HARVARD | BUSINESS | SCHOOL

De Gustibus non est Taxandum: Heterogeneity in Preferences and Optimal Redistribution

Benjamin B. Lockwood
Matthew Weinzierl

Working Paper

12-063

September 5, 2014

De Gustibus non est Taxandum: Heterogeneity in Preferences and Optimal Redistribution

Benjamin B. Lockwood and Matthew Weinzierl*

September 5, 2014

Abstract

The prominent but unproven intuition that preference heterogeneity reduces redistribution in a standard optimal tax model is shown to hold under the plausible condition that the distribution of preferences for consumption relative to leisure rises, in terms of first-order stochastic dominance, with income. Given familiar functional form assumptions on utility and the distributions of ability and preferences, a simple statistic for the effect of preference heterogeneity on marginal tax rates is derived. Numerical simulations and suggestive empirical evidence demonstrate the link between this potentially measurable statistic and the quantitative implications of preference heterogeneity for policy.

*Lockwood: Harvard University, lockwood@fas.harvard.edu; Weinzierl: Harvard University and NBER, mweinzierl@hbs.edu. We are grateful to Robert Barro, Rafael di Tella, Alex Gelber, Caroline Hoxby, Louis Kaplow, Narayana Kocherlakota, Erzo F.P. Luttmer (the editor), Greg Mankiw, David Moss, Eric Nelson, Julio Rotemberg, Dan Shoag, Aleh Tsyvinski, Glen Weyl, Danny Yagan, anonymous referees and seminar participants at Harvard and Michigan for helpful comments and suggestions on earlier versions of this paper. We especially appreciate insightful suggestions on the formal analysis of section 1 by one of the referees for this *Journal*.

Introduction

In the early years of modern optimal tax research, theorists assumed all individuals have the same preferences over consumption and leisure. James A. Mirrlees's (1971) second simplifying assumption was that "Differences in tastes...are ignored. These raise rather different kinds of problems, and it is natural to assume them away." This simplification freed Mirrlees to assume that the only way in which people differ is in their ability to earn income.[1] His powerful approach–along with his assumption of preference homogeneity–now dominates theoretical work on tax design.

Preference heterogeneity of this form, however, is an evident feature of reality. Kahneman (2011) reports that such preference differences are widespread among young adults and correlate with outcomes later in life. Data shown in this paper from the World Values Survey reveal that respondents report a wide range of views toward the value of material possessions. More anecdotally, people choose a wide range of consumption-leisure bundles, even conditional on apparent budget constraints.

Heterogeneous preferences for consumption relative to leisure can be included in a standard Mirrleesian model without any impact on the results if society's normative attitude toward those preferences is the same as that toward income-earning abilities. In fact, in that case the distinction between preferences and ability is merely semantic, as they are also observationally equivalent. That is, an individual may earn a low income, and respond to taxes the way he does, either because he has low ability or because he has a weak relative preference for consumption.

In contrast, if society does not view these preferences as normatively equivalent to abilities, the implications for optimal taxation may be dramatic, and these implications are the focus of our paper. We analyze the impact of society adopting the normative view that individuals are to be, in the influential terminology of Marc Fleurbaey and Francois Maniquet (2004), compensated for having low abilities but held responsible for their preferences.[2] In that case, society's preferred unconstrained policy could range from, for example, full equalization of outcomes (if income differences are entirely due to ability) to no redistribution (if income differences are entirely due to preferences).

[1] Mirrlees was not the first to adopt this simplification. Arthur Pigou (1928) wrote, in a classic text: "Of course, in so far as tastes and temperaments differ, allowance ought, in strictness, to be made for this fact...But, since it is impossible in practice to take account of variations between different people's capacity for enjoyment, this consideration must be ignored, and the assumption made, for want of a better, that temperamentally all taxpayers are alike."

[2] Other ways in which individuals vary may merit partial compensation. We limit our focus to the form of preference heterogeneity most clearly distinct from income-earning ability. See Kaplow (2008) for a discussion of other specific cases.

Despite an early demonstration of their potential importance by Agnar Sandmo (1993), results characterizing the effects of this form of preference heterogeneity on optimal tax policy in a general setting have proven elusive.[3] This lack of results has left us without a clear understanding of the conditions under which the prominent but unproven intuition that heterogeneity in preferences lowers optimal redistribution holds and, when it does hold, how large the effects are. For example, despite the arguments made by prominent critics of redistribution,[4] in principle adding preference heterogeneity to the model may *increase* optimal redistribution. Intuitively, if preferences for consumption relative to leisure are lower among those with high incomes, attributing income variation to ability alone will understate the income-earning abilities of high earners and imply an optimal extent of redistribution that is too small.

In this paper, we derive two novel results that clarify how the presence of preference heterogeneity affects the optimal extent of income redistribution. In both cases, we show that there is a transparent formal mechanism through which we can model the effects of preference heterogeneity: namely that it changes the pattern of welfare weights the social planner assigns along the income distribution. Throughout, we refer to the conventional case in which all income heterogeneity is treated as due to ability differences or, equivalently, to differences in characteristics with the same normative implications as ability, as the "homogeneous preferences" case.

Our first contribution is to show that heterogeneity in preferences lowers optimal redistribution under a specific, plausible condition: if the distribution of the relative preference for consumption over leisure rises with income (in terms of first-order stochastic dominance), then optimal marginal tax rates are lower at all incomes and the net transfer to the lowest earner is smaller than in the homogeneous preferences case. Using the standard optimal tax model, we show this result

[3] Mirrlees (1976, 1986) addressed the general case but obtained inconclusive results. Some prior work adopts specialized settings, such as Sandmo's (1993) insightful analysis with only preference (not ability) heterogeneity; Robin Boadway, Maurice Marchand, Pierre Pestieau, and Maria del Mar Racionero's (2002) results with two preference types, two ability levels, and quasilinear utility; and Fleurbaey and Maniquet's (2006) analysis with a specific normative approach. Other work has focused on numerical simulations, such as Ritva Tarkiainen and Matti Tuomala (2007) or Kenneth L. Judd and Che-Lin Su (2006), who explain the computational complexities associated with multiple dimensions of heterogeneity. Two other recent papers focus on related but somewhat different questions. Narayana Kocherlakota and Christopher Phelan (2009) focus on the policy implications of uncertainty over the relationship between individuals' preferences and another, welfare-relevant, dimension of heterogeneity such as wealth. They argue that such uncertainty causes a planner using a maximin objective to avoid redistributive policy that is optimal when no such uncertainty is present. Paul Beaudry, Charles Blackorby, and Dezso Szalay (2009) indirectly address preference differences by including in their optimal tax analysis differences in productivity of market and non-market labor effort. They show that the optimal redistributive policy makes transfers to the poor conditional on work.

[4] See Robert Nozick (1974), "Why should we treat the man whose happiness requires certain material goods or services differently from the man whose preferences and desires make such goods unnecessary for his happiness?" Or, Milton Friedman (1962), "Given individuals whom we are prepared to regard as alike in ability and initial resources, if some have a greater taste for leisure and others for marketable goods, inequality of return through the market is necessary to achieve equality of total return or equality of treatment."

analytically for the case of quasilinear utility studied in Diamond (1998) and isoelastic welfare weights that decrease with ability. We also show, through numerical simulations, that the result holds for more general functional forms of utility and social welfare. As a consequence, we argue that the conventional assumption of preference homogeneity likely overstates the optimal extent of redistribution.

Second, we derive a simple statistic for quantifying the effect of heterogeneity in preferences on optimal marginal tax rates and redistribution. That statistic–the coefficient on the best linear predictor of log preference conditional on log unified type–is a sufficient statistic for marginal tax rates if we assume certain familiar functional forms for the distributions of ability and preferences, but it also can be used more broadly as an intuitive guide to the role of preferences. We demonstrate the link between this statistic and the quantitative implications of preference heterogeneity for optimal policy using numerical simulations calibrated to the U.S. economy. We also generate empirical estimates of this statistic for OECD countries and use them to show suggestive evidence that existing policy variation appears to be consistent with our theoretical findings. Though this simple statistic is not observable through conventional economic data, our findings suggest it is a valuable target for future empirical work.

We obtain our novel analytical results by combining two recent innovations in the literature with a third innovation of our own. First, in a setting with a continuum of agents and standard utility functions, Philippe Choné and Guy Laroque (2010) show how heterogeneity in the opportunity cost of work can justify negative marginal tax rates at low incomes. They achieve this important finding in part by collapsing multiple dimensions of heterogeneity into a single composite characteristic that determines behavior.[5] We focus on a form of preferences–i.e. for utility from consumption relative to disutility of labor effort–that has the same effects on behavior as ability and therefore allows us, like Choné and Laroque, to obtain an analytically tractable model in which individuals differ in multiple ways.[6] Related, our formal approach has much in common with theirs. Second, we adopt the moral reasoning behind the "second fairness requirement" in the prominent work of Marc Fleurbaey and Francois Maniquet (2006), which states that "the laisser-faire (this is, the absence of redistribution) should be the social optimum in the hypothetical case when all agents have equal earning abilities" even if they have different preferences.[7] In other words, we adopt the normative perspective that

[5] This technique is similar to that used by Craig Brett and John Weymark (2003). Casey Rothschild and Florian Scheuer (2013) use a different method to avoid the technical problems with multi-dimensional income-earning ability.

[6] This technique cannot help with all dimensions of heterogeneity, such as time discounting as in Mikhail Golosov, Maxim Troshkin, Aleh Tsyvinski, and Matthew Weinzierl (2013) or Peter Diamond and Johannes Spinnewijn (2011).

[7] Fleurbaey and Maniquet (2006) impose informational constraints on the social planner which rule out conventional

preferences over consumption and leisure do not justify redistribution by themselves. Though specific, this interpretation is the natural one if preferences are thought of as tastes as opposed to, for example, needs (see Kaplow 2008 for a discussion). Third, and crucially, we introduce the technique of studying how optimal policy changes when a given distribution of income is attributed to more than one source of heterogeneity, rather than how optimal policy changes when ability is augmented with additional sources of heterogeneity that change the distribution of income. This shift makes possible our progress over prior results. It has the additional virtue of formulating the problem in a way resembling that faced by policy makers, who must decide the appropriate extent of redistribution in the face of an observable income distribution.

The paper proceeds as follows. Section 1 presents a standard optimal tax model that explicitly incorporates preference heterogeneity and derives our result on its implications for redistribution. Section 2 describes a simple summary statistic for quantifying the effect of preference heterogeneity on optimal policy and shows its usefulness through both calibrated numerical simulations and suggestive empirical evidence. After the Conclusion, proofs and numerical simulations demonstrating the robustness of our findings to the functional forms of utility and social welfare are collected in the Appendix, labeled Section 3.

1 Optimal Income Taxation with Heterogeneous Preferences

Our first novel analytic result is to show an intuitive and plausible condition under which the presence of preference heterogeneity reduces the optimal extent of redistribution. First we present a simple weighted utilitarian version of the standard Mirrlees (1971) model of optimal income taxation with generalized type-specific "welfare weights," and we demonstrate a relationship between the structure of those weights and optimal redistribution. We then modify that model to allow heterogeneity in type to be attributed to observationally equivalent but (possibly) normatively distinct differences in income-earning ability and preferences for consumption relative to leisure. We impose a natural normative requirement, "preference neutrality", on the relationship between welfare weights and preferences, and we show how preference-neutral welfare weights are determined endogenously from the joint distribution of ability and preferences. Finally, we show the conditions

utilitarian social welfare functions and which, in combination with particular fairness requirements on allocations, imply the use of a maximin social welfare function. They conclude that the optimal income tax should maximize the subsidies to the working poor: that is, it should be quite redistributive to those with low ability but who exert labor effort. Our analysis can be seen as a complement to theirs, studying the same type of preference heterogeneity in a setting closer to the more conventional Mirrleesian approach.

under which the presence of preference heterogeneity results in less (or more) redistribution at the optimum.

1.1 A standard model with homogeneous preferences

Individuals have utility of consumption c and labor effort ℓ given by $u(c, \ell) = c - \ell^{1+1/\varepsilon}$ where ε is the constant elasticity of labor supply. As in Mirrlees (1971), they are indexed by unobservable ability $n \geq 0$, equal to their (assumed constant) marginal product of labor effort so that gross income y is equal to $n\ell$. Thus we can write utility as a function of consumption, earnings, and type:

$$U(c, y, n) = c - (y/n)^{1+1/\varepsilon}. \tag{1}$$

Ability is distributed according to $F(n)$ with assumed density $f(n)$, and a planner selects the allocation $\{c(n), y(n)\}$ to maximize a weighted sum of utilities, solving

$$\max_{\{c(n),y(n)\}} \int_0^\infty g(n)U(c(n), y(n), n)f(n)\, dn, \tag{2}$$

where $g(n) \geq 0$ denotes the "welfare weight" for type n (sometimes called the Pareto weight) assumed to be continuous and decreasing in n but otherwise left exogenous until the next subsection.[8] The maximization in (2) is subject to a resource constraint,

$$\int_0^\infty (y(n) - c(n))f(n)\, dn \geq 0, \tag{3}$$

and incentive compatibility (IC) constraints

$$U(c(n), y(n), n) \geq U(c(m), y(m), n), \quad \forall m, n.$$

In this setup, an allocation $\{c(n), y(n)\}$ may be implemented by specifying a corresponding income tax function $T(y) = y - c$, in which case the IC constraints can be written as follows:

$$y(n) = \arg\max_y \left\{ y - T(y) - (y/n)^{1+1/\varepsilon} \right\}, \quad \forall n. \tag{4}$$

[8] This setup is similar to the structure in Diamond (1998), in which the planner maximizes $\int_0^\infty \Psi(U(c(n), y(n), n))f(n)\, dn$, for concave $\Psi(\cdot)$, in that any concave $\Psi(\cdot)$ can be used to construct type-specific welfare weights with $g(n) = \Psi'(U(c(n), y(n), n))$ at the optimal allocation. These weights will then give rise to the same second-best optimal allocation as $\Psi(\cdot)$ in our setting. See Werning (2007) and Salanié (2011) for related analyses.

As in Diamond (1998), we represent these IC constraints with the first-order conditions for each type's choice of y:[9]

$$1 - T'(y(n)) - (1 + 1/\varepsilon)\left(\frac{y(n)}{n^{1+\varepsilon}}\right)^{1/\varepsilon} = 0, \quad \forall n.$$

In this case, the optimal tax function is characterized the following first-order condition (Diamond, 1998).

$$\frac{T'(y(n))}{1 - T'(y(n))} = \frac{1 + 1/\varepsilon}{nf(n)}(G(n) - F(n)), \quad \forall n, \tag{5}$$

where

$$G(n) = \frac{\int_{m=0}^{n} g(m)f(m)\,dm}{\int_{m=0}^{\infty} g(m)f(m)\,dm} \tag{6}$$

is the "cumulative welfare weight" at n, normalized so that $G(0) = 0$ and $\lim_{n\to\infty} G(n) = 1$.

1.2 A relationship between welfare weights and optimal redistribution

We are interested in the relationship between welfare weights and the shape of the optimal tax function $T(y(n))$. Here it is useful to establish a partial ranking of tax functions based on their "redistributivity". Loosely, a tax is considered redistributive when it transfers resources from higher earners, for whom $T(y(n)) > 0$, to lower earners via a lump sum grant $-T(y(0)) > 0$. Formally, we will employ the following definition:

Definition 1 *The income tax function $T(y(n))$ is "less redistributive" than $\hat{T}(\hat{y}(n))$ if it imposes weakly lower marginal income tax rates on all types (i.e., $T'(y(n)) \leq \hat{T}'(\hat{y}(n))$ for all n) and provides a strictly smaller lump sum grant, $-T(y(0)) < -\hat{T}(\hat{y}(0))$.*

It is worth noting that this is a rather demanding definition which leaves many pairs of tax functions unranked in terms of redistributivity. For example, under this definition a function $T(y(n))$ that decreases the lump sum grant and most marginal tax rates but increases a subset of marginal tax rates relative to $\hat{T}(\hat{y}(n))$ would not qualify as less redistributive. The strictness of this definition helps us to avoid ambiguity in our results on the extent of optimal redistribution.

Using this definition, we obtain the following relationship between cumulative welfare weights and optimal redistribution.

[9] This assumption is equivalent to assuming that, at the optimum, $T(y)$ is differentiable and $y'(n) > 0$, the latter of which ensures that the individual's choice is globally optimal (note that the Spence-Mirrlees single crossing condition is ensured by the functional form of $U(c, y, n)$. See Salanié, 2011).

Lemma 1 *Consider income tax functions $T(y(n))$ and $\hat{T}(\hat{y}(n))$ that solve the planner's problem in* (2)*,* (3)*, and* (4) *given welfare weights $g(n)$ and $\hat{g}(n)$ that are everywhere continuous, positive, and decreasing in n. If the corresponding cumulative welfare weights $G(n)$ and $\hat{G}(n)$ defined in* (6) *satisfy $G(n) < \hat{G}(n)$ for all $n > 0$, then $T(y(n))$ is less redistributive than $\hat{T}(\hat{y}(n))$.*

As this lemma suggests, the shape of the welfare weights $g(n)$ will be key to our results. We now turn to the characterization of that shape when preferences are heterogeneous.

1.3 Welfare weights in the presence of heterogeneous preferences

We begin by introducing a modification to the model above. Individuals are now characterized by a two-dimensional type, (w, θ), where $w \geq 0$ is an individual's unobservable ability (their marginal product of labor effort) so that $y = w\ell$, while $\theta > 0$ is an unobservable preference parameter scaling the disutility that an individual experiences from exerting labor effort relative to the utility the individual experiences from consumption. We assume the units on θ are such that the population average of θ is equal to one. Whereas utility of consumption and labor effort, $u(c, \ell)$, was homogeneous in the previous section, it now depends on the preference parameter: $u(c, \ell, \theta) = c - (\ell/\theta)^{1+1/\varepsilon}$. We can also write individual utility analogously to (1) as follows:

$$U(c, y, w, \theta) = c - \left(\frac{y}{w\theta}\right)^{1+1/\varepsilon}. \tag{7}$$

The structure of (7) demonstrates that agents with different pairs of types face the same maximization problem. Specifically, an individual of type (w', θ') behaves exactly like another individual of type $(w'', \theta'') \neq (w', \theta')$ if $w'\theta' = w''\theta''$.

The product $w\theta$ is thus a sufficient statistic for labor supply behavior—we will call this product the individual's "unified type". Because it is impossible to distinguish between individuals of the same unified type, policy must treat them identically. Thus the planner's choice space is the set of allocations $\{c(w\theta), y(w\theta)\}$.

We again assume the planner seeks to maximize a weighted sum of utilities, and we denote the welfare weights $b(w, \theta)$ to reflect their possible dependence on both ability and preferences. Letting $H(w, \theta)$ denote the joint distribution of ability and preferences, with density $h(w, \theta)$, the planner's objective in this modified problem is

$$\max_{\{c(w\theta), y(w\theta)\}} \int_{\theta=0}^{\infty} \int_{w=0}^{\infty} b(w, \theta) U(c(w\theta), y(w\theta), w, \theta) h(w, \theta)\, dw\, d\theta. \tag{8}$$

This maximization is subject to the resource constraint

$$\int_{\theta=0}^{\infty}\int_{w=0}^{\infty}(y(w\theta)-c(w\theta))h(w,\theta)\,dw\,d\theta\geq 0, \tag{9}$$

and IC constraints, written in terms of the tax function $T(y(w\theta))=y(w\theta)-c(w\theta)$,

$$y(w\theta)=\arg\max_{y}\left\{y-T(y)-\left(\frac{y}{w\theta}\right)^{1+1/\varepsilon}\right\},\quad \forall w,\theta. \tag{10}$$

Our key normative assumption is a condition on $b(w,\theta)$, which can be stated as follows.

Preference neutrality. Welfare weights are independent of preferences, that is $b(w,\theta)=b(w,\theta')$ for all θ and θ', so we define $b\,(w)\equiv b\,(w,\theta)$ for all θ.

This condition, motivated by the ethical considerations axiomatized in Fleurbaey and Maniquet (2006), amounts to assuming that income differences arising from differences in budget constraints merit redistribution, whereas those arising from differences in preferences do not. Under preference neutrality, the objective in (8) can be written

$$\max_{\{c(w\theta),y(w\theta)\}}\int_{\theta=0}^{\infty}\int_{w=0}^{\infty}b(w)U(c(w\theta),y(w\theta),w,\theta)h(w,\theta)\,dw\,d\theta. \tag{11}$$

Letting n denote unified type, so that $n=w\theta$, we can employ a change of variables, using $\tilde{H}(\theta,n)$ to denote the joint distribution of preferences and unified type, with density

$$\tilde{h}(\theta,n)=h(n/\theta,\theta). \tag{12}$$

Further, we let $f(n)=\int_0^{\infty}\tilde{h}(\theta,n)\,d\theta$, denoting the density of unified types arising from a given joint distribution $H(w,\theta)$. Then, substitution shows that the resource constraint (9) and the IC constraints (10) are equivalent to (3) and (4) from the previous section. Moreover, the preference neutral planner's objective (11) can be written

$$\max_{\{c(n),y(n)\}}\int_{n=0}^{\infty}\int_{\theta=0}^{\infty}b(n/\theta)U(c(n),y(n),n)\tilde{h}(\theta,n)\,d\theta\,dn=$$
$$\max_{\{c(n),y(n)\}}\int_{n=0}^{\infty}\left(\frac{\int_{\theta=0}^{\infty}b(n/\theta)\tilde{h}(\theta,n)\,d\theta}{f(n)}\right)U(c(n),y(n),n)f(n)\,dn,$$

or simply

$$\max_{\{c(n),y(n)\}} \int_{n=0}^{\infty} \bar{b}(n) U(c(n), y(n), n) f(n)\, dn, \tag{13}$$

where $\bar{b}(n)$ is the mean welfare weight on individuals of unified type n under the distribution $H(w, \theta)$

$$\bar{b}(n) = \frac{\int_{\theta=0}^{\infty} b(n/\theta)\tilde{h}(\theta, n)\, d\theta}{f(n)}. \tag{14}$$

Note that the objective (13) is equivalent to (2), with $\bar{b}(n)$ replacing $g(n)$.

In principle, the distribution $H(w, \theta)$ could be such that $\bar{b}(n)$ would be increasing in n, even if $b(w)$ decreases in w. Such a situation would merit not only a reduction in redistributivity but in fact a reversal, i.e., redistribution to higher earners, and would require income and ability to be negatively correlated. Because we view this possibility as empirically implausible, we will set aside the technical complexities associated with this possibility and assume that $\bar{b}'(n) < 0$.

Assuming that $H(w, \theta)$ gives rise to a distribution of unified types $F(n)$ which satisfies the standard regularity assumptions as in section 1.1, the optimal tax function in this model with two dimensional heterogeneity satisfies the familiar condition (5), with cumulative welfare weights now given by

$$G(n) = \frac{\int_{m=0}^{n} \bar{b}(m) f(m)\, dm}{\int_{m=0}^{\infty} \bar{b}(m) f(m)\, dm}. \tag{15}$$

The solution to the planner's problem in this modified setup provides a useful deconstruction of the welfare weights $g(n)$ while being formally equivalent to the model of the previous section. That is, it allows us to distinguish between two possible sources of disagreement about the optimal extent of redistribution—the weights $b(w)$, and the joint distribution $H(w, \theta)$–that together produce the policy-relevant weights $g(n)$. In the next section we explore the implications of disagreements about the second of these sources as a simple way to capture the effects of preference heterogeneity on optimal policy.

1.4 Preference neutrality and optimal redistribution

We can now prove our first novel analytic result using the expressions from previous subsections. It may facilitate intuition to imagine two hypothetical planners. Those two planners agree on the distribution of unified types $F(n)$, the principle of preference neutrality, and the appropriate

ability-dependent welfare weights $b(w)$, which take an isoelastic form:

$$b(w) = \kappa w^{-\eta}, \tag{16}$$

for positive constants κ and η. But, these two planners have differing positive beliefs about the joint distribution of ability and preferences, $H(w, \theta)$. In particular, one planner knows the true distribution $H(w, \theta)$, while the other incorrectly believes $H(w, \theta)$ is degenerate along the θ dimension, with $\theta = 1$ for all individuals. We use carats to denote the latter planner's incorrect beliefs. This disagreement results in different policy relevant welfare weights $g(n)$ and $\hat{g}(n)$, and thus different preferred tax functions $T(y(n))$ and $\hat{T}(\hat{y}(n))$. Specifically, we can show the following proposition.

Proposition 1 *Assume the individual utility function is quasilinear as in* (1)*, and welfare weights* $b(w)$ *are isoelastic as in* (16)*. Consider the income tax function* $T(y(n))$ *that solves the planner's problem in* (13)*,* (3)*, and* (4) *assuming the joint distribution* $H(w, \theta)$*. Consider also the income tax function* $\hat{T}(\hat{y}(n))$ *that solves the same planner's problem but assuming homogeneous preferences, that is* $\theta = 1$ *for all* n*. Let* $\tilde{H}(\theta|n)$ *be determined by* $H(w, \theta)$ *according to* (12)*. If the conditional distribution* $\tilde{H}(\theta|n)$ *first-order stochastically dominates* $\tilde{H}(\theta|m)$ *whenever* $n > m$*, then* $T(y(n))$ *is less redistributive than* $\hat{T}(\hat{y}(n))$*.*

Given certain tractable functional forms for utility and social welfare, Proposition 1 establishes a simple and plausible condition under which preference heterogeneity reduces the optimal extent of redistribution. A direct corollary clarifies the related condition under which the opposite result holds, namely that the presence of preference heterogeneity increases the optimal extent of redistribution, a possibility noted in the Introduction.

Corollary 1 *Assume the same conditions as in Proposition 1. If the conditional distribution* $\tilde{H}(\theta|n)$ *is first-order stochastically dominated by* $\tilde{H}(\theta|m)$ *whenever* $n > m$*, then* $T(y(n))$ *is more redistributive than* $\hat{T}(\hat{y}(n))$*.*

These analytical results help us better understand the qualitative effect of preference heterogeneity on redistribution. In the next section we look for a similarly simple guide to the size of this effect.

2 A simple statistic for the quantitative effects of preference heterogeneity on redistribution

In this section, we introduce an intuitive summary statistic for the quantitative effects of preference heterogeneity on redistribution. Assuming certain functional forms for welfare weights, individual utility, and the distributions of ability and preferences, we show that this statistic is in fact sufficient to characterize the effects of preference heterogeneity on marginal tax rates. More generally, we show this statistic's usefulness through both calibrated numerical simulations of optimal policy in the United States and empirical evidence on existing policies and preference heterogeneity in OECD countries.

The statistic of interest is

$$\beta = \frac{\text{cov}(\ln \theta_i, \ln n_i)}{\text{var}(\ln n_i)}, \tag{17}$$

the coefficient on the best linear predictor of log preference conditional on log unified type. In other words, β captures the expected increase in preferences for an increase in unified type.

It is possible to provide a more formal characterization of the role of β given certain simplifying assumptions about the economy. In particular, we assume that the distributions of ability and preferences are jointly lognormal, so that the distribution of unified types n is also lognormal.[10] We can then show that marginal tax rates depend on the distribution of θ only through the statistic β, as in the following proposition:

Proposition 2 *Suppose the welfare weights $b(w)$ are isoelastic as in* (16)*, individual utility is quasilinear as in* (1)*, and $\ln w$ and $\ln \theta$ are jointly normally distributed so that the distribution $f(n)$ of unified type n is lognormal with $\sigma_N = Var[\ln(n)]$. Then, optimal marginal tax rates $T'(y(n))$ satisfy:*

$$\frac{T'(y(n))}{1 - T'(y(n))} = \frac{1 + 1/\varepsilon}{nf(n)} \left(\int_0^{n \exp(\eta(1-\beta)\sigma_N)} f(m)dm - \int_0^n f(m)dm \right). \tag{18}$$

As with the first proposition, the mechanism behind Proposition 2 is that β affects the shape of the welfare weights $g(n)$, which in turn determine the first integral in (18). In particular, that integral decreases with β, with extreme cases providing especially clear results. In the case

[10] Though evidence (see Saez 2001) shows that the upper tail of the income distribution is better described as a Pareto distribution, lognormality has long been used in the optimal tax literature to describe most of the income distribution (see Tuomala 1990) below the top tail.

where $\theta = 1$ for all individuals, $\beta = 0$ and the first integral in (18) can be shown to equal $\int_{m=0}^{n} b(m) f(m)\, dm / \int_{m=0}^{\infty} b(m) f(m)\, dm$ where $b(m)$ satisfies (16), so the integral equals $G(n)$ from the conventional homogeneous preferences case. At the opposite extreme, if ability is homogeneous and all behavioral heterogeneity is due to preferences, $\beta = 1$ and the first integral in (18) equals $\int_0^n f(n)dn$, so optimal tax rates are uniformly zero.

2.1 Numerical simulations of optimal policy

We now use calibrated numerical simulations to illustrate the potential quantitative effects of preference heterogeneity on optimal policy and the usefulness of the statistic β in measuring them.

Our calibration strategy is to match the income distribution chosen by individuals as modeled in Section 1, taking U.S. tax policy as given, to the empirical income distribution in the United States, and thus infer a distribution of unified types $F(n)$. We use a baseline labor supply elasticity value of $\varepsilon = 0.33$, the preferred estimate in Chetty (2012), accounting for optimization frictions.

To calibrate the ability distribution, we assume that unified types n are drawn from a lognormal distribution with parameters μ_N and σ_N, and we select these parameters so that resulting income distribution approximates the empirical distribution in the US in 2011.[11] The resulting parameter estimates, when incomes are reported in \$10,000s, are $\mu_N = 1.65$ and $\sigma_N = 0.65$. Our conceptual results are not sensitive to these values, but having a realistic calibration makes the magnitudes of our results easier to interpret.

The optimal policy naturally depends on the planner's welfare weights. We assume they take the iso-elastic form in (16) where the planner's inequality aversion is measured by η. We use a baseline value of $\eta = 1$. We then vary β as defined in (17) to see how optimal policy diverges from that which assumes no preference heterogeneity. Figure 1 plots marginal tax rates from (18) for four values for β, ranging from 0 (the conventional, homogeneous preferences case) to 0.75, which loosely corresponds to three fourths of income variation deriving from preferences. The extreme case of $\beta = 1$, in which all income variation is due to preferences and taxes are optimally zero, is omitted.

[11] Specifically, we select the parameters which minimize the sum of squared differences between incomes at percentiles 20, 40, 60, 80, and 90 under the simulated distribution and the actual income distribution in the US in 2011, as reported by the Tax Policy Center. For our computational expediency, we perform these simulations assuming flat taxes, as in Saez 2001, with a marginal tax rate of 30%.

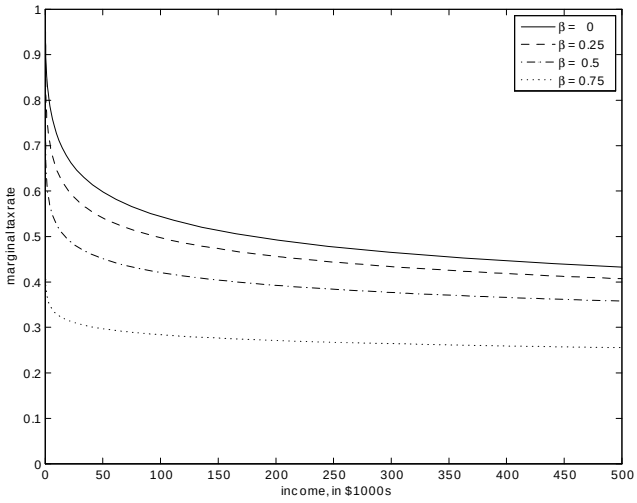

Figure 1: Optimal marginal tax rate schedules for four values of β.

These results expand quantitatively on the qualitative result in Proposition 1. Under this baseline specification, for example, if $\beta = 0.25$ (so that roughly one fourth of income heterogeneity is due to preferences) the optimal marginal tax rate falls by 5.6 percentage points for individuals earning \$50,000, and by 2.5 percentage points on those earning \$500,000. This represents a substantial change in redistributive policy—the net transfer to individuals at the 10th percentile of the income distribution falls by \$2180 annually, while net taxes levied on those at the 90th percentile decrease by \$9600.

The analytic proof of Proposition 1 imposed two requirements: an absence of income effects, and Pareto weights which are isoelastic in unified type n. In the appendix, we relax both assumptions and find that the inverse relationship between β and marginal tax rates still holds. Simulations are performed using (18) in the baseline case, and numerically using the Knitro nonlinear optimization package (see Byrd et al., 2006) in the appendix.

Though β is not conventionally observable, these simulations demonstrate its value as a straightforward way in which to modify a numerical version of the model to determine the potential quantitative implications of preference heterogeneity. Moreover, β may be a more plausible empirical target than has previously been identified, especially if unconventional sources of evidence are brought to bear, as we now show.

2.2 Suggestive empirical patterns

To demonstrate the empirical potential of our results, and to reinforce the usefulness of the population statistic β, we now provide suggestive evidence that preference heterogeneity may be related to real-world policy across modern developed countries in the way that our analytical results suggest. We emphasize that these results are admittedly far from conclusive and are vulnerable to a variety of criticisms. Our hope is that they stimulate further data gathering and empirical work that can more reliably test for the implications of the theory in existing policy.

Our cross-sectional[12] international data on preference heterogeneity and β comes from the World Values Survey, whose international coverage of attitudes toward such topics is unmatched. The World Values Survey asked the following question of respondents in a set of countries between 2005 and 2007: *"Now I will briefly describe some people. Using this card, would you please indicate for each description whether that person is very much like you, like you, somewhat like you, not like you, or not at all like you? ...It is important to this person to be rich; to have a lot of money and expensive things."* We will use the answers to the question to measure preferences θ.[13] This question is well-posed for our purposes, as it attempts to have the respondent reflect on his or her underlying preferences rather than how he or she feels in the status quo, i.e., "on the margin." Key for our purposes, the World Values Survey also asks respondents to report their place in the income distribution (it asks which of ten "steps" the respondent's household income falls into). Since income increases monotonically with unified type, we use these reported values as a measure of unified type n.[14] It is possible to calculate the covariance of these data within each country, giving us all of the components required to calculate β.[15]

We relate these values of β to a standard measure of redistribution, the difference between the

[12] Panel analysis would be desirable, but the survey data we use to measure preferences is available over at most a ten-year horizon. We believe this is too narrow a window over which to expect either meaningful changes in preference variation or a response to any such changes in policy, so we leave the analysis of panel data for future research.

[13] To be precise, we interpret these answers, which we rescale to run from 2 to 10 rather than 1 to 5 to more closely match the scale for the income question, as indicating values of $\ln\left(\theta^{1+1/\varepsilon}\right)$, the log of the observationally-equivalent preference factor that could be applied to the subutility of consumption rather than to the disutility of labor effort in the utility function (1). We therefore scale the responses to convert them to values of θ with an expected value of one, to match the assumption in Section 1.3, and then take the logarithms of those θ values, before calculating β. Very similar results are obtained if we use simply the reported levels of β, instead.

[14] The distribution of responses is far from uniform across deciles in most countries, which could reflect either the pattern of sampling or surveyor and respondent behavior toward such a question. We calculate β in this analysis using the levels of the answers to the questions on income and preferences (rescaled–see the preceding footnote), not the logs of those levels. The reason is that income is reported on a linear scale from 1 to 10, so that individuals are likely interpreting the scale roughly as deciles. If income is distributed roughly log-normally, taking logs of these levels would be redundant. For consistency, we assume that the preference scale represents a similar implicit transformation.

[15] Uncertainty over how respondents interpreted the scale of choices for the WVS question on preferences makes the absolute levels of β calculated in this subsection less meaningful than their relative levels across countries.

Gini coefficients on primary (pre-tax and pre-transfer) income and disposable (post-tax and post-transfer) income, as calculated in Wang and Caminada (2011) based on data from the Luxembourg Income Study.[16] We are able to calculate both β and this measure of redistribution for 13 countries with PPP-adjusted GDP per capita greater than $25,000 in 2005 U.S. dollars, a simple threshold that helps to control for the wide range of institutional variables that likely affect the scale of redistribution. Figure 2 shows the results visually.

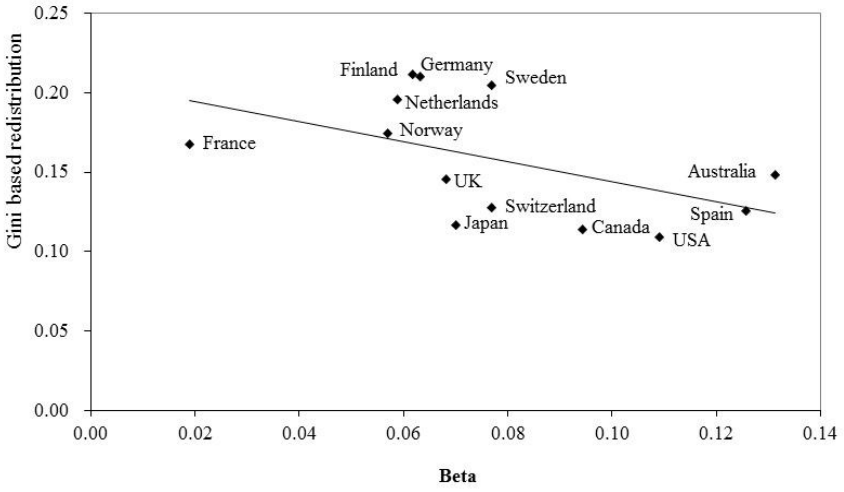

Figure 2: Redistribution and β in 13 OECD countries

A mild but noticeable negative relationship between redistribution and β is apparent in Figure 2, consistent with the theory developed above. That is, countries in which preference heterogeneity plays a larger role in explaining income variation appear to have less redistributive policies. The point estimate of the coefficient on β is negative and marginally significant (it is -0.67 with a standard error of 0.31); it is also the slope of the best-fit line shown in the figure. Though this evidence is far from definitive, this relationship is robust to controlling for the log of GDP per capita and the extent of inequality as measured by the pre-tax Gini coefficient. Of course, any results with such a limited sample are merely suggestive of a relationship that, given the potential feasibility of measuring the statistic β, may reward greater study.

To be clear, this relationship may very well reflect interdependence rather than unidirectional causality and be consistent with the theoretical results developed above. For example, it may be

[16] Japan is the one exception, as it does not participate in the LIS. We rely on Tachibanaki (2005), who reports primary and disposable income Gini coefficients based on Japan's Ministry of Welfare and Labor's Income Redistribution Survey in his Table 1.1.

that residents of countries with less redistributive policies tend to evolve toward having less similar preferences. Related, it may be of interest to note that the pattern in Figure 2 is consistent with the well-known results of Alesina and Angeletos (2005), who find that countries in which effort is perceived to be more important than luck in determining economic success have less redistributive policies. In the terminology of this paper, larger β implies that preferences play a larger role in determining income relative to ability–that is, the source of heterogeneity not deserving of redistribution is relatively more important.

Conclusion

Preference differences have played a relatively minor role in optimal tax research thus far, but they are readily apparent in the real world and have long been a staple of broader debates over taxation. We focus on the implications of heterogeneity in preferences for consumption relative to leisure that are observationally equivalent, but not normatively equivalent, to income-earning abilities, in that society does not wish to redistribute income based on these preferences. We show two novel results on how this heterogeneity affects the optimal extent of redistribution. In both cases, we isolate a transparent formal mechanism through which we can model the operation of preference heterogeneity, namely changing the pattern of welfare weights the social planner assigns along the income distribution. First, we show that long-standing intuitions about this form of preference heterogeneity reducing optimal redistribution are incomplete but correct given a plausible condition on how preferences relate to income. Second, we describe a simple, and under certain assumptions sufficient, statistic for measuring this effect, providing a potentially empirically-relevant way to gauge the quantitative implications of preference heterogeneity for redistribution.

3 Appendix

3.1 Proof of Lemma 1

It follows immediately from (5) and the assumption that $G(n) < \hat{G}(n)$ for all $n > 0$ that marginal tax rates are strictly less for all $n > 0$ under $T(y(n))$ than under $\hat{T}(\hat{y}(n))$. To show that this implies $-T(y(0)) < -\hat{T}(\hat{y}(0))$, suppose to the contrary that $-T(y(0)) \geq -\hat{T}(\hat{y}(0))$. Note that any positive perturbation to $T'(y(n^*))$ at some n^*, while holding the lump sum grant $-T(\hat{y}(0))$ fixed, lowers utility $y(n) - T(y(n)) - (y(n)/n)^{1+1/\varepsilon}$ for all $n > n^*$, while leaving utility unaffected for all $n \leq n^*$. By extension, an increase in $T'(y(n))$ for all $n > 0$, holding $-T(0)$ fixed, reduces utility for all $n > 0$ and leaves $U(c(0), y(0), 0)$ unchanged—a Pareto inferior reform. Note that $T'(y(n)) < \hat{T}'(\hat{y}(n))$ for all n, so we can conclude that $\hat{T}(\hat{y}(n))$ is Pareto inferior to $T(y(n))$ if $-T(y(0)) \geq -\hat{T}(\hat{y}(0))$. By assumption that $\hat{g}(n) > 0$, however, $\hat{T}(\hat{y}(n))$ is Pareto efficient. Therefore, the supposition that $-T(y(0)) \geq -\hat{T}(\hat{y}(0))$ must be false, implying $-T(y(0)) < -\hat{T}(\hat{y}(0))$. Therefore $T(y(n))$ is less redistributive than $\hat{T}(\hat{y}(n))$.

3.2 Proof of Proposition 1

The mistaken planner believes $w = n/\bar{\theta} = n$ for all individuals, and thus $\hat{g}(n) = b(n)$. The correct planner instead believes $g(n) = \bar{b}(n)$ as given by (14). Let $\alpha(n)$ denote the ratio of the latter to the former:

$$\alpha(n) = \frac{g(n)}{\hat{g}(n)} = \frac{\bar{b}(n)}{b(n)} = \frac{\int_{\theta=0}^{\infty} b(n/\theta)\tilde{h}(\theta|n)\, d\theta}{b(n)}, \tag{19}$$

where $\tilde{h}(\theta|n) = \tilde{h}(\theta, n)/f(n)$, the conditional density of θ given $w\theta = n$ under $H(w, \theta)$. Note that

$$\hat{G}(n) = \frac{\int_{m=0}^{n} \hat{g}(m)f(m)\, dm}{\int_{m=0}^{\infty} \hat{g}(m)f(m)\, dm} = \frac{\int_{m=0}^{n} b(m)f(m)\, dm}{\int_{m=0}^{\infty} b(m)f(m)\, dm}$$

and

$$G(n) = \frac{\int_{m=0}^{n} g(m)f(m)\, dm}{\int_{m=0}^{\infty} g(m)f(m)\, dm} = \frac{\int_{m=0}^{n} \alpha(m)b(m)f(m)\, dm}{\int_{m=0}^{\infty} \alpha(m)b(m)f(m)\, dm}.$$

Therefore we can write

$$\hat{G}(n) - G(n) = \frac{\int_{m=0}^{n} b(m)f(m)\, dm}{\int_{m=0}^{\infty} b(m)f(m)\, dm} - \frac{\int_{m=0}^{n} \alpha(m)b(m)f(m)\, dm}{\int_{m=0}^{\infty} \alpha(m)b(m)f(m)\, dm},$$

which is proportional to (and has the same sign as)

$$\int_{m=0}^{n} b(m)f(m)\,dm \int_{m=0}^{\infty} \alpha(m)b(m)f(m)\,dm - \int_{m=0}^{n} \alpha(m)b(m)f(m)\,dm \int_{m=0}^{\infty} b(m)f(m)\,dm.$$

Combining the integrals and factoring, this expression can be rewritten

$$\int_{m=0}^{\infty}\int_{r=0}^{n} b(r)f(r)b(m)f(m)\left(\alpha(m)-\alpha(r)\right)\,dr\,dm$$

and splitting the outer integral at n, this is equivalent to

$$\int_{m=0}^{n}\int_{r=0}^{n} b(r)f(r)b(m)f(m)\left(\alpha(m)-\alpha(r)\right)\,dr\,dm + \int_{m=n}^{\infty}\int_{r=0}^{n} b(r)f(r)b(m)f(m)\left(\alpha(m)-\alpha(r)\right)\,dr\,dm.$$

The first term integrates to zero, so we conclude

$$\hat{G}(n) - G(n) \propto \int_{m=n}^{\infty}\int_{r=0}^{n} b(r)f(r)b(m)f(m)\left(\alpha(m)-\alpha(r)\right)\,dr\,dm. \tag{20}$$

Thus a sufficient condition for $\hat{G}(n) > G(n)$, for all $n > 0$, is that $\alpha(n)$ is increasing in n. To show when this is true, we differentiate (19):

$$\alpha'(n) = \frac{b(n)\int_{\theta=0}^{\infty}\left[\tilde{h}(\theta|n)\frac{d}{dn}b(n/\theta) + b(n/\theta)\frac{d}{dn}\tilde{h}(\theta|n)\right]\,d\theta - b'(n)\int_{\theta=0}^{\infty} b(n/\theta)\tilde{h}(\theta|n)\,d\theta}{b(n)^2}.$$

The denominator is positive, so the sign of $\alpha'(n)$ is determined by the numerator above, which can be rearranged as

$$\int_{\theta=0}^{\infty}\left[b(n)\frac{d}{dn}b(n/\theta) - b'(n)b(n/\theta)\right]\tilde{h}(\theta|n)\,d\theta + b(n)\int_{\theta=0}^{\infty} b(n/\theta)\frac{d}{dn}\tilde{h}(\theta|n)\,d\theta \tag{21}$$

By the assumption that $b(\cdot)$ is isoelastic, i.e., that $b(x) = \kappa x^{-\eta}$ for constants κ and η, the term in brackets is equal to zero for all n and θ. And by the assumption that $\tilde{H}(\theta|n)$ first-order stochastically dominates $\tilde{H}(\theta|m)$ for all $n > m$, decreasing $b(\cdot)$ implies that the integral on the right above is positive. Thus $\alpha'(n) > 0$ for all n, so by (20), $\hat{G}(n) > G(n)$, for all $n > 0$. Therefore by Lemma (1), the tax function $T(y\,(n))$ is less redistributive than the tax function $\hat{T}(\hat{y}\,(n))$.

3.3 Proof of Corollary 1

The proof is identical to the proof of Proposition 1, except that by the assumption that $\tilde{H}(\theta|n)$ is first-order stochastically dominated by $\tilde{H}(\theta|m)$ whenever $n > m$, expression (21) is negative, and thus $\alpha'(n) < 0$ for all n, so by expression (20), $\hat{G}(n) > G(n)$ for all $n > 0$ and the tax function $T(y(n))$ is more redistributive than $\hat{T}(\hat{y}(n))$.

3.4 Proof of Proposition 2

Rewrite (15), using capitals to denote random variables and lowercase for particular realizations, and using the fact that $\bar{b}(n) = \mathbb{E}[b(W)|N = n]$:

$$\begin{aligned}
G(n) &= \frac{\int_{m=0}^{n} \mathbb{E}[b(W)|N = m]f(m)dm}{\int_{m=0}^{\infty} \mathbb{E}[b(W)|N = m]f(m)dm} \\
&= \frac{\int_{m=0}^{n} \mathbb{E}[\kappa(W)^{-\eta}|N = m]f(m)dm}{\int_{m=0}^{\infty} \mathbb{E}[\kappa(W)^{-\eta}|N = m]f(m)dm} \\
&= \frac{\int_{m=0}^{n} \mathbb{E}[\exp(-\eta \ln W)|N = m]f(m)dm}{\int_{m=0}^{\infty} \mathbb{E}[\exp(-\eta \ln W)|N = m]f(m)dm} \\
&= \frac{\int_{m=0}^{n} \exp(\mathbb{E}[-\eta \ln W|N = m] + \mathrm{Var}[-\eta \ln W|N = m]/2)f(m)dm}{\int_{m=0}^{\infty} \exp(\mathbb{E}[-\eta \ln W|N = m] + \mathrm{Var}[-\eta \ln W|N = m]/2)f(m)dm}.
\end{aligned}$$

We can use the fact that when random variables X and Y are jointly normal, the conditional variance $\mathrm{Var}[X|Y = y]$ is independent of y, to write

$$\begin{aligned}
G(n) &= \frac{\int_{m=0}^{n} \exp(\mathbb{E}[-\eta \ln W|N = m] + \mathrm{Var}[-\eta \ln W]/2)f(m)dm}{\int_{m=0}^{\infty} \exp(\mathbb{E}[-\eta \ln W|N = m] + \mathrm{Var}[-\eta \ln W]/2)f(m)dm} \\
&= \frac{\exp(\mathrm{Var}[-\eta \ln W]/2) \int_{m=0}^{n} \exp(\mathbb{E}[-\eta \ln W|N = m])f(m)dm}{\exp(\mathrm{Var}[-\eta \ln W]/2) \int_{m=0}^{\infty} \exp(\mathbb{E}[-\eta \ln W|N = m])f(m)dm} \\
&= \frac{\int_{m=0}^{n} \exp(\mathbb{E}[-\eta \ln W|N = m])f(m)dm}{\int_{m=0}^{\infty} \exp(\mathbb{E}[-\eta \ln W|N = m])f(m)dm} \\
&= \frac{\mathbb{E}_N[\exp(\mathbb{E}_{W|N}[-\eta \ln W|N])|N < n]F(n)}{\mathbb{E}_N[\exp(\mathbb{E}_{W|N}[-\eta \ln W|N])]} \\
&= \frac{\mathbb{E}_N[\exp(-\eta \mathbb{E}_{W|N}[\ln W|N])|N < n]F(n)}{\mathbb{E}_N[\exp(-\eta \mathbb{E}_{W|N}[\ln W|N])]}.
\end{aligned}$$

Using the notation $\mu_N = \mathbb{E}[\ln N]$, $\sigma_N = \mathrm{Var}[\ln N]$, and $\mu_W = \mathbb{E}[\ln W]$, we can use the fact that

when $\ln W$ and $\ln N$ are jointly normal, then $\mathbb{E}_{W|N}[\ln W|N] = \mu_W + (1-\beta)(\ln N - \mu_N)$:[17]

$$\begin{aligned} G(n) &= \frac{\mathbb{E}_N[\exp(-\eta(\mu_W + (1-\beta)(\ln N - \mu_N)))|N < n]F(n)}{\mathbb{E}_N[\exp(-\eta(\mu_W + (1-\beta)(\ln N - \mu_N)))]} \\ &= \frac{\exp(-\eta\mu_W + \eta(1-\beta)\mu_N)\mathbb{E}_N[\exp(-\eta(1-\beta)\ln N)|N < n]F(n)}{\exp(-\eta\mu_W + \eta(1-\beta)\mu_N)\mathbb{E}_N[\exp(-\eta(1-\beta)\ln N)]} \\ &= \frac{\mathbb{E}_N[\exp(-\eta(1-\beta)\ln N)|N < n]F(n)}{\mathbb{E}_N[\exp(-\eta(1-\beta)\ln N)]}. \end{aligned}$$

The expectation in the numerator is the mean of a truncated lognormal distribution. We can use the fact that if Y is lognormally distributed with log mean μ_Y and log variance σ_Y, then[18]

$$\mathbb{E}[Y^r|Y < y] = \mathbb{E}[Y^r]\frac{\Phi\left(-r\sqrt{\sigma_Y} + (\ln y - \mu_Y)/\sqrt{\sigma_Y}\right)}{\Phi\left((\ln y - \mu_Y)/\sqrt{\sigma_Y}\right)}.$$

Thus letting $Y = N$ and $r = -\eta(1-\beta)$, we can write

$$\begin{aligned} G(n) &= \frac{\mathbb{E}_N[\exp(-\eta(1-\beta)\ln N)]F(n)}{\mathbb{E}_N[\exp(-\eta(1-\beta)\ln N)]} \cdot \frac{\Phi\left(\eta(1-\beta)\sqrt{\sigma_N} + (\ln n - \mu_N)/\sqrt{\sigma_N}\right)}{\Phi\left((\ln n - \mu_N)/\sqrt{\sigma_N}\right)} \\ &= \frac{\Phi\left((\ln n + \eta(1-\beta)\sigma_N - \mu_N)/\sqrt{\sigma_N}\right)}{\Phi\left((\ln n - \mu_N)/\sqrt{\sigma_N}\right)}F(n). \end{aligned}$$

Since n is lognormally distributed with log mean μ_N and log variance σ_N, the denominator is equal to $F(n)$, so we have

$$G(n) = \Phi\left((\ln n + \eta(1-\beta)\sigma_N - \mu_N)/\sqrt{\sigma_N}\right) = \int_0^{n\exp(\eta(1-\beta)\sigma_N)} f(n)dn.$$

Substituting this into the first-order condition for optimal marginal tax rates in (5) yields the result

[17] Here we use the fact that

$$\frac{\text{Cov}(\ln w, \ln n)}{\text{Var}(\ln n)} = 1 - \frac{\text{Cov}(\ln \theta, \ln n)}{\text{Var}(\ln n)} = 1 - \beta.$$

This can be found by rearranging the identity $\text{Var}(\ln w) = \text{Var}(\ln n) + \text{Var}(\ln \theta) - 2\text{Cov}(\ln \theta, \ln n)$ to get

$$\frac{1}{2} + \frac{\text{Var}(\ln \theta) - \text{Var}(\ln w)}{2\text{Var}(\ln n)} = \frac{\text{Cov}(\ln w, \ln n)}{\text{Var}(\ln n)} = \beta,$$

and thus

$$1 - \beta = \frac{1}{2} + \frac{\text{Var}(\ln w) - \text{Var}(\ln \theta)}{2\text{Var}(\ln n)},$$

and rearranging the identity $\text{Var}(\ln \theta) = \text{Var}(\ln n) + \text{Var}(\ln w) - 2\text{Cov}(\ln w, \ln n)$ to get

$$\frac{\text{Cov}(\ln w, \ln n)}{\text{Var}(\ln n)} = \frac{1}{2} + \frac{\text{Var}(\ln w) - \text{Var}(\ln \theta)}{2\text{Var}(\ln n)}.$$

[18] See Larry Lee and R.G. Krutchkoff (1980).

in the proposition,

$$\frac{T'(y(n))}{1-T'(y(n))} = \frac{1+1/\varepsilon}{nf(n)}\left(\int_0^{n\exp(\eta(1-\beta)\sigma_N)} f(n)dn - \int_0^{n} f(n)dn\right).$$

3.5 Robustness of results on redistribution and β

Here we demonstrate that the qualitative nature of the results in Section 2.1—in which marginal tax rates decrease as β rises—are robust to using a non-isoelastic functional form for the welfare weights $b(w)$ and to the inclusion of income effects.

In both cases, we first compute the optimal tax code under the conventional assumption that unified type n is equal to ability w (i.e., that $\beta = 0$). This computation provides a mapping between ability w and marginal social welfare weights, $b(w)$. We let $\ln(\theta) = \mu_\theta + \beta \ln(n)$, which satisfies the relationship in (17) with $\mu_\theta = -\beta\mu_N - \beta^2\sigma_N$ satisfying the normalization that the population average of θ is equal to one. We can then compute $w = n/\theta$, and we use the already-computed $b(w)$ to assign marginal social welfare weights to the alternative, more compressed schedule of abilities w.

The results of the first robustness check are shown in Figure 3. Rather than using an isoelastic shape for $b(w)$, under this specification social welfare is the integral over the logarithmic transformation of individual utility (like the "Type I" utility function in Saez 2001) in the case where $\beta = 0$. The tax schedule is generally less redistributive under this specification because redistribution compresses the distribution of utility (and thus the distribution of welfare weights), endogenously reducing the motivation to redistribute, whereas type-specific weights are held fixed in the baseline specification. The qualitative result that marginal tax rates are uniformly lower for higher values of β remains present.

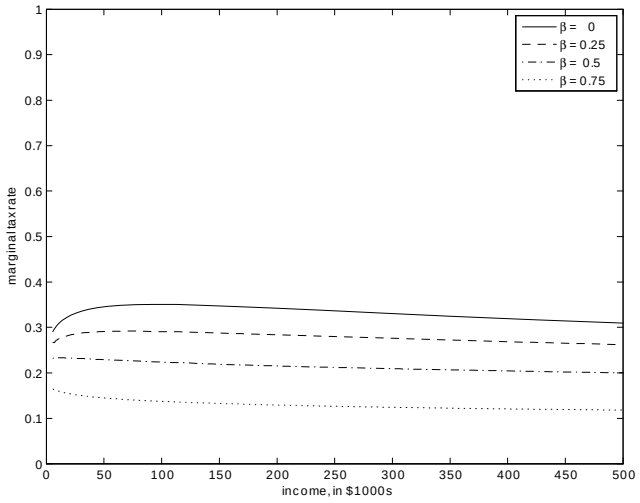

Figure 3: Optimal marginal tax rate schedules when the welfare weights $b(w)$ are set equal to those that arise by maximization the logarithmic transformation of individual utility in the case of homogeneous preferences.

The results of the second robustness check are shown in Figure 4, for which we use the "Type II" utility function from Saez 2001:

$$u(c, y, n) = \log(c) - \log\left(1 + \frac{(y/n)^{1+1/\varepsilon}}{1 + 1/\varepsilon}\right).$$

We apply the same type-specific isoelastic weights that were used in the baseline specification to this case. Note that in this case the welfare gains of redistribution—quantified by the social marginal value of consumption—are driven by both decreasing welfare weights and decreasing individual marginal utility of consumption, and so tax rates are not zero even when $\beta = 1$. Alternatively, one could specify a form of preference neutrality that took into account the functional form of individual utility. To see such an approach, see the working paper version of this paper, Lockwood and Weinzierl (2012).

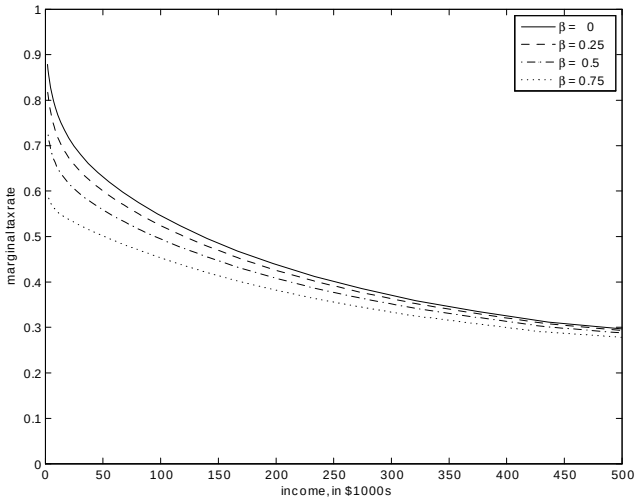

Figure 4: Optimal marginal tax rate schedules
when income effects are present.

Again, the qualitative result that marginal tax rates are uniformly lower for higher values of β obtains under this alternative.

References

[1] Alesina, Alberto and George-Marios Angeletos (2005). "Fairness and Redistribution," *American Economic Review*, 95(4), (September).

[2] Beaudry, Paul, Charles Blackorby, and Dezso Szalay (2009). "Taxes and Employment Subsidies in Optimal Redistribution Programs," *American Economic Review*, 99(1), pp. 216-242.

[3] Boadway, Robin, Maurice Marchand, Pierre Pestieau, and Maria del Mar Racionero (2002). "Optimal Redistribution with Heterogeneous Preferences for Leisure," *Journal of Public Economic Theory*, 4(4), pp. 475-498.

[4] Brett, Craig and John A. Weymark (2003). "Financing Education using Optimal Redistributive Taxation," *Journal of Public Economics.*

[5] Byrd, R. H.; J. Nocedal; and R. A. Waltz, "KNITRO: An Integrated Package for Nonlinear Optimization" in Large-Scale Nonlinear Optimization, G. di Pillo and M. Roma, eds, pp. 35-59 (2006), Springer-Verlag.

[6] Chetty, Raj (2012). "Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," Econometrica, Econometric Society, vol. 80(3), pages 969-1018.

[7] Choné, Philippe and Guy Laroque (2010). "Negative Marginal Tax Rates and Heterogeneity", *American Economic Review 100,* pp. 2532-2547.

[8] Diamond, Peter (1998). "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates", *American Economic Review*, 88(1) pp. 83-95.

[9] Diamond, Peter and Johannes Spinnewijn (2011) "Capital Income Taxes with Heterogeneous Discount Rates," *American Economic Journal: Economic Policy* 3(4), November.

[10] Feldstein, Martin and Marian Vaillard Wrobel (1998). "Can State Taxes Redistribute Income?" *Journal of Public Economics.*

[11] Fleurbaey, Marc and Francois Maniquet (2004). "Compensation and Responsibility," working paper from 2004 used. Forthcoming in K.J. Arrow, A.K. Sen and K. Suzumura (eds) *Handbook of Social Choice and Welfare.*

[12] Fleurbaey, Marc and Francois Maniquet (2006). "Fair Income Tax," *Review of Economic Studies*, 73, pp. 55-83.

[13] Friedman, Milton (1962). *Capitalism and Freedom.* Chicago: University of Chicago Press.

[14] Golosov, Mikhael, Maxim Troshkin, Aleh Tsyvinski, and Matthew Weinzierl (2013). "Preference Heterogeneity and Optimal Capital Income Taxation," *Journal of Public Economics.*

[15] Judd, Kenneth and Che-Lin Su (2006). "Optimal Income Taxation with Multidimensional Taxpayer Types," Working Paper, earliest available draft April, 2006; more recent draft November 22, 2006.

[16] Kahneman, Daniel (2011). *Thinking, Fast and Slow.* New York: Farrar, Straus, and Giroux (see page 401).

[17] Kaplow, Louis (2008). "Optimal Policy with Heterogeneous Preferences," *The B.E. Journal of Economic Analysis & Policy*, 8(1) (Advances), Article 40.

[18] Kocherlakota, Narayana and Christopher Phelan (2009). "On the robustness of laissez-faire," *Journal of Economic Theory* 144, pp. 2372-2387+

[19] Lee, Larry, and R.G. Krutchkoff (1980), "Mean and Variance of Partially-Truncated Distributions," Biometrics, Vol. 36, No. 3, pp. 531-536

[20] Lockwood, Benjamin B. and Matthew Weinzierl (2012). "De Gustibus non est Taxandum: Heterogeneity in Preferences and Optimal Redistribution," NBER Working Paper 17784.

[21] Mirrlees, James A. (1971). "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies.*

[22] Mirrlees, James A. (1976). "Optimal Tax Theory: A Synthesis," *Journal of Public Economics.*

[23] Mirrlees, James A. (1986) "The Theory of Optimal Taxation", in *Handbook of Mathematical Economics,* edited by K.J. Arrow and M.D. Intrilligator, vol. III, ch. 24, 1198-1249.

[24] Nozick, Robert (1974). *Anarchy, State, and Utopia.* New York: Basic Books.

[25] Pigou, Arthur C.(1928). *A Study in Public Finance.* Macmillan. (page 58, 1960 reprint).

[26] Rothschild, Casey and Florian Scheuer (2013). "Redistributive Taxation in the Roy Model," *Quarterly Journal of Economics.*

[27] Saez, Emmanuel (2001). "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies,* 68.

[28] Salanié, Bernard (2011). *The Economics of Taxation.* MIT.

[29] Sandmo, Agnar (1993). "Optimal Redistribution when Tastes Differ," *Finanzarchiv* 50(2).

[30] Tachibanaki, Toshiaki (2005), Confronting Income Inequality in Japan: A Comparative Analysis of Causes, Consequences, and Reform, MIT Press.

[31] Tarkiainen, Ritva and Matti Tuomala (2007). "On Optimal Income Taxation with Heterogeneous Work Preferences," International Journal of Economic Theory 3.1, pp. 35-46.

[32] Tuomala, Matti (1990). *Optimal Income Tax and Redistribution*, Oxford University Press.

[33] Vickrey, William (1945). "Measuring Marginal Utility by Reactions to Risk," *Econometrica*, 13(4), (October), pp. 319-333.

[34] Wang, Chen and Koen Caminada (2011), 'Disentangling income inequality and the redistributive effect of social transfers and taxes in 36 LIS countries', Leiden Department of Economics Research Memorandum #2011.02, and www.hsz.leidenuniv.nl

[35] Werning, Iván (2007). "Optimal Fiscal Policy with Redistribution," Quarterly Journal of Economics.