) f_i^{\psi} y_i^{\chi})(g_i - f_i) \\ &= \sum_{\{i|f_i>0\}} \kappa(g_i - f_i) + \sum_{\{i|f_i=0\}} c_i^{\alpha} g_i \\ &\leq \sum_{\{i|f_i>0\}} \kappa(g_i - f_i) + \kappa \sum_{\{i|f_i=0\}} g_i \\ &= \kappa - \kappa \sum_{\{i|f_i>0\}} f_i \\ &= \kappa - \kappa \\ &= 0 \end{aligned}$$

because:

$$c_i^\alpha \leq \kappa$$

if $f_i = 0$. This contradiction implies that f is uniquely optimal.

Next, we establish necessity. Suppose:

$$f \in \arg\max_{g \in \mathcal{F}} U(g;T)$$

and f satisfies the summability conditions in the Proposition. Because U is concave in its first argument, f must be unique. Since $|U(f;T)| < \infty$, we can then use perturbation methods to establish that, for some real k, f satisfies the necessary condition in the Proposition. Define:

$$m_i \equiv c_i^\alpha - \phi(1+\psi) f_i^\psi y_i^\chi, i \in \mathbb{N}$$

Suppose there exists i, j such that:

$$m_i > m_j$$
$$f_i, f_j > 0$$

Then consider a class of perturbations $\hat{f}(\delta)$ such that:

$$\hat{f}_i(\delta) = f_i + \delta$$
$$\hat{f}_j(\delta) = f_j - \delta$$
$$\hat{f}_k(\delta) = f_k, k \neq i, j$$

for $\delta \geq 0$. Then, define $\hat{u}$ by:

$$\hat{u}(\delta) = U(\hat{f}(\delta); T) - U(f;T)$$

for δ in a neighborhood around zero. Taking derivatives at $\delta = 0$, we get:

$$\hat{u}'(0) = (y_i - T_i)^\alpha - (y_j - T_j)^\alpha - \phi(1+\psi) f_i^\psi y_i^\chi + \phi(1+\psi) f_j^\psi y_j^\psi.$$
$$= (m_i - m_j) > 0.$$

Hence, if f is optimal, $m_i = m_j$ for all i, j such that $f_i, f_j > 0$.

Now define $\kappa = m_i$ for any $i \in \mathbb{N}$ such that $f_i > 0$. (There exists at least one such i

because f is in $\mathcal{F}$.) Suppose that there is $j \in \mathbb{N}$ such that:

$$f_j = 0$$
$$c_j^\alpha > \kappa.$$

Then consider a class of perturbations $\hat{f}(\delta)$:

$$\hat{f}_i(\delta) = f_i - \delta$$
$$\hat{f}_j(\delta) = f_j + \delta$$
$$\hat{f}_k(\delta) = f_k, k \neq i, j$$

Define:

$$\hat{u}(\delta) = U(\hat{f}(\delta); T) - U(f; T)$$

and take derivatives at $\delta = 0$:

$$\begin{aligned}\hat{u}'(0) &= -(y_i - T_i)^\alpha + (y_j - T_j)^\alpha + \phi(1+\psi) f_i^\psi y_i^\chi - \phi(1+\psi) f_j^\psi y_j^\psi \\ &= -\kappa + c_j^\alpha > 0.\end{aligned}$$

It follows that f is suboptimal (since setting $\delta > 0$ generates more utility).

Proof of Proposition 2

The proof makes use of the Intermediate Value Theorem. Specifically, let $I(\kappa)$ be defined as:

$$I(\kappa) = \sum_{i=1}^{\infty} \left(\frac{\max((y_i - T_i)^\alpha - \kappa, 0)}{\phi(\psi+1) y_i^\chi}\right)^{1/\psi}$$

We show that $I(\kappa)$ is finite for all κ, and is hence continuous as a function of κ. We show too that $I(\kappa)$ is near zero for large values of κ, and near infinity when κ is very low.

Suppose $\kappa \leq 0$. Then:

$$I(\kappa) = \left(\frac{1}{\phi(\psi+1)}\right)^{1/\psi} \sum_{i=1}^{\infty} ((y_i - T_i)^\alpha y_i^{-\chi} - \kappa y_i^{-\chi})^{1/\psi}$$

This is finite for any value of $\kappa \leq 0$. To see this, pick $\varepsilon > 0$ so that:

$$\frac{\chi - \alpha}{\psi} > (1 + \varepsilon).$$

Then:

$$\begin{aligned}
&((y_i - T_i)^\alpha y_i^{-\chi} - \kappa y_i^{-\chi})^{1/\psi} y_i^{1+\varepsilon} \\
\leq\ &((y_i - T_i)^\alpha y_i^{-\chi} y_i^{\psi+\psi\varepsilon} - \kappa y_i^{-\chi+\psi+\psi\varepsilon})^{1/\psi} \\
\leq\ &\max_{i\in\mathbb{N}}(b^\alpha y_i^\alpha y_i^{-\chi} y_i^{\psi+\psi\varepsilon} - \kappa y_i^{-\chi+\psi+\psi\varepsilon})^{1/\psi} \\
=\ &(b^\alpha y_1^\alpha y_1^{-\chi} y_1^{\psi+\psi\varepsilon} - \kappa y_1^{-\chi+\psi+\varepsilon\psi})^{1/\psi}
\end{aligned}$$

since:

$$\begin{aligned}
\alpha - \chi + \psi + \psi\varepsilon < 0 \\
-\chi + \psi + \psi\varepsilon < 0.
\end{aligned}$$

It follows that when $\kappa \leq 0$:

$$0 \leq I(\kappa) \leq (b^\alpha y_1^\alpha y_1^{-\chi} y_1^{\psi+\psi\varepsilon} - \kappa y_1^{-\chi+\psi+\varepsilon\psi})^{1/\psi} \sum_{i=1}^{\infty} y_i^{-1-\varepsilon} < \infty.$$

Now suppose $\kappa > 0$. I claim again that the sum is finite for any such κ. To see this, note that in this case the sum is bounded from above by:

$$(\frac{1}{\phi(\psi+1)})^{1/\psi} \sum_{i=1}^{\infty} (b^\alpha y_i^\alpha y_i^{-\chi})^{1/\psi}.$$

This is finite because $(\alpha - \chi)/\psi < -2 = -1 - 1$.

The integral $I(\kappa)$ is finite for all κ. It is a strictly decreasing function of κ. It approaches infinity for κ near negative infinity and approaches zero for κ sufficiently large. Hence, there is a unique value κ^* where $I(\kappa^*) = 1$.

Finally, we verify that the summability conditions in Proposition 1 are satisfied. Consider a sequence $\{g_i\}_{i=1}^{\infty}$ where:

$$g_i = (\frac{max((y_i - T_i)^\alpha - \kappa^*, 0)}{\phi(\psi+1) y_i^\chi})^{1/\psi} (y_i - T_i). i \in \mathbb{N}$$

Suppose first that $\kappa^* \geq 0$. Then g_i is bounded from above by:

$$\frac{b^{\alpha/\psi+1} y_i^{\alpha/\psi+1}}{(\phi(\psi+1))^{1/\psi} y_i^{\chi/\psi}}.$$

This is summable because:

$$\frac{\alpha - \chi}{\psi} + 1 < -1.$$

Suppose next that $\kappa^* < 0$. Then:

$$\begin{aligned} g_i &= (\frac{(y_i - T_i)^\alpha - \kappa^*}{\phi(\psi+1)y_i^\chi})^{1/\psi}(y_i - T_i) \\ &= (\frac{((y_i - T_i)^{\alpha+\psi} - \kappa^*(y_i - T_i)^\psi}{\phi(\psi+1)y^\chi})^{1/\psi} \\ &\leq (\frac{b^{\alpha+\psi}y_i^{\alpha+\psi} - \kappa^* b^\psi y_i^\psi}{\phi(\psi+1)y^\chi})^{1/\psi}. \end{aligned}$$

This is summable because:

$$\frac{(\alpha + \psi - \chi)}{\psi} < -1.$$

Now consider the sequence $g' = \{g'_i\}_{i=1}^\infty$ where:

$$g'_i = (\frac{max((y_i - T_i)^\alpha - \kappa^*, 0)}{\phi(\psi+1)y_i^\chi})^{(1+\psi)/\psi} y_i^\chi, i \in \mathbb{N}$$

If $\kappa^* \geq 0$, then g' is bounded from above by:

$$\frac{b^{\alpha(1+1/\psi)} y_i^{\alpha+\alpha/\psi-\chi/\psi-\chi} y_i^\chi}{\phi(\psi+1)^{(1+1/\psi)}}$$

which is summable because:

$$\alpha + \alpha/\psi - \chi/\psi < \alpha - 2 < -1.$$

If $\kappa^* < 0$, then g'_i is bounded from above by:

$$\bar{g}_i = (\frac{max(b^\alpha y_i^\alpha y_i^{\frac{\chi\psi}{1+\psi}-\chi} - \kappa^* y_i^{-\chi}, 0)}{\phi(\psi+1)})^{(1+\psi)/\psi}, i \in \mathbb{N}.$$

The sequence $\bar{g}'$ is summable because:

$$\begin{aligned} &\alpha(1+\psi)/\psi - \chi/\psi \\ &= \frac{\alpha - \chi}{\psi} + \alpha \\ &< -2 + \alpha \\ &< -1. \end{aligned}$$

Proof of Proposition 3

The first part of the proposition follows from Proposition 2. Consider:

$$b^* = \sup_{i \in \mathbb{N}} (1 - T_i/y_i).$$

Suppose $b^* = \infty$. Then:

$$\lim_{i \to \infty} (1 - T_i/y_i) = \infty.$$

But this implies that:

$$\lim_{i \to \infty} (\frac{(y_i - T_i)y_i^\eta}{y_i^{\eta+1}}) = \infty,$$

which contradictions the assumption about T. Hence, there exists $b^* < \infty$ such that $b^* \geq (1 - T_i/y_i)$ for all i in $\mathbb{N}$.

To prove the remainder of the proposition, define $c_i = (y_i - T_i)$. We first show that the posited tail behavior is a valid characterization for any κ. Define:

$$L = lim_{i \to \infty} c_i y_i^{-\eta}.$$

Then consider the limit:

$$\begin{aligned} & lim_{i \to \infty} y_i^{\frac{\chi - \alpha\eta}{\psi}} \hat{f}_i(\kappa) \\ = & lim_{i \to \infty} (\frac{max(y_i^{\chi - \alpha\eta} c_i^\alpha - \kappa y_i^{\chi - \alpha\eta}, 0)}{\theta(1+\psi) y^\chi})^{1/\psi} \\ = & lim_{i \to \infty} (\frac{max(y_i^{-\alpha\eta} c_i^\alpha - \kappa y_i^{-\alpha\eta}, 0)}{\theta(1+\psi)})^{1/\psi} \\ = & (\frac{L^\alpha}{\theta(1+\psi)})^{1/\psi}. \end{aligned}$$

It follows that:

$$\hat{f}(\kappa) \sim_\infty y^{\frac{\alpha\eta - \chi}{\psi}}.$$

Proof of Lemma 1

Let (f, κ) be in the constraint set to $P^*(M, \bar{\theta})$. Define T by:

$$T_i = y_i - (\phi(1+\psi) f_i^\psi y_i^\chi + \kappa)^{1/\alpha}, i \in \mathbb{N}.$$

Since:

$$\kappa \geq -\inf_{i\in\mathbb{N}} \phi(1+\psi)f_i^{\psi}y_i^{\chi},$$

the tax schedule satisfies the non-negativity requirement that, for all $i \in \mathbb{N}$, $(y_i - T_i) \geq 0$. As well, given this definition of (T, f), the resource constraint in $P^*(M, \bar{\theta})$ implies that (T, f) satisfies the resource constraint in $P(M, \bar{\theta})$.

The main issue, then, is to verify that T implements f. Since f satisfies the sufficient conditions in Proposition 1, we need only verify that it satisfies the summability conditions in Proposition 1. In terms of the first summability condition:

$$\begin{aligned}
&\sum_{i=1}^{\infty}(y_i - T_i)^{\alpha} f_i \\
&= \sum_{i=1}^{\infty}(\kappa + \phi(1+\psi)f_i^{\psi}y_i^{\chi})f_i \\
&= \sum_{i=1}^{\infty}((\phi(1+\psi)f_i^{\psi}y_i^{\chi} + \kappa)^{1/\alpha})^{\alpha} f_i \\
&\leq (\sum_{i=1}^{\infty}(\phi(1+\psi)f_i^{\psi}y_i^{\chi} + \kappa)^{1/\alpha} f_i)^{\alpha} \text{ by Jensen's inequality} \\
&\leq (\sum_{i=1}^{\infty} f_i y_i)^{\alpha} \text{ by the resource constraint} \\
&< \infty
\end{aligned}$$

where the last step follows from the upper bound on f.

In terms of the second summability condition:

$$\begin{aligned}
&\phi(1+\psi)\sum_{i=1}^{\infty} f_i^{1+\psi}y_i^{\chi} \\
&= \sum_{i=1}^{\infty}(\phi(1+\psi)f_i^{\psi}y_i^{\chi} + \kappa)f_i - \kappa \\
&\leq \sum_{i=1}^{\infty} y_i^{\alpha} f_i - \kappa \\
&\leq (\sum_{i=1}^{\infty} y_i f_i)^{\alpha} - \kappa
\end{aligned}$$

and this sum is finite because of the upper bound on f.

It is straightforward to verify that the objectives are the same:

$$U(f;T) = \sum_{i=1}^{\infty}(\phi(1+\psi)f_i^{\psi}y_i^{\chi} + \kappa)f_i - \phi\sum_{i=1}^{\infty}f_i^{1+\psi}y_i^{\chi}$$
$$= \phi\psi\sum_{i=1}^{\infty}f_i^{1+\psi}y_i^{\chi} + \kappa.$$

Proof of Lemma 2

Let (T, f) be in the constraint set to $P(M, \bar{\theta})$, and define $c_i = y_i - T_i, i \in \mathbb{N}$. T implements f and so f solves the problem

$$f \in \arg\max_{g\in\mathcal{F}} \sum_{i=1}^{\infty} c_i^{\alpha} g_i - \phi\sum_{i=1}^{\infty} g_i^{\psi+1}y_i^{\chi}.$$

We first verify the summability conditions in Proposition 1, and then apply the necessary conditions in that Proposition.

The pmf f satisfies the resource constraint in $P(M, \bar{\theta})$:

$$\sum_{i=1}^{\infty} T_i f_i \geq 0$$
$$\Rightarrow \sum_{i=1}^{\infty} y_i f_i \geq \sum_{i=1}^{\infty} f_i c_i$$

The left hand side is finite because f satisfies the upper bound constraint. Hence, by Jensen's inequality, we can conclude that:

$$\sum_{i=1}^{\infty} c_i^{\alpha} f_i < \infty$$

which verifies the first summability condition in Proposition 1.

In terms of the second summability condition, we know that f is a solution to the agent's problem. The agent could choose $g_1 = 1$ and $g_i = 0, i > 1$. This would result in a value for the objective that is larger than negative infinity. It follows that at the optimal pmf f:

$$\sum_{i=1}^{\infty} f_i^{1+\psi} y_i < \infty.$$

Since T implements f, and f satisfies the summability conditions in Proposition 1, f must satisfy the necessary conditions in Proposition 1. It follows that there exists κ such that:

$$f_i = (\frac{max(c_i^\alpha - \kappa, 0)}{y_i^\chi \phi(1+\psi)})^{1/\psi}, i \in \mathbb{N} \tag{21}$$

which implies that if $f_i > 0:$

$$\kappa = c_i^\alpha - \phi(1+\psi) f_i^\psi y_i^\chi.$$

Note that:

$$f_i^\psi y_i^\chi \phi(1+\psi) + \kappa \geq c_i^\alpha, i \in \mathbb{N}$$

and so:

$$\inf_{i \in \mathbb{N}} f_i^\psi y_i^\chi \phi(1+\psi) \geq -\kappa.$$

As well, since (T, f) satisfies the resource constraint in $P(M, \bar{\theta}):$

$$\begin{aligned} \sum_{i=1}^{\infty} T_i f_i &\geq 0 \\ \Rightarrow \sum_{\{i|f_i>0\}} (y_i - (f_i^\psi y_i^\chi \phi(1+\psi) + \kappa)^{1/\alpha})) f_i &\geq 0 \\ = \sum_{i=1}^{\infty} (y_i - (f_i^\psi y_i^\chi \phi(1+\psi) + \kappa)^{1/\alpha})) f_i &\geq 0 \end{aligned}$$

and so (f, κ) satisfy the resource constraint in $P^*(M, \bar{\theta})$.

We have shown that (f, κ) is in $P^*(M, \bar{\theta})$. We can then prove the lemma by checking that (f, κ) has the same value for the objective as $(T, f):$

$$\begin{aligned} \phi\psi \sum_{i=1}^{\infty} f_i^{1+\psi} y_i^\chi + \kappa. \\ = \sum_{i=1}^{\infty} (\phi(1+\psi) f_i^\psi y_i^\chi f_i + \kappa f_i) - \phi \sum_{i=1}^{\infty} f_i^\psi y_i^\chi f_i \\ = \sum_{i=1}^{\infty} ((y_i - T_i)^\alpha) f_i - \phi \sum_{i=1}^{\infty} f_i^\psi y_i^\chi f_i \\ = U(f; T). \end{aligned}$$

Proof of Theorem 1

The proof of the Theorem is divided into four parts:

1. In a solution to $P^*(M, \bar{\theta})$, the upper bound does not bind for large values of y
2. In a solution to $P^*(M, \bar{\theta})$, the lower bound does not bind for large values of y

3. In a solution to $P^*(M, \bar{\theta})$, optimal consumption is asymptotically linear as a function of y

4. In a solution to $P(M, \bar{\theta})$, the optimal tax schedule is asymptotically linear as a function of y.

Upper Bound Does Not Bind For Large y

We first prove that, given a solution to $P^*(M, \bar{\theta})$, the upper bound does not bind for large values of y. Consider a solution (f, κ) to problem $P^*(M, \bar{\theta})$. and define:

$$c_i = (\phi(1+\psi) f_i^{\psi} y_i^{\chi} + \kappa)^{1/\alpha}, i \in \mathbb{N}.$$

It is helpful to keep in mind that:

$$\begin{aligned}
\sum_{i=1}^{\infty} y_i f_i &< \infty \text{ (f satisfies the upper bound to } P^*(M, \bar{\theta})) \\
\sum_{i=1}^{\infty} c_i f_i &\leq \sum_{i=1}^{\infty} y_i f_i < \infty \text{ (resource constraint)} \\
\phi(1+\psi) \sum_{i=1}^{\infty} f_i^{1+\psi} y_i^{\chi} + \kappa = \sum_{i=1}^{\infty} c_i^{\alpha} f_i &\leq (\sum_{i=1}^{\infty} f_i c_i)^{\alpha} < \infty \text{ (Jensen's inequality)} \\
\sum_{i=1}^{\infty} c_i^{1-\alpha} f_i &< \infty \text{ (also Jensen's inequality).} \qquad (22)
\end{aligned}$$

Suppose that there is a monotone sequence $(i_k)_{k=1}^{\infty}$ in $\mathbb{N}$ that converges to infinity and such that:

$$f_{i_k} y_{i_k}^{\bar{\theta}} = M, k \in \mathbb{N}$$

We will see that this supposition leads to a contradiction. Note that:

$$c_{i_k} = (\phi(1+\psi) M^{\psi} y_{i_k}^{\chi - \bar{\theta}\psi} + \kappa)^{1/\alpha}, k \in \mathbb{N}$$

By assumption, M is sufficiently large that:

$$1 < \sum_{i=1}^{\infty} M y_i^{-\bar{\theta}}.$$

Since $f \in \mathcal{F}$, there exists some i^* such that:

$$f_{i^*} < M y_{i^*}^{-\bar{\theta}}.$$

For any j, consider the perturbation:

$$\hat{f}_{i_j} = f_{i_j} - \delta$$
$$\hat{f}_{i^*} = f_{i^*} + \delta$$

where δ is positive and near zero (so that $\hat{f}$ is in $\mathcal{F}$). Then, consider the optimization problem:

$$\begin{aligned}
&\max_{\delta,\Delta\geq 0} \phi\psi(f_{i^*}+\delta)^{1+\psi}y_{i^*}^{\chi} + \phi\psi(f_{i_j}-\delta)^{1+\psi}y_{i_j}^{\chi} + \Delta \\
&s.t.\ (y_{i^*} - (\phi(1+\psi)(f_{i^*}+\delta)^{\psi}y_{i^*}^{\chi} + \kappa + \Delta)^{1/\alpha})(f_{i^*}+\delta) \\
&\quad +(y_{i_j} - (\phi(1+\psi)(f_{i_j}-\delta)^{\psi}y_{i_j}^{\chi} + \kappa + \Delta)^{1/\alpha})(f_{i_j}-\delta) \\
&\quad + \sum_{s\in\mathbb{N}-\{i^*,i_j\}} (y_s - (\phi(1+\psi)f_s^{\psi}y_s^{\psi} + \kappa + \Delta)^{1/\alpha})f_s \geq 0
\end{aligned}$$

Since (f,κ) is optimal, then $(0,0)$ should be a solution to this problem. Given j, the first-order necessary condition to this problem with respect to δ is:

$$\begin{aligned}
&(1+\psi)\phi\psi(f_{i^*}^{\psi}y_{i^*}^{\chi} - f_{i_j}^{\psi}y_{i_j}^{\chi}) + \lambda^j(y_{i^*} - c_{i^*} - y_{i_j} + c_{i_j}) \\
&-\lambda^j\alpha^{-1}\phi\psi(1+\psi)f_{i^*}^{\psi}y_{i^*}^{\chi}c_{i^*}^{1-\alpha} + \lambda^j\alpha^{-1}\phi\psi(1+\psi)f_{i_j}^{\psi}y_{i_j}^{\chi}c_{i_j}^{1-\alpha} \leq 0
\end{aligned} \tag{23}$$

and the first-order necessary condition with respect to Δ is:

$$1 \leq \lambda^j\alpha^{-1}\sum_{k=1}^{\infty} c_k^{1-\alpha}f_k$$

where λ^j is the positive multiplier on the constraint.

We can show that the term in (23) being multiplied by λ^j converges to infinity as j grows large:

$$\begin{aligned}
&\lim_{j\to\infty}(-y_{i_j} + c_{i_j} + \alpha^{-1}\phi\psi(1+\psi)f_{i_j}^{\psi}y_{i_j}^{\chi}c_{i_j}^{1-\alpha}) \\
&= \lim_{j\to\infty}(-y_{i_j} + (\phi(1+\psi)f_{i_j}^{\psi}y_{i_j}^{\chi} + \kappa)^{1/\alpha} + \alpha^{-1}\phi\psi(1+\psi)f_{i_j}^{\psi}y_{i_j}^{\chi}(\phi(1+\psi)f_{i_j}^{\psi}y_{i_j}^{\chi} + \kappa)^{1/\alpha-1}) \\
&= \lim_{j\to\infty} y_{i_j}^{\frac{\chi-\psi\bar{\theta}}{\alpha}}(-y_{i_j}^{1-\frac{\chi-\psi\bar{\theta}}{\alpha}} + (\phi(1+\psi))^{1/\alpha}M^{\psi/\alpha} + \alpha^{-1}\psi((\phi(1+\psi))^{1/\alpha}M^{\psi/\alpha}) \\
&= M^{\psi/\alpha}((\phi(1+\psi))^{1/\alpha} + \alpha^{-1}\psi((\phi(1+\psi))^{1/\alpha})\lim_{j\to\infty} y_{i_j}^{\frac{\chi-\psi\bar{\theta}}{\alpha}} \\
&= \infty
\end{aligned}$$

since $(\frac{\chi-\psi\bar{\theta}}{\alpha}) > 1$. Hence, there exists j^* such that for all $j \geq j^*$:

$$
\begin{aligned}
0 \geq &(1+\psi)\phi\psi(f_{i^*}^{\psi}y_{i^*}^{\chi} - f_{i_j}^{\psi}y_{i_j}^{\chi}) \\
&+ (\frac{1}{\alpha^{-1}\sum_{k=1}^{\infty} c_k f_k})(y_{i^*} - c_{i^*} - y_{i_j} + c_{i_j} - \alpha^{-1}\phi\psi(1+\psi)f_{i^*}^{\psi}y_{i^*}^{\chi}c_{i^*}^{1-\alpha} + \alpha^{-1}\phi\psi(1+\psi)f_{i_j}^{\psi}y_{i_j}^{\chi}c_{i_j}^{1-\alpha}).
\end{aligned}
$$

Taking limits of the right hand side with respect to j, we get:

$$
\begin{aligned}
0 \geq\ &(1+\psi)\phi\psi f_{i^*}^{\psi}y_{i^*}^{\chi} + (\frac{1}{\alpha^{-1}\sum_{k=1}^{\infty} c_k f_k})(y_{i^*} - c_{i^*} - \alpha^{-1}\phi\psi(1+\psi)f_{i^*}^{\psi}y_{i^*}^{\chi}c_{i^*}^{1-\alpha}) \\
&+ \lim_{j\to\infty} y_{i_j}^{\frac{\chi-\psi\bar{\theta}}{\alpha}}(-(1+\psi)\phi\psi y_{i_j}^{(\chi-\bar{\theta}\psi)(1-1/\alpha)}M^{\psi}) \\
&+ \lim_{j\to\infty} y_{i_j}^{\frac{\chi-\psi\bar{\theta}}{\alpha}}(\frac{1}{\alpha^{-1}\sum_{k=1}^{\infty} c_k f_k})(-y_{i_j}^{1-\frac{\chi-\psi\bar{\theta}}{\alpha}} + (\phi(1+\psi)M^{\psi})^{1/\alpha} + \alpha^{-1}\psi((\phi(1+\psi)M^{\psi})^{1/\alpha}) \\
&= (1+\psi)\phi\psi f_{i^*}^{\psi}y_{i^*}^{\chi} + (\frac{1}{\alpha^{-1}\sum_{k=1}^{\infty} c_k f_k})(y_{i^*} - c_{i^*} - \alpha^{-1}\phi\psi(1+\psi)f_{i^*}^{\psi}y_{i^*}^{\chi}c_{i^*}^{1-\alpha}) \\
&+ \lim_{j\to\infty} y_{i_j}^{\frac{\chi-\psi\bar{\theta}}{\alpha}}(\frac{1}{\alpha^{-1}\sum_{k=1}^{\infty} c_k f_k})(\phi(1+\psi))^{1/\alpha} + \alpha^{-1}\psi((\phi(1+\psi))^{1/\alpha})M^{\psi/\alpha} \\
&= \infty.
\end{aligned}
$$

This is a contradiction and so there exists j^* such that $f_j < My_j^{-\bar{\theta}}$ for all $j \geq j^*$.

Lower Bound Does Not Bind For Large y

Next, we show that given a solution (f,κ) to $P^*(M,\bar{\theta})$, there exists i^* such that:

$$f_i > 0$$

for all $i \geq i^*$. Suppose not, and so there is a monotone sequence $\{i_j\}_{j=1}^{\infty}$ in $\mathbb{N}$ such that:

$$
\begin{aligned}
f_{i_j} &= 0, j = 1,2,3..... \\
\lim_{j\to\infty} i_j &= \infty
\end{aligned}
$$

There exists k such that $f_k > 0$. Then, for arbitrary j, consider the optimization problem:

$$\max_{\delta\geq 0,\Delta\geq 0} \phi\psi(f_k-\delta)^{1+\psi}y_k^{\chi}+\phi\psi\delta^{1+\psi}y_{i_j}^{\chi}+\Delta$$
$$s.t.\ (y_k-(\phi(1+\psi)(f_k-\delta)^{\psi}y_k^{\chi}+\kappa+\Delta)^{1/\alpha})(f_k-\delta)$$
$$+(y_{i_j}-(\phi(1+\psi)\delta^{\psi}y_{i_j}^{\chi}+\kappa+\Delta)^{1/\alpha})\delta$$
$$+\sum_{s\in\mathbb{N}-\{k,i_j\}}(y_s-(\phi(1+\psi)f_s^{\psi}y_s^{\psi}+\kappa+\Delta)^{1/\alpha})f_s\geq 0$$

If (f,κ) is optimal, then $(\Delta,\delta)=(0,0)$ solves the above problem. It satisfies the first-order necessary conditions:.

$$0\geq -\phi\psi(1+\psi)f_k^{\psi}y_k^{\chi}+\lambda^j(-y_k+c_k)+\lambda^j\alpha^{-1}\phi\psi(1+\psi)f_k^{\psi}y_k^{\chi}c_k^{1-\alpha}+\lambda^j y_{i_j}$$
$$\geq -\phi\psi(1+\psi)f_k^{\psi}y_k^{\chi}+\lambda^j(-y_k+c_k)+\lambda^j\alpha^{-1}\phi\psi(1+\psi)f_k^{\psi}y_k^{\chi}c_k^{1-\alpha}+\frac{1}{\alpha^{-1}\sum_{s=1}^{\infty}c_s^{1-\alpha}f_s}y_{i_j}.$$

where the inequality is a consequence of the first-order necessary condition with respect to Δ. But this inequality cannot be satisfied for large values of j, as y_{i_j} converges to infinity.

Consumption Asymptotes to a Linear Function of Income

The above two parts of the proof allow us to conclude that there exists i^{**} such that:

$$0<f_i<My_i^{-\bar{\theta}}$$

for all $i>i^{**}$. Let $i,j>i^{**}$ and consider the maximization problem:

$$\max_{\Delta\geq 0,\delta\in\mathbb{R}} \phi\psi(f_i+\delta)^{1+\psi}y_i^{\chi}+\phi\psi(f_j-\delta)^{1+\psi}y_j^{\chi}$$
$$s.t.\ (y_i-(\phi(1+\psi)(f_i+\delta)^{\psi}y_i^{\chi}+\kappa+\Delta)^{1/\alpha})(f_i+\delta)$$
$$+(y_j-(\phi(1+\psi)(f_j-\delta)^{\psi}y_j^{\chi}+\kappa+\Delta)^{1/\alpha})(f_j-\delta)$$
$$+\sum_{s\in\mathbb{N}-\{i,j\}}(y_s-(\phi(1+\psi)f_s^{\psi}y_s^{\psi}+\kappa+\Delta)^{1/\alpha})f_s\geq 0$$

As we saw earlier, the first-order necessary conditions to this problem with respect to δ are:

$$\begin{aligned}(1+\psi)\phi\psi(f_i^{\psi}y_i^{\chi}-f_j^{\psi}y_j^{\chi})+\lambda^{ij}(y_i-c_i-y_j+c_j)\\ -\lambda^{ij}\alpha^{-1}\phi\psi(1+\psi)f_i^{\psi}y_i^{\chi}c_i^{1-\alpha}+\lambda^{ij}\alpha^{-1}\phi\psi(1+\psi)f_j^{\psi}y_j^{\chi}c_j^{1-\alpha}=0\end{aligned} \tag{24}$$

and with respect to Δ:

$$1\leq\lambda^{ij}\alpha^{-1}\sum_{s=1}^{\infty}c_s^{1-\alpha}f_s$$

where λ^{ij} is the multiplier on the resource constraint.

Our goal in the remainder of this subsection is to show that:

$$lim_{j\to\infty}c_j/y_j=(1+\psi/\alpha)^{-1}. \tag{25}$$

We first obtain a contradiction when the LHS of (25) is less than the RHS of (25) along some subsequence, and then obtain a contradiction when the RHS is less than the LHS along some subsequence.

Suppose that there exists a monotone sequence $\{i_j\}_{j=1}^{\infty}$, $i_j\geq i^{**}$, such that it converges to infinity and:

$$\lim_{j\to\infty}c_{i_j}/y_{i_j}=r^*<(1+\psi/\alpha)^{-1}$$

The term in (24) being multiplied by λ^{ij} converges to negative infinity along this subsequence:

$$\begin{aligned}&\lim_{k\to\infty}(-y_{i_k}+c_{i_k}+\alpha^{-1}\phi\psi(1+\psi)f_{i_k}^{\psi}y_{i_k}^{\chi}c_{i_k}^{1-\alpha})\\ &=\lim_{k\to\infty}y_{i_k}(-1+c_{i_k}/y_{i_k}+\alpha^{-1}\phi\psi(1+\psi)c_{i_k}/y_{i_k})\\ &=\lim_{k\to\infty}y_{i_k}(-1+r^*(1+\alpha^{-1}\psi))\\ &=-\infty.\end{aligned}$$

and $\lambda^{ij}\geq\frac{1}{\alpha^{-1}\sum_{s=1}^{\infty}c_s^{1-\alpha}f_s}$. Hence, for k sufficiently large

$$\begin{aligned}0&=(1+\psi)\phi\psi(f_i^{\psi}y_i^{\chi}-f_{i_k}^{\psi}y_{i_k}^{\chi})+\lambda^{ij}(y_i-c_i-y_{i_k}+c_{i_k})\\ &-\lambda^{ij}\alpha^{-1}\phi\psi(1+\psi)f_i^{\psi}y_i^{\chi}c_i^{1-\alpha}+\lambda^{ij}\alpha^{-1}\phi\psi(1+\psi)f_{i_k}^{\psi}y_{i_k}^{\chi}c_{i_k}^{1-\alpha}\\ &\leq(1+\psi)\phi\psi(f_i^{\psi}y_i^{\chi}-f_{i_k}^{\psi}y_{i_k}^{\chi})\\ &+\frac{1}{\alpha^{-1}\sum_{s=1}^{\infty}c_s^{1-\alpha}f_s}((y_i-c_i-y_{i_k}+c_{i_k})-\alpha^{-1}\phi\psi(1+\psi)f_i^{\psi}y_i^{\chi}c_i^{1-\alpha}+\alpha^{-1}\phi\psi(1+\psi)f_{i_k}^{\psi}y_{i_k}^{\chi}c_{i_k}^{1-\alpha})\end{aligned} \tag{26}$$

Recall that:

$$c_{i_k}^{\alpha} - \kappa = \phi(1+\psi) f_{i_k}^{\psi} y_{i_k}^{\chi}$$

and so:

$$\lim_{k\to\infty} \frac{\phi(1+\psi) f_{i_k}^{\psi} y_{i_k}^{\chi} c_{i_k}^{1-\alpha}}{y_{i_k}} = \lim_{k\to\infty} \frac{c_{i_k}}{y_{i_k}} = r^*$$
$$\lim_{k\to\infty} \frac{\phi(1+\psi) f_{i_k}^{\psi} y_{i_k}^{\chi}}{y_{i_k}} = \lim_{k\to\infty} (r^*)^{\alpha} / y_{i_k}^{1-\alpha} = 0$$

Hence, if we divide the RHS of (26) by y_{i_k} and take the limit with respect to k converging to infinity, we get:

$$0 \le \frac{1}{\alpha^{-1} \sum_{s=1}^{\infty} c_s^{1-\alpha} f_s} (-1 + r^*(1+\psi/\alpha)) < 0$$

which is a contradiction.

Suppose instead that there exists a monotone sequence $\{i_j\}_{j=1}^{\infty}, i_j \ge i^{**}$, such that it converges to infinity and:

$$\lim_{j\to\infty} y_{i_j}/c_{i_j} = r^{**} < (1+\psi/\alpha)$$

Note that this implies that $lim_{j\to\infty} c_{i_j} = \infty$ and so $lim_{j\to\infty} c_{i_j}^{\alpha-1} = 0$.

The term in (24) being multiplied by λ^{ij} converges to infinity along this subsequence:

$$\begin{aligned}
&\lim_{k\to\infty} (-y_{i_k} + c_{i_k} + \alpha^{-1}\phi\psi(1+\psi) f_{i_k}^{\psi} y_{i_k}^{\chi} c_{i_k}^{1-\alpha}) \\
&= \lim_{k\to\infty} c_{i_k}(-y_{i_k}/c_{i_k} + 1 + \alpha^{-1}\phi\psi(1+\psi)) \\
&= \lim_{k\to\infty} c_{i_k}(-r^{**} + (1+\alpha^{-1}\psi)) \\
&= \infty.
\end{aligned}$$

Hence, for k sufficiently large:

$$\begin{aligned}
0 &= (1+\psi)\phi\psi(f_i^{\psi} y_i^{\chi} - f_{i_k}^{\psi} y_{i_k}^{\chi}) + \lambda^{ij}(y_i - c_i - y_{i_k} + c_{i_k}) \\
&- \lambda^{ij}\alpha^{-1}\phi\psi(1+\psi) f_i^{\psi} y_i^{\chi} c_i^{1-\alpha} + \lambda^{ij}\alpha^{-1}\phi\psi(1+\psi) f_{i_k}^{\psi} y_{i_k}^{\chi} c_{i_k}^{1-\alpha} \\
&\ge (1+\psi)\phi\psi(f_i^{\psi} y_i^{\chi} - f_{i_k}^{\psi} y_{i_k}^{\chi}) \\
&+ \frac{1}{\alpha^{-1}\sum_{s=1}^{\infty} c_s^{1-\alpha} f_s}((y_i - c_i - y_{i_k} + c_{i_k}) - \alpha^{-1}\phi\psi(1+\psi) f_i^{\psi} y_i^{\chi} c_i^{1-\alpha} + \alpha^{-1}\phi\psi(1+\psi) f_{i_k}^{\psi} y_{i_k}^{\chi} c_{i_k}^{1-\alpha})
\end{aligned}$$

But if we divide the RHS by c_{i_k} and take the limit with respect to k converging to infinity, we get:

$$0 \geq -\lim_{k\to\infty} c_{i_k}^{\alpha}/c_{i_k} + \frac{1}{\alpha^{-1}\sum_{s=1}^{\infty} c_s^{1-\alpha} f_s}(-r^{**} + (1+\psi/\alpha))$$
$$= \frac{1}{\alpha^{-1}\sum_{s=1}^{\infty} c_s^{1-\alpha} f_s}(-r^{**} + (1+\psi/\alpha)) > 0$$

which is a contradiction.

It follows that for every monotone sequence $\{i_j\}_{j=1}^{\infty}, i_j \geq i^{**}$, such that it converges to infinity:

$$\lim_{j\to\infty} y_{i_j}/c_{i_j} \geq (1+\psi/\alpha)$$
$$\lim_{j\to\infty} c_{i_j}/y_{i_j} \geq (1+\psi/\alpha)^{-1}.$$

Hence:

$$lim_{i\to\infty} c_i/y_i = (1+\psi/\alpha)^{-1}.$$

Asymptotic Properties of Optimal Taxes

We now use the above characterization of solutions to $P^*(M,\bar{\theta})$ to consider the properties of solutions to $P(M,\bar{\theta})$. So, let (T^*, f^*) be a solution to $P(M,\bar{\theta})$. Then, Lemmas 1-2 imply that (f,κ) is a solution to $P^*(M,\bar{\theta})$, where:

$$f_i = (\frac{\max((y_i - T_i)^{\alpha} - \kappa, 0)}{y_i^{\chi}\phi(1+\psi)})^{1/\psi}, i \in \mathbb{N}$$

Given this specification of f_i, we can (as above) define for all $i \in \mathbb{N}$:

$$c_i = (\phi(1+\psi) f_i^{\psi} y_i^{\chi} + \kappa)^{1/\alpha}$$
$$= (y_i - T_i) \geq 0$$

We have shown that:

$$\lim_{i\to\infty} c_i/y_i = (1+\psi/\alpha)^{-1}.$$

Hence:

$$\lim_{i\to\infty} T_i/y_i = 1 - (1+\psi/\alpha)^{-1}$$
$$= \frac{1}{\alpha/\psi + 1}.$$

As well:

$$\lim_{i\to\infty} f_i y_i^{\frac{\chi-\alpha}{\psi}}$$
$$= \lim_{i\to\infty}((f_i^{\psi} y_i^{\chi})^{1/\alpha} y_i^{-1})^{\alpha/\psi}$$
$$= \text{const} * \lim_{i\to\infty}(c_i/y_i)^{\alpha/\psi}$$

which proves the theorem:

$$f \sim_\infty y^{\frac{\alpha-\chi}{\psi}}$$

Proof of Proposition 4

We rewrite problem $P^*(M,\theta)$ as:

$$\max_{\kappa\in\mathbb{R}, f\in\mathbb{R}^\infty} \phi\psi \sum_{i=1}^{\infty} f_i^{1+\psi} y_i^{\chi} + \kappa$$

$$s.t. \sum_{i=1}^{\infty} y_i f_i \geq \sum_{i=1}^{\infty} (\phi(1+\psi) f_i^{\psi} y_i^{\chi} + \kappa)^{1/\alpha} f_i \quad (27)$$

$$M y_i^{-\bar{\theta}} \geq f_i \geq 0, i \in \mathbb{N} \quad (28)$$

$$\kappa \geq -\inf_{i\in\mathbb{N}} \phi(1+\psi) f_i^{\psi} y_i^{\chi} \quad (29)$$

$$\sum_{i=1}^{\infty} f_i = 1 \quad (30)$$

We first establish that the constraint set of the problem is compact in the product topology over $\mathbb{R}^\infty$. Consider the set $\hat{\mathcal{L}}$ of all elements of $\mathbb{R}^\infty$ such that:

$$0 \leq f_i \leq M y_i^{-\bar{\theta}} \text{ for all } i \in \mathbb{N}.$$

By Tychonoff's Theorem, the set $\hat{\mathcal{L}}$ is compact in the product topology.

We next prove that the constraint set is a closed (in the product topology) subset of $\hat{\mathcal{L}}$.

Consider a sequence $(f^n, \kappa_n)_{n=1}^{\infty}$ in $\hat{\mathcal{L}} \times \mathbb{R}$ such that:

$$\begin{aligned}
f^n &\in \hat{\mathcal{L}}, n = 1, 2, 3, ... \\
-\kappa^n &\leq \inf_{i \in \mathbb{N}} \phi(1+\psi) y_i^{\chi} (f_i^n)^{\psi}, n = 1, 2, 3., ... \\
\sum_{i=1}^{\infty} y_i f_i^n &\geq \sum_{i=1}^{\infty} (\phi(1+\psi) f_i^n y_i^{\chi} + \kappa^n)^{1/\alpha} f_i^n, n = 1, 2, 3, .. \\
\sum_{i=1}^{\infty} f_i^n &= 1, n = 1, 2, 3, ... \\
\lim_{n \to \infty} f_i^n &= f_i^* \text{ for all } i \in \mathbb{N} \\
\lim_{n \to \infty} \kappa_n &= \kappa^*.
\end{aligned}$$

Note that κ_n is bounded from above and below:

$$\begin{aligned}
M \sum_{i=1}^{\infty} y_i^{1-\bar{\theta}} &\geq \sum_{i=1}^{\infty} y_i f_i^n \geq \kappa_n^{1/\alpha} \\
-\kappa_n &\leq \inf_{i \in \mathbb{N}} \phi(1+\psi) y_i^{\chi} M^{\psi} y_i^{-\bar{\theta}\psi} = \phi(1+\psi) M^{\psi} y_1^{\chi - \bar{\theta}\psi}
\end{aligned}$$

where the last equality follows from the hypothesis that $(\chi - \bar{\theta}\psi) > \alpha > 0$. Hence:

$$-\phi(1+\psi) M^{\psi} y_1^{\chi - \bar{\theta}\psi} \equiv \kappa_{LB} \leq \kappa^* \leq \kappa_{UB} \equiv M^{1/\alpha} (\sum_{i=1}^{\infty} y_i^{1-\bar{\theta}})^{1/\alpha}.$$

Define a sequence $(h_n)_{n=1}^{\infty}$ in $\mathbb{R}^{\infty}$ by:

$$h_i^n = (\phi(1+\psi)(f_i^n)^{\psi} y^{\chi} + \kappa^n)^{1/\alpha} f_i^n, i \in \mathbb{N}.$$

Then the upper bound on f implies that:

$$h_i^n \leq \bar{h}_i \equiv (\phi(1+\psi) M^{\psi} y_i^{-\psi\bar{\theta}} y_i^{\chi} + \kappa_{UB})^{1/\alpha} M y_i^{-\bar{\theta}}, i \in \mathbb{N}$$

We can verify that $\bar{h}$ is a summable sequence as follows. The proposition hypothesizes that $(\bar{\theta} - 2)\alpha + \bar{\theta}\psi > (\chi - \alpha)$ or equivalently that $(\bar{\theta} - 1)\alpha - \chi + \psi\bar{\theta} > 0$. We can choose ε so that $\frac{\alpha\bar{\theta} - \alpha - \chi + \psi\bar{\theta}}{\alpha} > \varepsilon > 0$. Then:

$$\begin{aligned}
lim_{i \to \infty} \bar{h}_i y_i^{1+\varepsilon} &= lim_{i \to \infty} (\phi(1+\psi) M^{\psi} y_i^{-\psi\bar{\theta}} y_i^{\chi} + \kappa_{UB})^{1/\alpha} M y_i^{1-\bar{\theta}+\varepsilon} \\
&= lim_{i \to \infty} (\phi(1+\psi) M^{\psi+\alpha} y_i^{\chi - \psi\bar{\theta} + \alpha - \alpha\bar{\theta} + \alpha\varepsilon} + \kappa_{UB} M^{\alpha} y^{\alpha - \alpha\bar{\theta} + \alpha\varepsilon})^{1/\alpha} = 0
\end{aligned}$$

which proves $\bar{h}$ is summable.

We can then apply the Dominated Convergence Theorem to conclude that:

$$-\kappa^* \leq \inf_{i\in\mathbb{N}} \phi(1+\psi)y_i^\chi f_i^*$$
$$\sum_{i=1}^{\infty} y_i f_i^* \geq \sum_{i=1}^{\infty}(\phi(1+\psi)f_i^{*\psi}y_i^\chi + \kappa^*)^{1/\alpha} f_i^*$$
$$\sum_{i=1}^{\infty} f_i^* = 1.$$

This means that the constraint set to $P^*(M,\bar{\theta})$ is a closed (in the product topology) subset of a compact set, and so the constraint set is itself compact.

We now need to verify that the objective function to $P^*(M,\bar{\theta})$ is continuous in the product topology. Let $(f^n)_{n=1}^{\infty}$ be a sequence in the constraint set to $P^*(M,\bar{\theta})$ that converges to f^*. Consider the sequence $(g^n)_{n=1}^{\infty}$ of sequences defined as:

$$\begin{aligned} g_i^n &= \phi\psi(f_i^n)^{1+\psi}y_i^\chi, i \in \mathbb{N} \\ &\leq \phi\psi M^{1+\phi}y_i^{-\bar{\theta}(1+\psi)+\chi}, i \in \mathbb{N} \\ &\leq \phi\psi M^{1+\phi}y_i^{-\bar{\theta}+\alpha\bar{\theta}-\alpha}, i \in \mathbb{N} \\ &= \phi\psi M^{1+\psi}y_i^{(-1+(1-\bar{\theta}+\alpha(\bar{\theta}-1)))}, i \in \mathbb{N} \\ &= \phi\psi M^{1+\psi}y_i^{-1+(1-\bar{\theta})(1-\alpha)}, i \in \mathbb{N} \\ &\equiv \bar{g}_i, i \in \mathbb{N} \end{aligned}$$

The sequence $\bar{g}$ is summable, because:

$$-1-(\bar{\theta}-1)(1-\alpha) < -1.$$

It follows from the Dominated Convergence Theorem that:

$$lim_{n\to\infty}\sum_{i=1}^{\infty}(f_i^n)^{1+\psi}y_i^\chi = \sum_{i=1}^{\infty} f_i^{*1+\psi}y_i^\chi$$

which proves that the objective is continuous.

Thus, for any M, there is a solution to $P^*(M,\bar{\theta})$ and so a solution to $P(M,\bar{\theta})$.

Data Availability Statement

The data and code underlying this paper are available at https://doi.org/10.5281/zenodo.17360085.

References

[1] Aiyagari, S. Rao, 1994, "Uninsured Idiosyncratic Risk and Aggregate Saving," *The Quarterly Journal of Economics* 109, 659-684.

[2] Badel, A., and Huggett, M., 2017, "The Sufficient Statistic Approach: Predicting the Top of the Laffer Curve," *Journal of Monetary Economics* 87, 1–12.

[3] Badel, A., Huggett, M., and Luo, W., 2020, "Taxing Top Earners: A Human Capital Perspective," *Economic Journal* 130, 1200-1225.

[4] de Vries, T. and Toda, A., 2022, "Capital and Labor Pareto Exponents Across Space and Time," *Review of Income and Wealth* 22, 1058-1078.

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

[6] Feldstein, M., 1969, "The Effects of Taxation on Risk Taking," *Journal of Political Economy* 77, 755–764.

[7] Georgiadis, G., Ravid, D., and Szentes, B., 2024, "Flexible Moral Hazard Problems," *Econometrica* 92, 387-409.

[8] Heathcote, J., Storesletten, K., and Violante, G., 2017, "Optimal Tax Progressivity: An Analytical Framework," *Quarterly Journal of Economics* 132, 1693-1754.

[9] Huggett, M., 1993, "The Risk-Free rate in Heterogeneous Agent Incomplete-Insurance Economies," *Journal of Economic Dynamics and Control* 17, 953-969.

[10] Kocherlakota, N., 2010, *The New Dynamic Public Finance,* Princeton, NJ: Princeton University Press.

[11] Kopytov, A., Taschereau-Dumouchel, M., and Xu, Z., 2024, "The Origin of Risk," Simon Business School working paper.

[12] Mattsson, L.-G. and Weibull, J., 2023, "An Analytically Solvable Principal-Agent Problem," *Games and Economic Behavior* 140, 33-49.

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

[14] Mirrlees, J. A., 1974, "Notes on welfare economics, information, and uncertainty," in *Essays in Equilibrium Behavior Under Uncertainty,* ed. Balch, M., Fishburn, P., and Wu, S., Amsterdam: New Holland.

[15] Mirrlees, J. A., 1996, "Information and Incentives: The Economics of Carrots and Sticks," Nobel Prize Lecture, https://www.nobelprize.org/uploads/2018/06/mirrlees-lecture.pdf.

[16] Piketty, T., 2014, Supplement to Figure 14.1 of *Capital in the Twenty-First Century*, http://piketty.pse.ens.fr/files/capital21c/en/pdf/supp/TS14.1.pdf, accessed January 5, 2024.

[17] Rogerson, W., 1985, "The First-Order Approach to Principal Agent-Problems," *Econometrica* 53, 1357-1367.

[18] Saez, E., 2001, "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies* 68, 205-229.

[19] Scheuer, F., and Werning, I., 2017, "The Taxation of Superstars," *Quarterly Journal of Economics* 132, 211-270.

[20] Smith, S., and Timmerman, A., 2022, "Have Risk Premia Vanished?" *Journal of Financial Economics* 145, 553-576.

[21] Sun, Y., 2006, "The Exact Law of Large Numbers via Fubini Extension and Characterization of Insurable Risks," *Journal of Economic Theory* 126, 31-69.

[22] Stantcheva, S., 2017, "Optimal Taxation and Human Capital Policies over the Life-Cycle," *Journal of Political Economy* 125, 1931-1989.

[23] Varian, H., 1980, "Redistributive Taxation as Social Insurance," *Journal of Public Economics* 14, 49-68.