



European
University
Institute

MAX WEBER
PROGRAMME
FOR
POSTDOCTORAL
STUDIES

WORKING PAPERS

MWP 2013/22
Max Weber Programme

Efficiency, Equity, and Optimal Income Taxation

Charles Brendon

European University Institute
Max Weber Programme

Efficiency, Equity, and Optimal Income Taxation

Charles Brendon

EUI Working Paper **MWP** 2013/22

This text may be downloaded for personal research purposes only. Any additional reproduction for other purposes, whether in hard copy or electronically, requires the consent of the author(s), editor(s). If cited or quoted, reference should be made to the full name of the author(s), editor(s), the title, the working paper or other series, the year, and the publisher.

ISSN 1830-7728

© Charles Brendon, 2013

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

Abstract

Social insurance schemes must resolve a trade-off between competing efficiency and equity considerations. Yet there are few direct statements of this trade-off that could be used for practical policymaking. To this end, this paper re-assesses optimal tax policy in the celebrated Mirrlees (1971) model. It provides a new, intuitive characterisation of the optimum, relating two cost terms that are directly interpretable as the marginal costs of inefficiency and of inequality respectively. An empirical exercise then shows how our characterisation can be used to determine whether existing tax systems give too much weight to efficiency or to equity considerations, relative to a benchmark constrained optimum. Based on earnings, consumption and tax data from 2008 we show that social insurance policy in the US systematically gives insufficient weight to equality considerations when conventional assumptions are made about individual preference structures and the policymaker is utilitarian.

Keywords

Optimal income tax, Mirrlees model, Efficiency-equity trade-offs, Pareto efficient taxation, Primal approach

JEL codes: D63, D82, H21, H24

I would like to thank Árpád Ábrahám, Piero Gottardi, Ramon Marimon and Evi Pappa for detailed discussions and comments on this work. All errors are mine.

Charles Brendon

Max Weber Fellow 2012-2013 and 2013-2014

Email: charles.brendon@eui.eu.

1 Introduction

It has long been argued that social insurance schemes must resolve a trade-off between ‘efficiency’ and ‘equity’. Policy intervention is generally needed if substantial variation in welfare is to be avoided across members of the same society, but the greater the degree of intervention the more likely it is that productive behaviour will be discouraged.¹ This trade-off is central to income tax policy, where the key issue is whether the distortionary impact of raising taxes offsets the benefits of having more resources to redistribute. A natural question one might ask, therefore, is whether real-world tax systems do a good job in managing these competing concerns.

An important framework for analysing this question is the celebrated model of Mirrlees (1971), in which the efficiency-equity trade-off derives more specifically from an informational asymmetry. Individuals are assumed to differ in their underlying productivity levels, but productivity itself cannot be observed – only income. The government is concerned to see an even distribution of consumption across the population, but must always ensure that more productive agents are given sufficient incentives to ‘reveal’ their status, and produce more output. This model is notoriously complex, and a number of equivalent analytical characterisations of its optimum are possible. By far the most celebrated has been that of Saez (2001), who provided a solution in terms of a limited number of interpretable, and potentially estimable, objects – notably compensated and uncompensated labour supply elasticities, the empirical earnings distribution, and social preferences. In improving the accessibility of the Mirrlees model, this work provided a key foundation for a large applied literature.²

This paper provides a new, alternative characterisation that captures the efficiency-equity trade-off simply and intuitively, and in a far more direct manner than Saez (2001). The aim in doing so is to give a new dimension to the applied policy debate, allowing observed social insurance systems to be assessed relative to this characterisation – that is, assessed directly in terms of the balance that they are striking between the two competing concerns. Specifically, we define and motivate two cost terms that are directly interpretable as the marginal costs of inequity and inefficiency respectively. We show that an optimum can then be described solely in terms of these cost terms, plus the exogenous distribution of productivity types and derivatives of the social welfare criterion. Our key optimality result states that the marginal cost of introducing greater productive

¹A form of this argument can be traced at least to Smith’s *Wealth of Nations* (Book V, Ch 2), where four maxims for a desirable tax system are presented. The first captures a contemporary notion of equity: “The subjects of every state ought to contribute towards the support of the government, as nearly as possible, in proportion to their respective abilities.”

The fourth maxim, meanwhile, captured the need to minimise productive losses from tax distortions: “Every tax ought to be so contrived as both to take out and to keep out of the pockets of the people as little as possible over and above what it brings into the public treasury of the state.” The other two maxims related to the timing of taxation and the predictability of one’s liabilities – issues that have subsequently faded in importance.

²See in particular the recent survey by Piketty and Saez (2013). Mirrlees (2011) is the clearest example of the lessons from this literature being directly incorporated into the policy debate.

distortions at a given point in the income distribution should exactly equal the marginal benefits from reducing the costly ‘information rents’ that more productive agents must be given in order to incentivise their production. This latter cost follows from the policymaker’s desire to provide welfare in the least-cost manner across all agents, which generally means delivering resources to those who value them most – which in turn usually implies a desire for equality, by the logic of diminishing marginal utility.

Naturally it is very unlikely that the trade-off will be perfectly struck by any real-world tax system – at least under standard social and individual preference assumptions. But one of the advantages of our analysis is that it can indicate the manner in which there is a departure from optimality at different points in observed earnings distributions – of the form: ‘Insufficient concern given to efficiency for medium earners’, for instance. It also allows for a direct comparison across different earnings levels of the marginal benefits from improving the trade-off; and since the cost terms that we define are measured in units of real output, these marginal benefits can be stated in the same units, easing their interpretation. This seems particularly useful for applied policy purposes, as it can show precisely what sorts of tax reforms would yield the greatest returns, and provide a quantification of the extra benefits they might yield at the margin. In this way we can help inform ongoing debates about the appropriateness of the efficiency burden of taxation.

Technically our characterisation is based on viewing the Mirrlees problem through the lens of mechanism design theory, with the government choosing direct allocations of consumption and work effort for all agents as functions of their revealed productivities, subject to an incentive compatibility requirement.³ This differs analytically from the approach of Saez (2001), which was to consider choice across alternative tax schedules subject to the known responses of individuals to changes in their post-tax wage rates. This may seem an obscure technicality, but the difference is substantive for applied purposes because the ‘primal’ approach taken here delivers optimality conditions that are expressed in terms of the arguments of the direct utility function, whereas Saez’s ‘dual’ method involves objects more closely related to the indirect utility function – notably Hicksian and Marshallian elasticities.⁴ Hence this paper’s results are particularly useful when the analyst prefers to impose a structural, parametric form on preferences, rather than applying estimates of Hicksian and Marshallian elasticities directly. We discuss below why this might be desirable in some settings.

Our results are likely to be of most use for applied work, but they also shine new theoretical light on two particular issues of practical importance. The first

³The interpretation of the Mirrlees model as a mechanism design problem is common in the optimal contracting literature – see, for instance, the textbook discussions by Laffont and Martimort (2002), Chapter 2, or Bolton and Dewatripont (2005), Chapter 2. Yet analytical characterisations in that literature have been restricted to the relatively simple case of quasi-linear preferences. In this paper we allow for general preference structures.

⁴The terms ‘primal’ and ‘dual’ are applied by analogy to their usage in the Ramsey taxation literature, where the same analytical distinction exists. See the discussion in Section 3.4.

is the character of Pareto efficient income taxation – a topic first considered by Werning (2007). We show below how our two cost terms can be used to infer a set of restrictions that are both necessary and sufficient for a tax system to be Pareto efficient, and that must therefore be satisfied by any optimal scheme devised by a social planner whose objective criterion is strictly increasing in the utility levels of all agents – a fairly innocuous restriction. In this regard we generalise Werning’s earlier results, which obtained only for a simplified version of the Mirrlees model. Perhaps most surprisingly, we show as part of this analysis that it is generally Pareto inefficient for marginal income tax rates to jump upwards by discrete amounts. Such jumps are common features of means-tested benefit schemes that see the absolute value of benefits withdrawn as incomes grow above some threshold level – examples being the Earned Income Tax Credit in the United States and Working Tax Credit in the United Kingdom. Their Pareto inefficiency suggests they should be uncontroversial candidates for reform.

The second theoretical contribution is to provide clarification on the character of optimal taxes at the top of the income distribution, when preferences take the commonly-used isoelastic, separable form. In particular, we show that if the distribution of productivity types is consistent with stylised facts on the upper tail of the earnings distribution – as presented, for instance, by Diamond and Saez (2011) – then for a surprisingly broad subset of the parameter space the top rate of income tax should approach 100 per cent as incomes grow. That is, for relatively ‘mild’ preference calibrations we obtain the complete opposite to the traditional ‘zero distortion at the top’ result that featured in early analyses of the Mirrlees model. This result relies heavily on the particular cardinalisation of preferences used – less curvature in social or individual welfare functions will always undermine it – but it is surprising that it should go through for such conventional parameterisations as it does.⁵

Following these results, an applied section then provides an illustrative attempt to quantify the way efficiency-equity trade-offs are managed in practice, of the sort outlined above. Using data from the 2009 wave of the PSID survey we estimate a distribution of productivity types consistent with observed cross-sectional income and consumption patterns for the US economy in 2008. This distribution satisfies the requirement that each agent’s observed consumption-output choices in the dataset must be optimal given (a) the marginal tax rate that they are likely to have faced, and (b) an assumed, parametric form for the utility function. Given this type distribution, for the same utility structure we can then infer the marginal benefits to changing the income tax schedule at different points, and in particular to changing the way in which efficiency and equity considerations are balanced against one another.

The main qualitative result of this exercise is that the US tax system appears systematically to introduce too few productive distortions relative to the information rents that it leaves in place. Put differently, equity concerns are

⁵A Frisch elasticity of labour supply and elasticity of intertemporal substitution both equal to 0.5, for instance, will suffice.

under-valued relative to the Mirrleesian optimum. This result is surprisingly general: it is true at all points along the income distribution for realistic calibrations of the isoelastic, separable preference structure that we assume. The direction of the bias in observed policy only reverses for relatively extreme parameterisations: a Frisch elasticity of labour supply above 1, for instance, or a coefficient of relative risk aversion less than 0.65. Quantitatively, the benchmark results suggest that a utilitarian policymaker could obtain the same level of social welfare whilst raising an additional quantity of tax revenue of up to \$3500 per worker for every unit by which inequity (in utility units) is reduced.

These results are clearly contentious, and before proceeding it is worth reminding the reader of some of the many factors that the Mirrlees model omits, and which may be distorting our conclusions. First, and most importantly, the model is static. It assumes that all post-tax income will be consumed immediately, and consumption in excess of post-tax income is impossible. Agents with a low earnings potential have no means to provide themselves with insurance – whether through accumulated savings or more explicit schemes – independently of the tax system. Clearly this is not true in practice, but since the model’s optimality conditions are derived under the assumption that a tax system links gross income directly to *consumption* (not disposable income), a meaningful comparison with real-world tax systems can only be obtained under the assumption that they do likewise. If formal and informal insurance schemes mean that income variations across individuals do not translate into large consumption differentials then the benefits from greater post-tax income equality may not be as great as we find.

A second limitation of the model is that it allows only one type of behavioural response to changes in the marginal tax rate: labour supply changes at the intensive margin. There is no scope for tax avoidance, for migration, or for significant extensive-margin decisions. All of these omissions are likely to understate the costs from introducing higher tax distortions. This is particularly the case because decisions to leave the labour market entirely, whether through migration or simple retirement, result in large, discrete losses of revenue to the government, different from intensive-margin reductions in labour supply – so that relatively large effects can be driven by a small number of agents. Adjusting our solution procedure for such effects seems a fruitful avenue for future work; it would allow in particular for a consideration of optimal transfer policy at the bottom of the income distribution – a problem highlighted by Saez (2002).

Finally, any optimal tax problem with heterogeneous agents must assume an objective criterion that allows changes in different agents’ welfare to be compared against one another. The results that we derive rely, for the most part, on a utilitarian criterion, aggregating utility functions that are themselves concave, and this concavity property is certainly necessary for the costs of inequality to be considered important. If the marginal social value of providing an individual with additional consumable resources is not diminishing in the level of welfare that the individual already enjoys then there are no marginal costs from allowing higher-type agents to consume more: this is only costly to a utilitarian because it implies a wasteful allocation under diminishing returns. What is important

is how far diminishing marginal utility must be exploited for the costs of inequality to dominate. On this matter, the parametric forms for utility that we assume are certainly not unusual in similar branches of the literature – though how much concavity is ‘too much’ is not an easily answered question, and for many readers may not be separable from the distributional results implied.

The rest of the paper proceeds as follows. Section 2 outlines the basic form of the static Mirrlees problem that we study. Section 3 presents our main characterisation result when a specific, generalised utilitarian welfare criterion is applied, and relates it to a weaker set of restrictions that follow simply from the Pareto criterion. We provide a brief discussion linking our analytical approach to the ‘primal’ method familiar from Ramsey tax theory. Section 4 provides intuition for the general result by applying it to the well-known isoelastic, separable preference structure, deriving novel results for the top rate of income tax in this case. Section 5 contains our main empirical exercise, testing the efficiency-equity balance struck by the US tax system. Section 6 concludes.

2 Model setup

2.1 Preferences and technology

We use a variant of the model set out in Mirrlees (1971). The economy is populated by a continuum of individuals indexed by their productivity type $\theta \in \Theta \subset \mathbb{R}$. The type set Θ is closed and has a finite lower bound denoted $\underline{\theta}$, but is possibly unbounded above: $\Theta = [\underline{\theta}, \bar{\theta}]$ or $\Theta = [\underline{\theta}, \infty)$. Agents derive utility from consumption and disutility from production, in a manner that depends on θ . Their utility function is denoted $u : \mathbb{R}_+^2 \times \Theta \rightarrow \mathbb{R}$, where u is assumed to be C^2 in all three of its arguments (respectively consumption, output and type). Demand for both consumption and leisure is assumed to be normal, where leisure can be understood as the negative of output. Types are assumed to be private information to individuals, with only output publicly observable; this will provide the government with a non-trivial screening problem in selecting among possible allocations.

To impose structure on the problem we endow u with the usual single crossing property:

Assumption 1 *For any distinct pair of allocations (c', y') and (c'', y'') such that $(c', y') < (c'', y'')$ (in the product order sense) and $\theta' < \theta''$, if $u(c'', y''; \theta') \geq u(c', y'; \theta')$ then $u(c'', y''; \theta'') > u(c', y'; \theta'')$.*

Geometrically this condition is implied by the fact that indifference curves in consumption-output space are ‘flattening’ in θ , in the sense:

$$\frac{d}{d\theta} \left(-\frac{u_y(c, y; \theta)}{u_c(c, y; \theta)} \right) < 0 \quad (1)$$

Single crossing is an important restriction: it will provide justification for the common practice in the mechanism design literature of relaxing the constraint set implied by incentive compatibility when determining a constrained-optimal

allocation. In the appendix we show that it is implied if all individuals share common preferences over consumption and labour supply, with labour supply then being converted into output in a manner that in turn depends on θ . Preference homogeneity of this form remains fairly contentious in the literature – criticised, for instance, by Diamond and Saez (2011) for being too strong a restriction. But at this stage it is an indispensable simplification for deriving our main results. We hope that these in turn may later prove of use to the ongoing project of understanding optimal taxes under *arbitrary* preference structures.

2.2 Government problem

2.2.1 Objective

We define an *allocation* as a pair of functions $c : \Theta \rightarrow \mathbb{R}_+$ and $y : \Theta \rightarrow \mathbb{R}_+$ specifying consumption and output levels for each type in Θ . The government’s problem will be to choose from a set of possible allocations in order to maximise a generalised social welfare function, W , defined on the utility levels that obtain for the chosen allocation:

$$W := \int_{\theta \in \Theta} G(u(\theta), \theta) f(\theta) d\theta$$

where $f(\theta)$ is the density of types at θ and we use $u(\theta)$ as shorthand for $u(c(\theta), y(\theta); \theta)$. $G(u, \theta)$ is assumed to be weakly increasing in u for all θ . This general formulation nests three important possibilities:

1. Utilitarianism: $G(u(\theta), \theta) = u(\theta)$.
2. Symmetric inequality aversion: $G(u(\theta), \theta) = g(u(\theta))$, for some concave, increasing function $g : \mathbb{R} \rightarrow \mathbb{R}$
3. Pareto weights: $G(u(\theta), \theta) = \alpha(\theta) u(\theta)$ for some $\alpha : \Theta \rightarrow \mathbb{R}_{++}$.

Most presentations of the model use the second of these, following the original treatment by Mirrlees (1971). Utilitarianism is a simpler approach to take, but is often avoided because it undermines any redistributive motive when agents’ preferences are restricted to be quasi-linear in consumption – a case that Diamond (1998) showed to be particularly tractable. Werning (2007) considers the case in which Pareto efficiency is the sole consideration used to assess tax schedules. In general an allocation A Pareto-dominates an alternative allocation B if and only if W is (weakly) higher under A than B for all admissible choices of the function G . Any restrictions on the optimal tax schedule implied by Pareto efficiency alone are thus robust to the controversial question of the appropriate welfare metric – at least within the class of metrics that satisfy the Pareto criterion. This makes them of interest as a potential means for generating ‘consensus’ reforms. We will highlight one such reform in Section 3.3 below, which follows from generalising Werning’s results.

Notice that the objective W is ‘welfarist’ in the traditional sense used in the social choice literature: it maximises a known function of individual-level utilities alone. A recent critique of this approach by Mankiw and Weizierl (2010) and Weizierl (2012) has claimed that it does not account for observed policy decisions – notably the absence of ‘tagging’ that would allow tax liabilities to vary on the basis of observable characteristics, such as height, that correlate with individuals’ earnings potentials. Saez and Stantcheva (2013) seek to accommodate this critique by allowing the marginal social value of providing income to a given individual itself to be endogenous to the tax system chosen – on the grounds that certain forms of redistribution might be seen as rewarding the ‘deserving’ more than others.⁶ To keep the problem simple this generalisation is not admitted here, but it may be useful in future to explore its incorporation into the characterisation that we set out.

2.2.2 Constraints

The government seeks to maximise W subject to two (sets of) constraints, which together will define the set of incentive-feasible allocations. The first is a restriction on resources:

$$\int_{\theta \in \Theta} [c(\theta) - y(\theta)] f(\theta) d\theta \leq -R \quad (2)$$

where R is an exogenous revenue requirement on the part of the government. An allocation that satisfies (2) will be called *feasible*.

The second requirement is a restriction on incentive compatibility. Since the government can only observe output, not types, it will have to satisfy the restriction that no agent can obtain strictly higher utility by mimicking another at the chosen allocation. The setting is one in which the revelation principle is well known to hold, and so we lose no generality by focusing exclusively on direct revelation mechanisms. If $(c(\sigma), y(\sigma))$ is the allocation of an agent who reports $\sigma \in \Theta$, incentive compatibility then requires that truthful reporting should be optimal:

$$u(c(\theta), y(\theta); \theta) \geq u(c(\sigma), y(\sigma); \theta) \quad \forall (\theta, \sigma) \in \Theta^2 \quad (3)$$

A feasible allocation that satisfies (3) is *incentive feasible*. The policymaker’s problem is to maximise W on the set of incentive-feasible allocations. An allocation that solves this problem is called a *constrained-optimal allocation*.

Condition (3) provides a continuum of constraints at each point in Θ . Such high dimensionality is unmanageable by direct means, and so we instead exploit the single crossing condition to re-cast the constrained choice problem using a technique familiar from the optimal contracting literature.⁷ We prove the following in the appendix:

⁶This marginal value takes a central role in optimality statements for tax rates derived under the dual approach. See Piketty and Saez (2013) for a general discussion and presentation of these formulae.

⁷See, for instance, Bolton and Dewatripont, Chapter 2.

Proposition 1 *An allocation is incentive feasible if and only if (a) the schedules $c(\theta)$ and $y(\theta)$ are weakly increasing in θ , and (b) it satisfies:*

$$\frac{d}{d\theta} [u(c(\theta), y(\theta); \theta)] = \frac{\partial}{\partial \theta} [u(c(\sigma), y(\sigma); \theta)]|_{\sigma=\theta} \quad (4)$$

where the derivatives here are replaced by their right- and left-hand variants at $\underline{\theta}$ and $\bar{\theta}$ respectively.

This envelope condition is a common feature in screening models. It accounts for the ‘information rents’ that higher types are able to enjoy as a consequence of their privileged informational position. As an agent’s true productivity is increased at the margin, any incentive-compatible scheme must provide enough extra utility under truthful reporting to compensate the agent for the additional welfare he or she can now obtain at a *given* report. Milgrom and Segal (2002) demonstrate the general applicability of the integrated version of this condition:

$$u(\theta) = u(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} u_{\theta}(\tilde{\theta}) d\tilde{\theta} \quad (5)$$

In what follows we will work with condition (5) in place of (3). By Proposition 1 a feasible allocation that satisfies (5) *and is increasing* must be incentive feasible. But increasingness will prove easier to check *ex post*, after finding the best feasible allocation in the set that satisfies (5) alone. Thus we will study the *relaxed problem* of maximising W subject to (2) and (5) alone. A feasible allocation that satisfies (5) – but is not necessarily increasing – we call *relaxed incentive feasible*. An allocation that maximises W on the set of relaxed incentive-feasible allocations is *constrained-optimal for the relaxed problem*.

2.2.3 Equivalent representations

An important feature of this framework is that the constraint set of the problem relies only on the ordinal properties of the utility function. In particular, we could always replace the general incentive compatibility restriction (3) with the following:

$$V(u(c(\theta), y(\theta); \theta)) \geq V(u(c(\sigma), y(\sigma); \theta)) \quad \forall (\theta, \sigma) \in \Theta^2 \quad (6)$$

for any monotonically increasing function $V: \mathbb{R} \rightarrow \mathbb{R}$. If we define the resulting utility function $v := V(u(\cdot))$ it is clear that this v inherits the basic structure of u , notably single crossing. The constrained-optimal allocation for the problem of maximising W on the set of incentive-compatible allocations must therefore be identical to the constrained-optimal allocation for the problem of maximising \widetilde{W} on the set of allocations that satisfy (6) and the resource constraint (2), where \widetilde{W} is defined by:

$$\widetilde{W} := \int_{\theta \in \Theta} \widetilde{G}(v(\theta), \theta) f(\theta) d\theta \quad (7)$$

and:

$$\tilde{G}(v(\theta), \theta) := G(V^{-1}(v(\theta)), \theta)$$

That is, the social objective must be adjusted to incorporate the inverse of the V transformation, but once this change is made the full problem becomes equivalent to our initial representation. What *does* change is the precise specification of the relaxed problem. In particular the derivative of v satisfies:

$$v_\theta(\theta) = V_u(u(\theta)) u_\theta(\theta) \quad (8)$$

Thus the equivalent of the envelope condition (5) is:

$$V(u(\theta)) = V(u(\underline{\theta})) + \int_{\underline{\theta}}^{\theta} V_u(u(\tilde{\theta})) u_\theta(\tilde{\theta}) d\tilde{\theta} \quad (9)$$

This is not directly equivalent to (5) except in the trivial case when V is a linear function.⁸ Yet if an allocation maximises \tilde{W} subject to (2) and (9) and it satisfies increasingness of $c(\theta)$ and $y(\theta)$ in θ then, by identical logic to before, this allocation must solve the problem of maximising \tilde{W} on the set of incentive-feasible allocations characterised by (2) and (6). But then it must also solve the original problem of maximising W on the set of incentive-feasible allocations characterised by (2) and (3).

This is important for what follows because we will introduce into the analysis objects that are defined directly by reference to the marginal information rents u_θ . But these information rents themselves depend on a particular normalisation of the problem – that is, a particular choice for V . Some normalisations may yield cleaner representations than others – notably when ordinal preferences can be described by a utility function that is additively separable between consumption and labour supply. We exploit such transformations wherever possible.

3 Characterising the equity-efficiency trade-off

In this section we show how the solution to the primal problem can be characterised in a form that isolates the model's central efficiency-equity trade-off. To

⁸ Consider, for instance, preferences of the Greenwood, Hercowitz and Huffman (1988) form:

$$u(c, y; \theta) = \frac{[c - \omega(\frac{y}{\theta})]^{1-\sigma}}{1-\sigma}$$

and the transformation V given by:

$$V(u) = [(1-\sigma)u]^{1/(1-\sigma)}$$

Clearly the associated v satisfies:

$$v_\theta = \frac{y}{\theta^2} \omega' \left(\frac{y}{\theta} \right)$$

whereas the expression for u_θ is far more complicated:

$$u_\theta = \left[c - \omega \left(\frac{y}{\theta} \right) \right]^{-\sigma} \frac{y}{\theta^2} \omega' \left(\frac{y}{\theta} \right)$$

In particular v_θ is independent of c , whereas u_θ is not.

understand heuristically why this trade-off arises, consider the solution to the ‘first-best’ problem of maximising W on the set of feasible allocations alone – ignoring incentive compatibility. Assuming interiority, this can be fully characterised by the resource constraint (2) together with two first-order conditions:

$$u_c(\theta) + u_y(\theta) = 0 \quad \forall \theta \in \Theta \quad (10)$$

$$G_u(u(\theta'), \theta') \cdot u_c(\theta') = G_u(u(\theta''), \theta'') \cdot u_c(\theta'') \quad \forall (\theta', \theta'') \in \Theta^2 \quad (11)$$

The first of these is a productive efficiency condition at the level of individual agents. It equates the marginal rate of substitution between consumption and production to the marginal rate of transformation, which is 1. The second condition deals with the optimal allocation (under W) of resources *across* individuals in the economy. There can be no marginal benefits from additional redistribution at the optimum.

Suppose that $G(u, \theta)$ takes the form $g(u)$ for some weakly concave, increasing function g – that is, the social welfare criterion is anonymous, and it exhibits weak aversion to utility disparities. Then under the assumed preference restrictions it is well known that the first-best allocation must involve decreasing utility in type. This is because higher-type agents in general draw the same benefits from consumption as lower types, but are more effective producers. The latter means that the policymaker has an incentive to induce more hours of work from high types; but there is no corresponding reason to provide them with greater consumption. High productivity thus becomes a curse rather than a blessing.

Such an allocation is clearly not consistent with incentive compatibility. In particular, since $u_\theta > 0$ always holds, utility will have to be increasing in θ at an allocation that satisfies the envelope condition (5). Productive efficiency, as characterised by equation (10), does remain possible, but (11) cannot simultaneously obtain. More significantly, it may be desirable to break condition (10) and introduce inefficiencies at the individual level as a means to ensure a more desirable *cross-sectional* distribution of resources. This will be true in particular if productive inefficiencies can be used to reduce the value of the ‘information rents’ captured in (5), which grow at rate u_θ as type increases. A positive marginal income tax can achieve just this: by restricting the production levels of lower types it reduces the marginal benefits to being a higher type, i.e. u_θ , since these benefits follow from being able to produce the same quantity of output with less effort. The lower is the output level in question, the lower are the marginal benefits to being more productive. From here there emerges a trade-off between ‘efficiency’ and ‘equity’: distorting allocations is likely to incur a direct resource cost, even as it yields benefits from a more even distribution of utility across the population.

3.1 Two cost terms

To characterise this trade-off more formally we define two cost terms that will be used throughout the subsequent analysis to describe the optimal allocation. They are easiest to rationalise in terms of the envelope condition (5), since each

gives the marginal resource cost of varying one of this condition's components, either $u_\theta(\theta)$ or $u(\theta)$, holding the other constant. First, consider the marginal cost of distorting the allocation of type θ by an amount just sufficient to reduce $u_\theta(\theta)$ by a unit, whilst holding constant $u(\theta)$. We label this $DC(\theta)$ – the ‘distortion cost’. It will be given by:

$$DC(\theta) := \frac{u_c(\theta) + u_y(\theta)}{u_c(\theta)u_{y\theta}(\theta) - u_y(\theta)u_{c\theta}(\theta)} \quad (12)$$

Useful intuition for this object can be obtained by defining $\tau(\theta)$ as the implicit marginal income tax rate faced by type θ :

$$\tau(\theta) := 1 + \frac{u_y(\theta)}{u_c(\theta)} \quad (13)$$

We then have:

$$DC(\theta) = \frac{\tau(\theta)}{u_{y\theta}(\theta) + (1 - \tau(\theta))u_{c\theta}(\theta)} \quad (14)$$

Consider a marginal change in the allocation given to type θ that reduces this agent's output by one unit whilst holding constant their utility. The corresponding reduction in consumption must be $(1 - \tau(\theta))$ units, since this is the agent's marginal rate of substitution between consumption and output. Thus the policymaker loses $\tau(\theta)$ units of resources for every unit by which output falls. This accounts for the numerator in (14). Meanwhile for every unit decrease in output and $(1 - \tau(\theta))$ decrease in consumption, the value of u_θ will decrease by an amount $u_{y\theta}(\theta) + (1 - \tau(\theta))u_{c\theta}(\theta)$ – the term in the denominator. Thus the overall expression gives the marginal resource loss to the policymaker per unit by which information rents at θ are reduced.

The other cost term that will prove relevant is a particular measure of the marginal cost to the policymaker of providing a unit of utility to an agent of type θ . In general there is clearly an infinity of possible vectors in consumption-output space along which θ 's utility will be increasing: an increase in $c(\theta)$ by a unit at the margin can be coupled with a reduction in output by α units for any $\alpha > \frac{1}{1 - \tau(\theta)}$, for instance. We are interested in the cost of raising utility by moving along just one such vector: that which will hold constant the value of $u_\theta(\theta)$. Again, at this stage it is not directly apparent why this should warrant special attention, but heuristically one can see from (5) that a measure of the marginal cost of allowing utility differentials, holding constant information rents, may help in separating out competing concerns. The cost in question is denoted $MC(\theta)$, defined by:

$$MC(\theta) := \frac{u_{c\theta}(\theta) + u_{y\theta}(\theta)}{u_c(\theta)u_{y\theta}(\theta) - u_y(\theta)u_{c\theta}(\theta)} \quad (15)$$

To develop intuition regarding this object, first note that if utility is additively separable in consumption and output then $u_{c\theta} = 0$, and $MC(\theta)$ collapses to $u_c(\theta)^{-1}$ – the inverse marginal utility of consumption. Separability of this

strong form means that consumption utility is entirely type-independent, and thus u_θ must be unaffected by any changes to allocations that involve consumption alone. Indeed, it is not just the value to being a *marginally* higher type at a given allocation that remains invariant to consumption utility changes: the value of

$$u(c + \delta, y; \theta'') - u(c + \delta, y; \theta')$$

will be invariant to δ for all $(\theta', \theta'') \in \Theta^2$. This observation has been exploited to prove the celebrated ‘inverse Euler condition’ in dynamic versions of the model, as it allows for a class of perturbations to be constructed in that setting that respect global incentive compatibility.⁹

More generally, u_θ will remain constant provided that for every unit increase in the consumption allocation of type θ there is an increase in that agent’s output allocation of $-\frac{u_{c\theta}}{u_{y\theta}}$ units: clearly this means

$$\frac{d}{dc}(u_\theta) = u_{c\theta} + u_{y\theta} \frac{dy}{dc} = 0$$

Using this insight we can rewrite $MC(\theta)$ as:

$$MC(\theta) := \frac{1 + \frac{u_{c\theta}(\theta)}{u_{y\theta}(\theta)}}{u_c(\theta) - u_y(\theta) \frac{u_{c\theta}(\theta)}{u_{y\theta}(\theta)}}$$

The numerator here can then be identified as the cost to the policymaker of increasing consumption by a unit, assuming that output is adjusted by $-\frac{u_{c\theta}}{u_{y\theta}}$ units simultaneously – increases in output being a resource benefit to the policymaker. The denominator is the marginal impact that this change has on the agent’s utility, so that the overall term is the marginal cost of utility provision that we seek.

The reason for defining these two cost terms is that it will be possible to express the entire set of necessary optimality conditions for the Mirrlees model in terms of them, plus the exogenous type distribution and the government welfare criterion. No additional arguments of the utility function, or elasticities, or Lagrange multipliers need feature.

3.1.1 Example: isoelastic, separable utility

To fix ideas it is useful to illustrate the form taken by our two cost objects when utility takes a specific functional form. One of the simplest cases arises when preferences are isoelastic and additively separable between consumption and labour supply:

$$u(c, y; \theta) = \frac{c^{1-\sigma} - 1}{1 - \sigma} - \frac{(ye^{-\theta})^{1+\frac{1}{\varepsilon}}}{1 + \frac{1}{\varepsilon}} \quad (16)$$

⁹See Golosov, Kocherlakota and Tsyvinski (2003) for a full discussion of the role of separability in the inverse Euler condition.

where θ here can be understood as the log of labour productivity, ε is the Frisch elasticity of labour supply, and σ is the coefficient of relative risk aversion. Separability gives $MC(\theta)$ a straightforward definition:

$$MC(\theta) = c(\theta)^\sigma \quad (17)$$

whilst – with some trivial manipulation – $DC(\theta)$ can be shown to satisfy:

$$DC(\theta) = \frac{\tau(\theta)}{1 - \tau(\theta)} \frac{\varepsilon}{1 + \varepsilon} c(\theta)^\sigma \quad (18)$$

Heuristically, the marginal cost of providing utility is the inverse of the marginal utility value of additional resources. If utility is being provided through consumption alone, which is the relevant vector to consider in the separable case, then this corresponds simply to the inverse marginal utility of consumption. As for the marginal cost of distorting allocations, $DC(\theta)$, this is increasing in the existing marginal tax rate, since reducing the output of an agent who is already paying high taxes is relatively costly to the public purse. The cost is also higher the higher is ε , the Frisch elasticity of labour supply. This is because a higher elasticity generally means a greater reduction in output will be induced for a given reduction in information rents, which raises the associated productive distortions. Finally, the term $c(\theta)^\sigma$ in the definition of $DC(\theta)$ follows from the utility scale being applied: $DC(\theta)$ is the marginal cost of reducing the marginal utility benefit from being a higher type, u_θ , by a unit. In general the lower is the marginal utility of consumption (i.e., the higher is $c(\theta)^\sigma$), the more resources will have to change to effect the desired change to u_θ – and thus the higher will be the distortion costs.

Clearly this is just one possible specification among many, but it gives useful intuition for the sorts of underlying factors that will be affecting the optimality trade-offs expressed in terms of $MC(\theta)$ and $DC(\theta)$ below.

3.2 An optimal trade-off

We now present our main characterisation result, which is novel to this paper, and provide a heuristic discussion of why it must hold. The full proof is algebraically intensive, and relegated to the appendix.

Proposition 2 *Any interior allocation that is constrained-optimal for the relaxed problem and satisfies the transversality restriction:*

$$\lim_{\theta \rightarrow \infty} DC(\theta) f(\theta) \leq 0$$

must satisfy the following condition for all $\theta' \in \Theta$:

$$DC(\theta') \cdot \frac{f(\theta')}{1 - F(\theta')} = E[MC(\theta) | \theta > \theta'] - \frac{E[G_u(u(\theta)) | \theta > \theta']}{E[G_u(u(\theta))]} \cdot E[MC(\theta)] \quad (19)$$

This result follows intuitively by considering the consequences of raising the marginal tax rate on an agent whose type is θ' , in order to pull down the information rents of higher types. The term on the left-hand side measures the cost of the extra productive distortions induced at the margin as a consequence: $DC(\theta')$ is the per-agent marginal loss in resources for the policymaker for every unit by which information rents are reduced at θ' , whilst the hazard rate term $\frac{f(\theta')}{1-F(\theta')}$ provides a measure of the number of agents who are of type θ' , relative to the number whose types are higher – that is, for whom information rents are reduced. The right-hand side can be read in two parts. The first term is the marginal quantity of resources that are gained by the policymaker when information rents above θ' can be reduced uniformly by a unit: by construction this is the expected value of $MC(\theta)$ above θ' . The second term corrects for the fact that these resources were not being completely wasted before: they were providing utility to agents above θ' , which is of value to any policymaker placing strictly positive weight on some or all of these agents' welfare in W . To offset the reduction to W it will be sufficient to provide a uniform quantity of utility across all agents, in an amount proportional to the relative weight given to agents above θ' under the planner's objective. Uniform utility provision will respect the incentive compatibility restrictions of the relaxed problem provided it is delivered in a manner that holds $DC(\theta)$ constant for all θ . The cost of providing it in this way is the expected value of $MC(\theta)$ across all $\theta \in \Theta$, interacted with the relative weight term $\frac{E[G_u(u(\theta))|\theta > \theta']}{E[G_u(u(\theta))]}$. Note that in the utilitarian case this weight term will equal one, whilst a Rawlsian 'min-max' objective corresponds to the limiting case in which it goes to zero.

The transversality condition deserves brief mention. It is akin to a 'no Ponzi games' restriction in dynamic savings problems, in the sense that it rules out perturbations that raise the utility of all agents above some type θ' if these perturbations cannot be approximated by changing the utility only of agents between θ' and some higher type $\theta'' > \theta'$, whilst keeping the utility of agents above θ'' constant, and then taking the limiting case as $\theta'' \rightarrow \infty$. Its inclusion is necessary because by imposing a sufficiently concave utility function to describe a given ordinal preference map it will often be possible to ensure that $MC(\theta) < 0$ holds for all θ sufficiently large – in which case raising high-types' utility levels by common increments would seem to guarantee a Pareto improvement. The negative marginal cost arises because it is possible to provide utility along a dimension such that output increases more than consumption, whilst holding u_θ constant. But such a scheme may only be possible if there is no upper type θ'' whose allocation must be distorted, at cost $DC(\theta'')$, to keep utility constant above θ'' – just as a household may be able to raise consumption in every time period only if it never reaches a time period T at which the associated debts must be settled. We only admit the class of perturbations that do not violate this type of transversality argument.

Overall, condition (19) therefore states that the government should trade off the marginal costs of greater inefficiency imposed on lower types against the marginal benefits of being able to channel resources to those who benefit most

from them under W . In this sense it can be read as a direct efficiency-equity trade-off. One of the most useful consequences of reading it in this way is that it implies model-consistent measures of concepts such as the degree of progressivity in the income tax schedule. Indeed, it reveals an aspect of the Mirrlees model that is initially quite counter-intuitive: higher marginal tax rates imposed on agents at points low down in the type distribution are a means for achieving greater cross-sectional equality, by reducing the rents of the better-off. This means that associating the shape of the marginal tax schedule with the degree of ‘progressivity’ implied by policy – as is commonly done in popular discussions about tax policy – is likely to be a deeply misleading exercise. A far better, model-informed set of measures of the degree of progressivity would follow by evaluating the object on the right-hand side of (19) for different values of θ' , with lower values associated with a more equitable distribution of resources. For instance, if one assumes a utilitarian objective with additively separable utility that is logarithmic in consumption then this measure would be:

$$E [c(\theta) | \theta > \theta'] - E [c(\theta)]$$

for each $\theta' \in \Theta$. This object has a direct interpretation as the marginal cost of allowing utility inequality above type θ' , and in this regard is far more closely connected with the underlying trade-off that the policymaker faces than any measure of the increasingness of the marginal tax rate; for this reason it would be a far better focal point in analysing ‘progressivity’. Similarly the relevance of the object $DC(\theta)$ shows that it is not just the tax rate *per se* that determines the inefficiency associated with a given schedule, but the relative effect that this rate has on information rents; this will depend in particular on the size of the labour supply elasticity.

3.3 Pareto efficiency

Proposition 2 provides a necessary optimality condition when social preferences across possible allocations correspond to the complete ordering induced by some objective W . But it is of interest also to consider whether any useful policy prescriptions may arise under more parsimonious, incomplete orderings of allocations – notably the partial ordering induced by a standard Pareto criterion. An allocation A is Pareto-dominated by an alternative B if all agents in the economy prefer B to A , with the preference strict in at least one case. Among the set of allocations that are relaxed incentive-feasible some may not lie on the Pareto frontier, in the sense that they are Pareto-dominated by others in the same set. The partial social preference ordering induced over allocations by the Pareto criterion is relatively uncontroversial by comparison with the (complete) ordering induced by a specific choice of W , such as utilitarianism or Rawlsianism. For this reason it is of interest to see how far the Pareto criterion can guide optimal tax rates.

Werning (2007) first discussed the usefulness of this criterion in an optimal tax setting, characterising the requirements of Pareto efficiency in a simplified

version of the Mirrlees model with additively separable utility. The cost objects that we have defined above can be manipulated to provide a more general statement, which follows with a little extra work from the proof of Proposition 2. The focus will be on ‘local’ Pareto efficiency, which we define as follows: an allocation $(c(\theta), y(\theta))$ is locally Pareto efficient within a given set if for any $\delta > 0$ there does not exist an alternative allocation $(c'(\theta), y'(\theta))$ in the same set that Pareto-dominates $(c(\theta), y(\theta))$, and for which $|c'(\theta) - c(\theta)| < \delta$ and $|y'(\theta) - y(\theta)| < \delta$ for all $\theta \in \Theta$. An allocation being locally Pareto efficient among the set of relaxed incentive-feasible allocations does not rule out that it might be Pareto-dominated by an alternative allocation in that set that is not local to it, just as differential optimality conditions do not guarantee global optima. But a necessary condition for local Pareto efficiency is clearly also necessary for global efficiency, so local arguments can still deliver potentially useful policy restrictions.

We have the following result. Its proof is in the appendix.

Proposition 3 *Consider any interior allocation that satisfies the transversality condition of Proposition 2. This allocation is locally Pareto efficient in the set of relaxed incentive-feasible allocations if and only if the following three conditions hold:*

1. For all $\theta' \in \Theta$:

$$E [MC(\theta) | \theta > \theta'] \cdot (1 - F(\theta')) - DC(\theta') \cdot f(\theta') \geq 0 \quad (20)$$

2. The left-hand side of (20) is monotonically decreasing (weakly) in θ' .

3.

$$E [MC(\theta)] \geq 0$$

The first and third conditions in the Proposition are not that surprising given the definitions of the cost terms. Clearly if the utility of all agents above some θ' – or across the entire distribution – can be increased at negative marginal cost then a Pareto improvement can be made. Non-decreasingness of the cost-gap term is perhaps less obvious. Intuitively if it didn’t hold then even with (20) satisfied it would be possible to increase the utility rents earned above θ' by a unit, decrease those earned above $\theta'' > \theta'$ by an offsetting unit (so that utility above θ'' remains constant), and generate surplus resources at the margin equal to the difference between the two cost gaps. The impact on utility would be zero for all agents outside the interval $[\theta', \theta'']$ and positive for those within it. Hence we would have a Pareto improvement.

As noted by Werning, there is a strong link between the question of Pareto efficient taxation and optimal taxation with a Rawlsian objective, which can be seen by comparing (20) with the main optimality condition (19). A Rawlsian optimum will satisfy inequality (20) exactly for all $\theta' > \underline{\theta}$, since the point at which it is satisfied is the point at which tax revenue would fall if still more productive distortions were introduced at θ' . That is, it characterises the peak

of the famous ‘Laffer curve’ specific to agent θ' . Going beyond that peak implies Pareto inefficiency – utility is reduced without raising any compensating resources. A Rawlsian ‘maxmin’ criterion treats taxpayers above $\underline{\theta}$ as revenue sources alone, and thus will seek the peak of the Laffer curve when trading off equity and efficiency considerations for each taxpayer above $\underline{\theta}$. More general welfare criteria that put strictly positive weight on the utility of all agents in Θ can be expected to satisfy the inequality strictly: this follows trivially from (19) when the last term is positive.

How likely is it that the Pareto criterion will be satisfied in practice? In general the non-negativity restriction (20) will simply place an upper bound on the level of the productive distortion that is tolerable at θ' , which in turn will depend on the deeper properties of the utility function. Higher labour supply elasticities, for instance, are more likely to be associated with a violation of the Pareto criterion by any given decentralised tax system. But our empirical exercise below suggests such violations are not likely to be a feature of the US income tax system at present: marginal tax rates are not so high as to be the ‘wrong side’ of the Laffer curve.

3.3.1 Implication: the Pareto inefficiency of linear benefit withdrawal

More interesting is the decreasingness in θ' that we require of the left-hand side of the inequality. Provided the type distribution is continuous over the relevant subset of Θ , this condition will be violated by any piecewise-linear tax schedule $T(y)$ that incorporates *decreases* in the marginal rate at threshold income levels. Such thresholds imply a non-convex, kinked budget set, and thus induce discrete differences in the allocations of individuals whose types are arbitrarily close to one another. At this point the agent moves from a higher to a lower marginal tax rate, and $DC(\theta')$ will jump discretely downwards as a consequence, whilst the first cost term in (20) is relatively unaffected (by continuity). Thus non-decreasingness will be violated.

Decreases in piecewise-linear effective tax schedules are a common feature of benefit programmes such as the Earned Income Tax Credit in the US and the Working Tax Credit in the UK, which augment the salary of low income earners but ‘withdraw’ the associated transfer at a fixed marginal rate as earnings rise above a certain threshold. At the upper limit of this withdrawal phase the effective marginal tax rate can drop substantially,¹⁰ inducing a non-convexity into the budget set. This will generally be Pareto inefficient. Specifically, it should be possible to deliver a strict improvement in the welfare of a subset of the agents who presently have earnings towards (but below) the upper end of the withdrawal band, by promising them a slightly lower marginal rate were

¹⁰For instance a single taxpayer with three or more children claiming EITC in the US in 2013 will pay an effective marginal rate of 21.06 per cent (in addition to other obligations) on incomes between \$17,530 and \$46,227, as the total quantity of benefits for which he or she is eligible falls with every extra dollar earned. At this upper threshold benefits are fully withdrawn, and the effective marginal rate thus drops by 21.06 percentage points.

they to work a small quantity of extra hours. ‘Smoothing out’ the kink in the tax schedule would have the effect of incentivising higher earnings from those in the upper end of the withdrawal band – and thus delivering higher tax revenue from them – whilst leaving all others unaffected.

Notice that this argument is very similar to the case for a zero top marginal rate when there is a finite upper type $\bar{\theta}$. There too, if $\bar{\theta}$ is stopping work with a strictly positive marginal rate there can be no loss to a slight cut in any taxes paid on still higher earnings, since these taxes are not affecting the choice of any other agent. If $\bar{\theta}$ chooses to work harder she must be strictly better off, and the extra work delivers extra revenue to the policymaker. The argument may be repeated until the marginal rate paid on the last cent earned is zero. Indeed, it is clear from (20) that if there is a finite upper type with strictly positive density then any Pareto efficient tax system will not distort the allocation of that type: $DC(\bar{\theta}) = 0$, corresponding to a zero marginal tax rate.

3.4 Discussion: primal and dual approaches

As noted in the introduction, the existing literature on the Mirrlees model contains a number of insightful optimality statements, and it is instructive to consider how ours relates to them. A useful way to understand condition (19) is as a ‘primal’ characterisation of the optimum, contrasting with the ‘dual’ approach taken by, for instance, Roberts (2000) and Saez (2001). The primal/dual distinction here is used by analogy to the closely related literature on Ramsey taxation models – in which second-best market allocations are found within a pre-specified set of distorted ‘competitive equilibria with taxes’.¹¹ The primal approach to these problems is to maximise consumer utility directly over the set of real (consumption and leisure) allocations, subject to resource constraints and so-called ‘implementability’ restrictions, where the latter ensure that the allocation can be decentralised. In Mirrleesian problems the equivalent restriction is incentive compatibility. Prices (and taxes) are then implicit in the solution; they are not treated as the objects of choice. The dual approach, by contrast, optimises welfare by choice of market prices, given the known response of consumers to these prices. The resulting expressions are able to exploit well-known results from consumer theory to express optimal taxes in terms of Hicksian and Marshallian demand elasticities.

Roberts (2000) and Saez (2001) independently showed how a dual approach to the Mirrlees model could be taken, considering perturbations to a decentralised non-linear tax system. As in Ramsey problems, the resulting expressions can be manipulated to be written in terms of compensated and uncompensated labour supply elasticities. This was the key insight of Saez (2001), and it has proved extremely useful for empirical work: it implies optimal taxes can be calculated from estimable elasticities. A large applied literature has emerged in response, surveyed comprehensively by Piketty and Saez (2013). These authors

¹¹See Atkinson and Stiglitz (1980) and Ljungqvist and Sargent (2012) for useful discussions of this distinction.

follow Diamond and Saez (2011) in emphasising the practical benefits of optimality statements that depend on estimable ‘sufficient statistics’ – notably the behavioural elasticities that feature in dual characterisations.

A lesson we hope will be drawn from the current paper is that a primal characterisation may be just as tractable as the dual, and thus of complementary value in drawing applied policy lessons. Though condition (19) is novel, the primal approach more generally is dominant in the growing ‘New Dynamic Public Finance’ literature, which considers the difficult problem of optimal Mirrleesian taxation over time.¹² Given the complexity of this literature, one can understand Piketty and Saez’s comment that the primal approach “tends to generate tax structures that are highly complex and results that are sensitive to the exact primitives of the model.”¹³ But in light of the present results this judgement seems a little rash: though we have made important structural assumptions on preferences, particularly single crossing, we believe condition (19) gives a simple and intuitive characterisation of the optimum. With structural (parametric) forms assumed for preferences it can also link optimal taxes to a small number of estimable parameters, such as the Frisch elasticity of labour supply and the elasticity of intertemporal substitution. We demonstrate this in the subsequent section. Therefore we hope that our primal method might be seen not in contrast to a ‘sufficient statistics’ approach to tax policy, but rather as contributing to it.

Whether the dual or primal characterisation will be simpler in general depends – here as in Ramsey models – on the structure of consumer preferences. As a general rule additive separability in the direct utility function tends to yield greater tractability in the primal problem, as relevant cross-derivatives of the utility function can then be set to zero. This accounts for the dominant use of the primal approach in studying *dynamic* taxation problems of both Ramsey and Mirrleesian form, where preferences are generally assumed separable across time and states of the world.¹⁴ Dual representations, by contrast, generally depend on the complete set of cross-price elasticities in different time periods and states of the world – that is, on an intractably large Slutsky substitution matrix. For this reason the primal approach is likely to continue to dominate the literature on dynamic Mirrleesian problems. Indeed, Brendon (2012) shows how the main theoretical results of the present paper can be generalised to that setting with only minor changes.

4 A parametric example

The main advantage of our approach is the tractability of condition (19) when one is willing to impose parametric structure on individual preferences: in this

¹²See Kocherlakota (2010) and Golosov, Tsyvinski and Werning (2006) for introductions to this literature.

¹³Piketty and Saez use the terminology ‘mechanism design approach’ in place of what is labelled the ‘primal approach’ here.

¹⁴A good example of the former is Lucas and Stokey (1983).

case it generates simple restrictions that can be used for empirical work. In this section we demonstrate how the condition simplifies when preferences take the isoelastic, additively separable form already discussed above. This delivers a particularly simple expression for the main optimality trade-off that highlights some important consequences for optimal taxes of income effects in labour supply.

Throughout the exercise it is important to remember that the particular choice of utility function combines a substantive statement about the structure of ordinal preferences with a normalisation to a particular cardinal form. As discussed in section 2.2.3, the optimal allocation is invariant to equivalent utility representations provided the social preference function G is adjusted appropriately. For this reason it is advantageous to fix on a representation for which the objects MC and DC take the simplest forms available, which will be achieved by using additively separable representations of the direct utility function where possible – so that MC will equal the inverse marginal utility of consumption. For example, the set of utility functions characterised by King, Plosser and Rebelo (1988) all describe the same ordinal consumption-labour supply preference map, and so will deliver the same optimum with the relevant adjustment to G . It is therefore simplest to focus on the special case of these preferences that is additively separable, with log consumption utility.¹⁵

4.1 Isoelastic, separable preferences

Suppose that the utility function again takes the form:

$$u(c, y; \theta) = \frac{c^{1-\sigma} - 1}{1 - \sigma} - \frac{(ye^{-\theta})^{1+\frac{1}{\varepsilon}}}{1 + \frac{1}{\varepsilon}} \quad (21)$$

As noted above, this means that $MC(\theta)$ collapses to the simple object $c(\theta)^\sigma$, whilst $DC(\theta)$ satisfies:

$$DC(\theta) = \frac{\tau(\theta)}{1 - \tau(\theta)} \frac{\varepsilon}{1 + \varepsilon} c(\theta)^\sigma$$

The main optimality condition (19) can then be expressed in the following form:

$$\frac{\tau(\theta')}{1 - \tau(\theta')} = \left(1 + \frac{1}{\varepsilon}\right) \cdot \frac{1 - F(\theta')}{f(\theta')} \cdot \frac{E[c(\theta)^\sigma | \theta > \theta'] - g(\theta') E[c(\theta)^\sigma]}{c(\theta')^\sigma} \quad (22)$$

where we write $g(\theta')$ as shorthand for $\frac{E[G_u(u(\theta)) | \theta > \theta']}{E[G_u(u(\theta))]}$.

This expression has dissected the term $DC(\theta')$ in order to express taxes as a function of all other variables, but it remains a succinct statement of the model's key efficiency/information-rent trade-off. It is particularly useful for drawing attention to four distinct factors that affect this trade-off:

¹⁵In dynamic models these arguments clearly no longer apply, as curvature in the period utility function governs dynamic preferences.

1. The Frisch elasticity of labour supply: *ceteris paribus* a higher value for ε will result in lower tax rates – a manifestation of the usual ‘inverse elasticity’ rule.
2. The (inverse) hazard rate term $\frac{1-F(\theta')}{f(\theta')}$: in general a lower value for this implies lower taxes. This is the mechanical effect of giving greater weight to efficiency costs when the distribution of types at θ' is dense relative to the measure of agents above that point. As an effect it was first highlighted by Diamond (1998), and has driven much work on the shape of empirical earnings distributions in the US and elsewhere.
3. The character of social preferences, as captured by the term $g(\theta')$: the higher is g the lower will be optimal taxes, reflecting the fact that reductions in information rents carry a higher direct cost to the government the more the welfare of relatively productive agents is valued.¹⁶
4. The curvature of the utility function: for a given (increasing) consumption allocation and given set of values for $g(\theta')$ the final fraction is easily shown to be greater the higher is σ , and so too will be taxes. Intuitively, more curvature in the utility function makes it more costly to provide welfare to high types. The benefits from reducing information rents will consequently be higher, and higher marginal taxes are desirable as a means to reduce them.

The first three of these factors are well understood, but the fourth has received relatively little attention in the literature to date. This is largely because the dual characterisation of the optimum provided by Saez (2001) is far simpler in the case of no income effects – nested here when $\sigma = 0$. This is no longer so acutely the case for our representation: whilst the last fraction in (22) reduces to $(1 - g(\theta))$ under quasi-linearity, for non-zero σ it remains manageable and has an intuitive interpretation as a measure of the relative marginal cost of providing information rents to types above θ' .

4.1.1 Optimal top rates

Turning to specific policy prescriptions, if preferences do take the form (21) then allowing $\sigma > 0$ has potentially significant implications for optimal top income tax rates if the upper tail of the empirical earnings distribution takes a Pareto form. To see this, we need first to explore the implications of a Pareto earnings tail for the upper tail of the underlying type distribution. Suppose there is a fixed upper rate of income tax $\bar{\tau} \in (0, 1)$, where $\bar{\tau}$ need not necessarily be optimal. For all types who pay this tax rate we have the usual first-order optimality condition:

$$(1 - \bar{\tau}) c(\theta)^{-\sigma} = e^{-\theta} (y(\theta) e^{-\theta})^{\frac{1}{\varepsilon}} \quad (23)$$

¹⁶If $G(u, \theta)$ exhibits symmetric inequality aversion then $g(\theta)$ will range between 0 and 1, the former corresponding to a Rawlsian social objective and the latter utilitarianism.

In the static Mirrlees model consumption must equal disposable income, so that if the support of the type distribution is unbounded we will have for high enough θ :

$$\frac{c(\theta)}{y(\theta)} \simeq (1 - \bar{\tau}) \quad (24)$$

Combining these two expressions and rearranging gives:

$$(1 - \bar{\tau})^{\frac{1-\sigma}{\sigma+\frac{1}{\varepsilon}}} e^{\frac{1+\varepsilon}{1+\sigma\varepsilon}\theta} \simeq y(\theta) \quad (25)$$

Suppose that top incomes are Pareto distributed, as argued by Diamond and Saez (2011), so that:

$$P\left(y(\theta) > y(\theta') \mid y(\theta) > y(\tilde{\theta})\right) = \left(\frac{y(\theta')}{y(\tilde{\theta})}\right)^{-\alpha} \quad (26)$$

for some parameter α and threshold type $\tilde{\theta} < \theta'$. Then it follows from (25) that the distribution of *types* in the upper tail satisfies:

$$F\left(\theta \mid \theta > \tilde{\theta}\right) \simeq 1 - e^{-\frac{1+\varepsilon}{1+\sigma\varepsilon}\alpha(\theta-\tilde{\theta})} \quad (27)$$

That is, types are distributed approximately exponentially at the top, with exponential parameter $\frac{1+\varepsilon}{1+\sigma\varepsilon}\alpha$. Reversing the logic, it follows that if types are exponentially distributed at the top then top incomes will be approximately Pareto, provided there is a fixed upper tax rate $\bar{\tau}$. This is of interest because of the following result, which we prove in the appendix.

Proposition 4 *Suppose preferences take the form (21), $g(\theta) \in \left[0, \frac{1}{1-F(\theta)}\right)$ for all $\theta \in \Theta$, and that the type distribution has an exponential upper tail with exponential parameter $\frac{1+\varepsilon}{1+\sigma\varepsilon}\alpha$. Then if $\sigma > \alpha$ optimal marginal tax rates at the top of the income distribution cannot be bounded below 100 per cent. Hence if optimal marginal tax rates converge at the top, it is to 100 per cent.*

This is a complete reversal of the ‘zero distortion at the top’ result, which demonstrates once more the importance to tax policy of the shape of the upper tails of the type distribution. Diamond and Saez (2011) provide evidence from 2005 US tax return data that earnings do indeed have a Pareto tail, with an estimated Pareto parameter of approximately 1.5. If true this implies optimal marginal tax rates will be confiscatory at the top of the income distribution whenever the coefficient of relative risk aversion exceeds 1.5, provided social preferences are sufficiently averse to inequality. Estimates of the relative risk aversion coefficient are famously variable. Chetty (2006) argues that relatively weak observed income effects on labour supply are consistent with a value no greater than 1.25, whilst Attanasio, Banks and Tanner (2002) use UK portfolio allocation data to estimate a value of 1.44. But experimental results can be substantially higher – see, for instance, Barsky et al. (1997), who find a mean

elasticity of intertemporal substitution of 0.2 (i.e., $\sigma = 5$). On top of this there are the double-digit (and even triple-digit) values commonly obtained from ‘macro’ analyses of the equity premium puzzle – see, for instance, Kocherlakota (1996) for a survey.

Thus *if* the isoelastic, separable preference structure is a reasonable benchmark then confiscatory taxation at the top will be best for a quite plausible subset of the parameter space – provided the social preference restriction $g(\theta) < \frac{1}{1-F(\theta)}$ is satisfied. This merits some comment. On the surface it seems quite mild: it will be violated only if *higher* social preference weight is put on marginal utility increments for high earners than low. But note that the cardinalisation of (consumption) utility itself admits more and more curvature as σ grows – there is a sense in which at high values of σ there is less for social preferences to ‘compensate for’. For instance, if agent A is consuming ten times as much as agent B then a utilitarian will be willing to reduce the consumption of A by ten units in order to increase that of B by one unit when $\sigma = 1$, whilst the ratio becomes 100 to one when $\sigma = 2$. Perhaps increasing social preference weights can be justified in such circumstances as a proxy for less concave, non-separable direct utility functions. Notice, though, that a Rawlsian would certainly allow the top tax rate to limit to 100 per cent – implying that the top of the Laffer curve is not reached for any finite top rate. Put differently, if ordinal preferences are consistent with the isoelastic, separable form and $\sigma > 1.5$ then a top tax rate below 100 per cent can only be justified by an intrinsic concern for the welfare of upper-rate taxpayers. Efficiency concerns alone will not do.

The intuition behind the Proposition is that when σ is large the marginal cost of allowing information rents grows rapidly as high types’ consumption grows. A small reduction in information rents at any given point in the distribution allows a large quantity of resources to be released from those higher up. If the tail of the type distribution is sufficiently dense – i.e., if α is low enough – then the total value of the marginal resources gained when information rents fall will be infinitely large whenever the rate of income tax on high types has an upper bound less than 100 per cent. Put differently, if the tax rate is bounded by some $\tilde{\tau} < 1$ then for large θ consumption must grow at least proportionally in income, by some factor that exceeds $(1 - \tilde{\tau})$. But if consumption grows linearly in income it will inherit a Pareto distribution – meaning that $MC(\theta)$, which here equals c^σ , will have an unbounded expectation whenever σ exceeds the Pareto parameter.

Care should certainly be taken in interpreting the result. In particular, a key step in the logic is to move from a Pareto distribution in *earnings* to a Pareto distribution in *consumption* when taxes are bounded. In the static model this is immediate, since all post-tax earnings must be consumed. But if the possibility of savings were present then it is far less clear what the relationship would be between the consumption and earnings distributions – or, indeed, whether the optimal shape of either obtained in this static model should carry over. Indeed, the results of Krueger and Perri (2006) suggest that consumption inequality may be relatively weakly related to earnings inequality, so that a focus on the

latter as a proxy for the former is unlikely to be appropriate.

Arguably this reflects a more general inadequacy of the static Mirrleesian framework: tax schedules are chosen annually, and earnings potentials evolve regularly over time, so that the government’s problem is far more complex in reality than choosing one single consumption-output schedule. The validity of our extreme progressive result in a dynamic setting that more closely parallels real-world government choice remains an open question.

5 Calibration results

5.1 General approach

In this section we conduct a more direct calibration exercise to assess the appropriateness of recent US tax policy on the basis of our theoretical characterisation. Specifically, we look to place direct numerical values on the objects in the main optimality condition (19) in order to answer the question: How might the balance between efficiency and equity considerations be better struck? To motivate this we multiply (19) through by $[1 - F(\theta')]$ to express it as:

$$\begin{aligned} & f(\theta') \cdot DC(\theta') & (28) \\ = & [1 - F(\theta')] \cdot \left\{ E[MC(\theta) | \theta > \theta'] - \frac{E[G_u(u(\theta)) | \theta > \theta']}{E[G_u(u(\theta))]} \cdot E[MC(\theta)] \right\} \end{aligned}$$

This expression is useful because the objects on the left- and right-hand sides can be interpreted directly as population-weighted cost terms, expressible in monetary units: the left-hand side is the per-capita marginal quantity of resources lost if the choices of an agent of type θ' are distorted in order to reduce information rents above θ' by a unit at the margin. We can label it $EfC(\theta')$. The object on the right-hand side is the per-capita marginal quantity of resources gained from so doing, assuming that the total value of the social welfare criterion is being held constant. We can label it $EqC(\theta')$. In these terms the equation states:

$$EfC(\theta') = EqC(\theta') \quad (29)$$

To operationalise the expression we need: (a) an individual (cardinal) preference structure, (b) a social objective, (c) a distribution of types, and (d) a tax schedule. It will then be possible to infer an optimal consumption-income choice for all types θ , and to study the properties of the (simulated) consumption-income distribution that is induced across types. All of the objects in (28) can then be evaluated, allowing for a direct comparison between the two types of cost for each possible θ' . This in turn allows for both qualitative predictions about the appropriate direction of tax reform at each given point in the earnings distribution (of the form: ‘Insufficient weight is being given to efficiency considerations for those earning $\$X$ ’), and quantitative predictions about the magnitude of gains that could thereby be obtained. We proceed by setting out in turn the manner in which we determine the objects listed (a) to (d) above.

5.1.1 Individual preferences

At the individual level we retain our earlier focus on isoelastic, separable preferences of the form given in (21). This means that $EfC(\theta)$ will be given by:

$$f(\theta) \cdot \frac{\tau(\theta)}{1 - \tau(\theta)} \frac{\varepsilon}{1 + \varepsilon} c(\theta)^\sigma$$

whilst $EqC(\theta')$ will be:

$$[1 - F(\theta')] \cdot \left\{ E[c(\theta)^\sigma | \theta > \theta'] - \frac{E[G_u(u(\theta)) | \theta > \theta']}{E[G_u(u(\theta))]} \cdot E[c(\theta)^\sigma] \right\}$$

We will consider a number of alternative values for ε and σ . The main restriction we face in doing so is the one implied by Proposition (4) above: if a Pareto distribution for top earnings is to be matched by the estimated type distribution then σ is restricted to be below the Pareto parameter. If it is not then the marginal benefits from reducing information rents will always be measured as infinite, and both qualitative and quantitative conclusions about appropriate reform will thus be trivial.

Again, these expressions for the marginal cost terms will be unique only up to monotone transformations of the utility function. Strictly what we have here is a pair of expressions normalised so that the cost of providing information rents is interpreted as a consumption cost alone. The main justification for our chosen normalisation is the intuitive appeal of separating the level of *consumption* utility from the information rents gained by greater *productive* capability where the preference structure allows this. In general it is not possible to find an additively separable representation for arbitrary preference forms, but where it is possible, as here, there seems little harm in exploiting it.

5.1.2 Social preferences

We assume that the function $G(u, \theta)$ is independent of θ , and concave in u , taking the form:

$$G(u; \theta) = \frac{1 - e^{-\lambda u}}{\lambda}$$

which limits to the utilitarian case when $\lambda \rightarrow 0$ and exhibits inequality aversion more generally for $\lambda > 0$. As a benchmark we will focus on the utilitarian case, but positive values for λ are subsequently considered.

5.1.3 Type distribution

To obtain a distribution of types we use cross-sectional US earnings and expenditure data from the 2009 wave of the PSID (relating to 2008 tax year).¹⁷ Making use of the more detailed expenditure data provided in recent waves of

¹⁷We take 2008 calendar-year data as the best available proxy for 2008 tax-year data.

the PSID, we associate a household's expenditure on non-durable goods with its consumption, and use this to approximate individual (rather than household) consumption levels by standard equivalence scales. This is then combined with the PSID labour income data for all agents earning in excess of \$1000, together with an estimate of the marginal tax rate τ that the agent is likely to have faced, to infer a value for their type under the assumption that they satisfy the optimality condition:

$$c^{-\sigma} (1 - \tau) = e^{-(1+\frac{1}{\varepsilon})\theta} y^{\frac{1}{\varepsilon}} \quad (30)$$

for a given σ and ε parameterisation. The marginal income tax rates are approximated using the NBER's TAXSIM programme,¹⁸ and augmented by state-level consumption taxes to approximate the total effective wedge at the labour-consumption margin. Notice that when individuals reside in larger households condition (30) implicitly assumes that they are nonetheless able to be the sole consumer of the goods purchased with the last dollar that they earn at the margin.

This procedure provides us with a cross-section of values for θ , which we then use to estimate the distribution of types in the 2008 US workforce non-parametrically.¹⁹ Since we use a normal smoothing kernel the estimated distribution cannot possibly exhibit the Pareto earnings tails that Saez (2001) and Diamond and Saez (2011) argue is a feature of US tax return data: with a finite amount of data the upper tail of a distribution given by a weighted sum of normal densities will never be Pareto, no matter what is the true data generating process. To overcome the problem we calibrate the upper tail of the type distribution to an exponential with parameter $\frac{1+\varepsilon}{1+\sigma\varepsilon}\alpha$, where α is the Pareto parameter. We set α to 1.5, in line with the evidence provided in Diamond and Saez (2011). The calibration is imposed only for the top 0.1 per cent of earners in the resulting estimated distribution $\hat{F}(\theta)$. It is nonetheless of significance for our results for high earners, since it ensures the estimated value of $1 - \hat{F}(\theta)$ remains large relative to the associated density estimate $\hat{f}(\theta)$, and thus that the marginal benefits of reducing information rents continue to be assessed at a non-negligible value even for very high types. If we were instead to use an estimated upper tail obtained by a normal kernel it would always be optimal to let the top rate of tax limit to zero, with $Efc(\theta') > Eqc(\theta')$ for all θ' sufficiently large when this does not hold. In this sense our results are certainly sensitive to the calibration of the tail distribution.

Note finally that because we have not restricted attention exclusively to quasi-linear preferences the inference procedure for the type distribution is more complex than it would otherwise be. When $\sigma = 0$ is imposed consumption does not feature in (30), and so θ can be inferred from earnings data alone. This allows, in particular, for detailed government data on tax returns to be applied to the exercise, as in Saez (2001) and many subsequent papers. For the general case this is not possible, as tax returns do not provide information on consumption.

¹⁸See <http://www.nber.org/taxsim/>

¹⁹We correct for sampling bias by use of the PSID population weights.

One could perhaps proceed by inferring post-tax income and imposing that this must equal consumption – a procedure followed, for instance, by Blundell and Shephard (2012) in their assessment of the optimal tax treatment of single mothers in the UK. But this does risk biasing the results. In particular it implies that agents with very high incomes should also have very high consumption levels, and thus very low marginal utilities of consumption. This would tend to over-state θ for high earners, for whom savings are generally substantial. This is the main reason for our use of the PSID dataset, whose detailed consumption data allow us to infer a model-consistent type distribution without such bias.

5.1.4 Tax schedule

The final object that must be inputted into our analysis is an estimate for the tax schedule facing individuals in the US in 2008. This will allow us to infer a hypothetical consumption-income allocation induced by that schedule for any given preference structure, and that allocation in turn can be used to gauge the balance between efficiency and equity considerations that the estimated tax schedule strikes. Given the complexity of the actual US tax code, as well as its variations from state to state, we must clearly simplify substantially if we are to keep the analysis consistent with the Mirrlees model’s assumption of a *single* schedule. Our approach is to estimate a non-linear, parametric schedule linking reported income in the PSID series to the marginal rates inferred from TAXSIM. The precise schedule that we fit follows Gouveia and Strauss (1994), who propose the following form for the marginal rate:

$$\tau(y) = b \left[1 - (sy^\rho + 1)^{-\frac{1}{\rho} - 1} \right] \quad (31)$$

This allows for three degrees of freedom, with b , s and ρ to be determined. Note that as $y \rightarrow \infty$ $\tau(y) \rightarrow b$, so this parameter can be interpreted as the limiting top rate of tax. ρ then controls the degree of curvature in the tax schedule, and s is a scaling parameter. We select b , s and ρ by minimising a sum of squared residuals between $\tau(y)$ and the tax rate reported by TAXSIM for each individual in the PSID series.²⁰ The values we obtain are (approximately) $b = 0.46$, $s = 0.015$ and $\rho = 0.6$. The resulting marginal tax schedule takes the following form: We will assume that agents in the economy must choose subject to this schedule, and analyse the marginal efficiency and equity costs of small changes to it. Note that the concavity of marginal tax rates in income will imply a convex budget set in consumption-income space for all individuals, so we may solve the model by imposing individual first-order conditions alone.

Clearly this estimated schedule will only provide a first approximation to tax policy in the US, so our focus will be on robust *qualitative* results that emerge for large classes of alternative preference structures. If, for instance, we find that the marginal costs of additional inefficiency outweigh the marginal costs of additional inequality for all plausible preference structures when incomes are within a certain range then this will suggest improvements can readily be made

²⁰ Again, we also include state-level consumption taxes in the definition of τ .

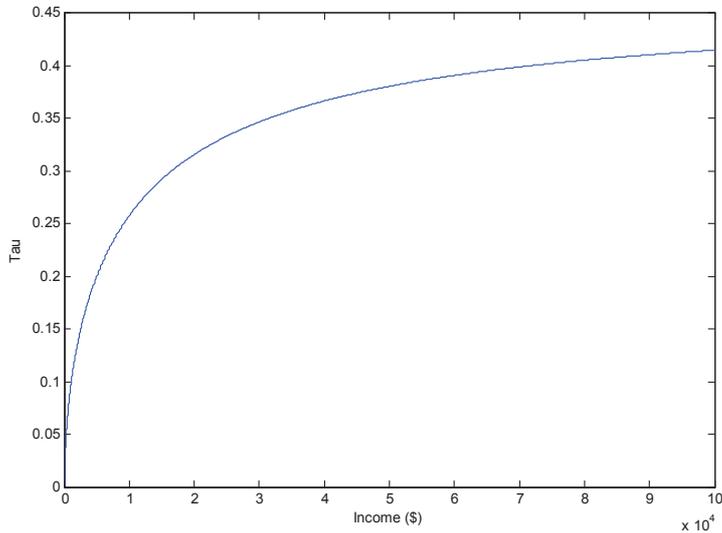


Figure 1: Estimated marginal tax schedule

by reducing actual federal tax rates applicable at such incomes – particularly if actual rates exceed those of the estimated schedule.

5.2 Benchmark results

In this subsection we carry out the benchmark policy experiment, comparing the marginal costs of efficiency and equity at each point in the earnings distribution for a simulated allocation obtained by the procedure explained above. For our benchmark case we consider a utilitarian objective criterion with log utility ($\sigma = 1$) and Frisch elasticity ε of 0.5. The former is well-known to be consistent with the stylised fact of balanced growth, and the latter falls well within the range of conventional micro estimates for the intensive-margin Frisch elasticity.²¹

5.2.1 Productivity distribution

Given these preferences our first step is to estimate a type distribution consistent with the observed consumption-income data. In Figure 2 we chart the estimated distribution of labour productivity, e^θ , which has a clearer economic interpretation than θ itself. This distribution incorporates the calibrated exponential tail for θ for the top 0.1 per cent of productivity draws, as discussed above. As one would expect, the productivity distribution replicates well-known features of the income distribution, notably a large degree of positive skewness.

²¹0.5 is the value recommended for the intensive-margin Frisch by Chetty et al. (2011).

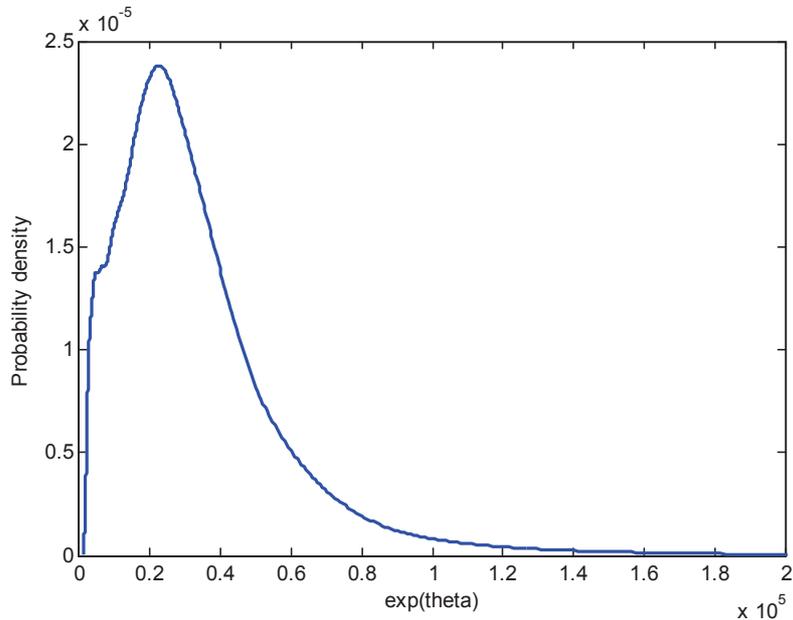


Figure 2: Estimated distribution of labour productivity

5.2.2 Efficiency-equity gap

Given this type distribution our next step is to simulate a consumption-income allocation across all types, assuming that the tax schedule takes the parametric form plotted above. We can then use this to graph the associated cost difference term:

$$\Delta(\theta) \equiv EqC(\theta) - EfC(\theta)$$

against $y(\theta)$ for all θ that induce incomes up to \$500,000. The value of this term can be understood as the marginal quantity of resources that would be gained per worker in the economy from a tax reform that held constant the value of the social objective but reduced information rents by a unit at θ . When it is positive the implication is that greater focus should be given to reducing inequality in the consumption distribution relative to θ , since information rents are more costly than a utilitarian policymaker should be willing to tolerate. When negative the implication is that too great a degree of productive distortion is witnessed at θ , and more information rents above that point ought to be tolerated.

Figure 3 plots this object for the benchmark calibration. The most notable feature of it is that $\Delta(\theta)$ is positive at all income levels, and this remains true for arbitrarily large income levels. The implication is that the information rents that are being permitted under the approximated tax schedule are far greater than a utilitarian policymaker should be willing to tolerate. By increasing the marginal distortion faced by an individual earning around \$26,000 and redistributing the

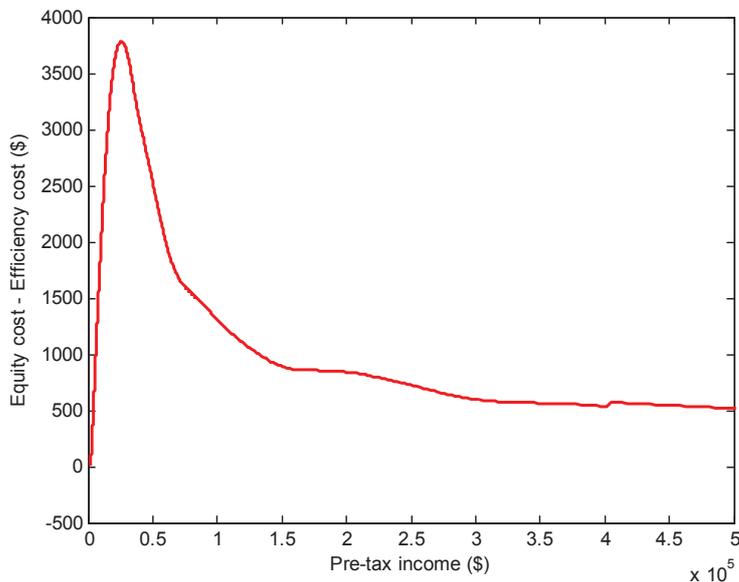


Figure 3: Cost difference: Benchmark calibration

proceeds uniformly (in utility terms) across the population the government could – if the calibration is accurate – generate an additional \$3,800 per taxpayer at the margin, for every unit by which information rents above \$26,000 are reduced.

The cost gap term $\Delta(\theta)$ is maximised for intermediate income levels: \$26,000 is very close to the median of the simulated distribution, which is approximately \$27,500. In itself this is an interesting result, since the focus of policy work in the optimal tax literature often gives special treatment to the problem of finding the optimal *top* rate, distinct from more general questions about the rest of the schedule.²² Yet it is not clear that the resources to be gained from getting the top rate right are any more significant than those available from striking a better trade-off elsewhere. The results here instead suggest intermediate tax bands may be of much greater consequence – and, for a utilitarian at least, improvements to them may offer far greater scope for improvement.

There are three reasons why changes to information rents at intermediate incomes are offering the greatest marginal gains. First there is the quite mechanistic fact that additional inefficiency induced at an intermediate point in the distribution reduces the information rents enjoyed by a comparatively large share of the total population – all those earning a greater amount. If distortions are increased only in the upper tail of the income distribution then the aggregate reduction in rents is lower, and so too is the quantity of resources

²²See, for instance, Chapter 2 of the first volume of the Mirrlees Review (Mirrlees, 2010).

raised. Second, marginal tax rates – and thus the marginal costs of additional distortions – remain relatively low for precisely the income range at which $\Delta(\theta)$ is maximised: around 33 per cent. This is significantly below the limiting top rate of 45.5 per cent implied by the estimated schedule, so the costs of additional distortions are not so great here. Third, and by contrast, for those on *low* incomes the average cost of utility provision to higher types is very close to the average cost across the entire distribution – and it is the comparison between the two that governs equity costs for a utilitarian policymaker. This means that $EqC(\theta)$ is also relatively low at this point.

5.3 Sensitivity

The strongest result above is the qualitative one, that regardless of the tax schedule considered the marginal costs from additional inefficiency in agents’ production decisions are always outweighed by the marginal benefits from reducing information rents, from a utilitarian perspective. In this subsection we test how sensitive these results are to alternative individual and social preference assumptions. We consider variations along three dimensions: an increase in the (Frisch) elasticity of labour supply above 0.5, a reduction in σ below 1, and a shift in policy preferences towards greater inequality aversion. As we will discuss, the last of these also provides an opportunity to assess the simulated distribution against the restrictions implied by Pareto efficiency, derived in Section 3.3 above.

5.3.1 High labour supply elasticity

We first consider the impact on our results of raising the elasticity of labour supply. Figure 4 shows the effect on the cost comparison for the approximated tax schedule of increasing ε from 0.5 to 1 – a value well above most empirical estimates. A higher labour supply elasticity induces greater productive distortions for all types, and so it is not surprising that the high-elasticity schedule for $\Delta(\theta)$ lies strictly below the benchmark case. But the qualitative result that marginal inequality costs are greater than inefficiency costs across the entire pre-tax income schedule still goes through, with the basic shape of the $\Delta(\theta)$ schedule effectively unchanged.

Higher values of the Frisch elasticity will induce $\Delta(\theta)$ become negative for some θ , but a value of 1 is already unusually large. For the given preference structure it can be shown to imply a Hicksian (‘steady-state’) labour supply elasticity in excess of 0.5 for most agents – well above even the value of 0.3 suggested by Chetty (2012), which itself allowed for the possibility of downward bias in prior estimates due to optimisation frictions. We consider it as generating an upper bound on the marginal cost of additional distortions, and infer that our general conclusion of excessive weight being given to efficiency relative to equity in the design of the tax schedule is robust to plausible changes in labour supply elasticity.

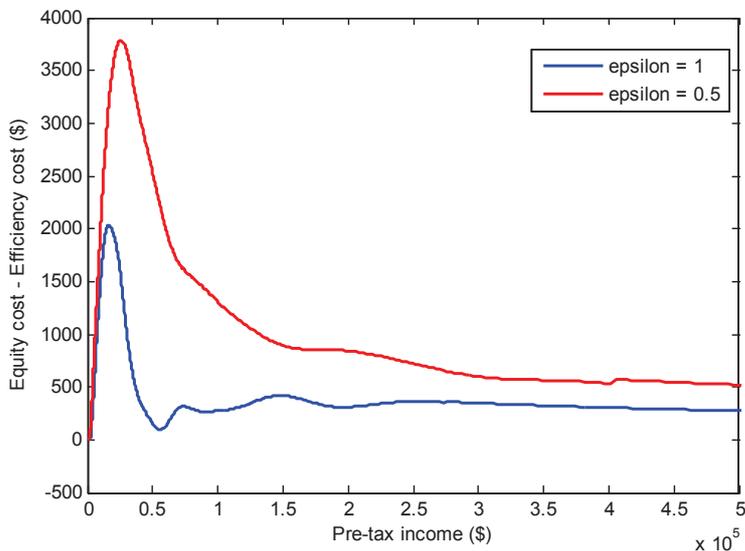


Figure 4: Sensitivity: higher labour supply elasticity

5.3.2 Low income effects

We next consider the effect on the results of reducing the coefficient of relative risk aversion, lowering σ from 1 to 0.5. We keep the social objective utilitarian, so that the net effect is to reduce the marginal (social) cost of providing consumption utility to high types, and thus make it more likely that efficiency costs will dominate equity costs. Figure 5 charts the difference in marginal cost terms in this case.²³ As anticipated, the cost difference term is no longer positive across the entire range of incomes: for earnings between around \$40,000 and \$130,000 it would now be preferable to reduce marginal tax distortions. For all other incomes higher distortions would still be preferable.

$\sigma < 1$ is a justifiable possibility if the uncompensated elasticity of labour supply with respect to the real wage is positive, as has been found in most micro studies – a point made by Chetty (2006). Nonetheless, empirical estimates of the uncompensated elasticity large enough to justify $\sigma = 0.5$ are rare,²⁴ and contradict the stylised fact of balanced growth. The range of incomes with a negative cost difference $\Delta(\theta)$ disappears once σ is increased above around 0.65.

We have already noted that increasing σ above the value taken by the Pareto parameter in the upper tail of the empirical earnings distribution would imply infinite marginal inequality costs, so testing the numerical sensitivity of our

²³We omit a direct comparison with the benchmark because the reduction in σ lowers all marginal cost terms by an order of magnitude.

²⁴The mean of the estimates inferred by Chetty (2006) from labour supply data was $\sigma = 0.71$.

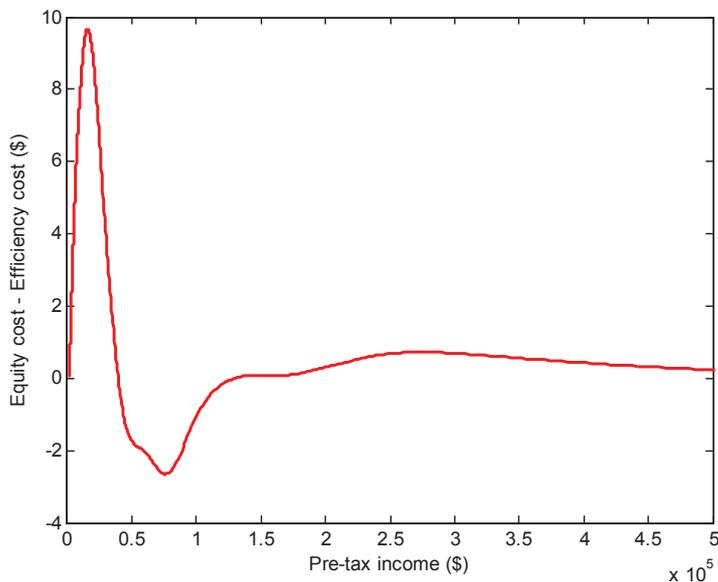


Figure 5: Sensitivity: lower coefficient of relative risk aversion

results to high income effects – i.e., values for σ substantially larger than 1 – is not necessary. We take from this exercise that under utilitarianism and isoelastic preferences there is a robust case for claiming too much weight is given to efficiency over equity considerations in the US tax system for both high incomes (over, say, \$150,000) and low-to-medium incomes (less than \$35,000), and that the conclusion additionally applies elsewhere in the income distribution unless the coefficient of relative risk aversion is lower than conventionally assumed.

5.3.3 Inequality aversion

We next consider the impact on our results of the social preference structure. Figure 6 shows the effect of raising the social preference parameter λ above 0 (the utilitarian case), through 1 to the limiting Rawlsian case in which $\lambda \rightarrow \infty$. Unsurprisingly a greater degree of inequality aversion increases the marginal costs of additional inequality at all income levels, relative to marginal inefficiency costs. This a necessary result, as changes to the social preference structure only affect the value of $EqC(\theta)$ – not the efficiency cost $EfC(\theta)$ – and can only serve to increase it.

Perhaps a less obvious feature of the results is the decline as λ increases in the income level at which the maximum gains from additional inefficiency are realised. To see why this occurs, note that for θ close to the lower end of the

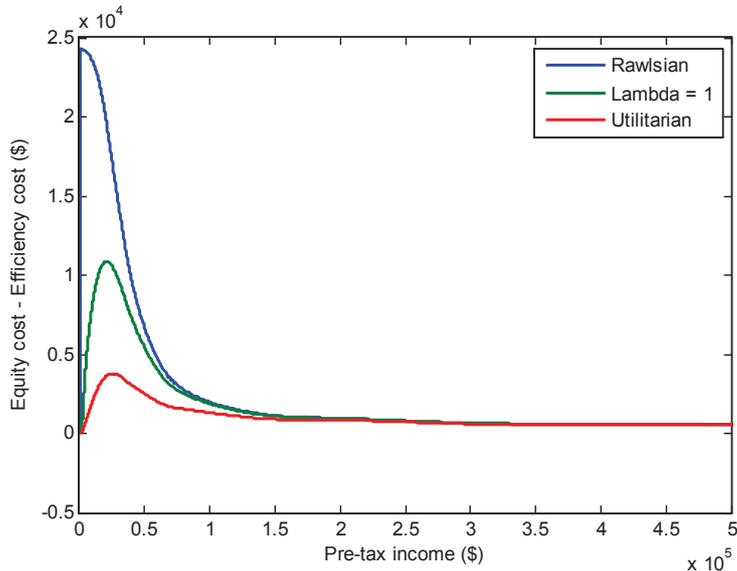


Figure 6: Sensitivity: varying social preferences

type distribution the value of $EqC(\theta)$ will approximately equal:

$$E[c(\theta)^\sigma] \cdot (1 - g(\theta))$$

where $g(\theta')$ is again shorthand for $\frac{E[G_u(u(\theta))|\theta > \theta']}{E[G_u(u(\theta))]}$. Moving from utilitarianism to Rawlsianism is equivalent to moving $g(\theta)$ from 1 to 0 for values of θ close to the minimum of the type distribution, which clearly induces a substantial change in $EqC(\theta)$ for such types. Once the policymaker has no intrinsic concern for the welfare of low (but not minimal) types, the fact that reducing their information rents saves on incentive costs paid to almost all agents in the rest of the distribution means that the gains from additional distortions become much greater, and higher marginal taxes are justified.

5.4 Testing Pareto efficiency

A final exercise motivated by our theoretical results is to test the Pareto efficiency of the estimated tax schedule, which we do in this subsection. Section 3.3 demonstrated that a tax schedule can only be Pareto efficient if the cost gap term $\Delta(\theta)$ taken for the Rawlsian case is positive and decreasing for all θ . Positivity implies that a reduction in productive inefficiencies at θ will not raise tax revenue, and decreasingness guarantees that utility cannot be provided to an interval of Θ whilst generating a resource surplus. We noted above that decreasingness would almost certainly be violated by a piecewise-linear tax

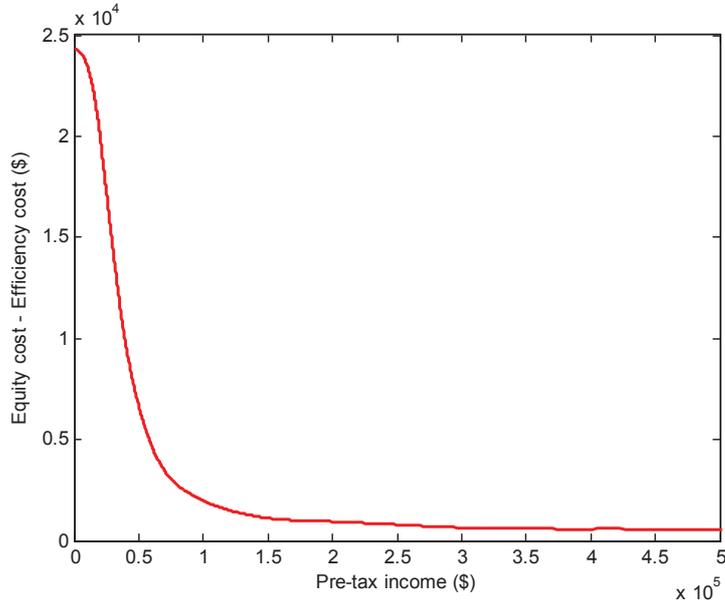


Figure 7: Checking Pareto efficiency

schedule that exhibited discrete effective reductions in the marginal tax rate – a fairly common feature of the withdrawal of income assistance programmes. Here we are analysing the approximated schedule graphed in Section 5.1.4, which by construction contains no such anomalies – it is smooth and concave. The aim here is thus to check for possible violations of Pareto efficiency beyond these more obvious cases.

To this end, Figure 7 reproduces the relevant cost gap term for our benchmark calibration. This clearly satisfies positivity, and decreasingness is exhibited everywhere apart from a small interval of earnings in the region of \$400,000. The approximated marginal tax schedule is essentially flat at this point, with the small violation driven by changes in the fitted density of types. Since density estimation is extremely imprecise in the tails we do not put too much weight on this finding, and conclude that our smoothed approximation to the US tax system does not imply any clear violation of Pareto efficiency, even if the same may not be said of precise benefit systems. Further simulations, not reported here, show that this result is not sensitive to realistic alternative calibrations.

6 Conclusion

Despite pioneering the genre the Mirrleesian optimal income tax model stands strangely apart from most other screening problems in the manner in which

it is conventionally analysed. Since the seminal work of Saez (2001) the focus of the tax literature has been on the direct role of labour supply elasticities, social preference parameters and the structure of the type distribution in shaping optimal rates. It has been rare to frame these objects in terms of the key trade-off emphasised by the parallel mechanism design literature – between the competing costs of information rents and productive efficiency. This paper has presented a new characterisation of the Mirrlees problem that is directly interpretable in terms of these latter objects, and explores the practical insights that it provides.

Specifically, we defined two marginal cost variables that capture, in turn, the cost of providing utility in a manner that keeps information rents constant, and the cost of reducing information rents in a manner that leaves utility constant. We demonstrated that these objects must satisfy a number of simple and intuitive relationships to one another at any optimum. Moreover, as cost terms they can be used to analyse existing tax schedules, providing meaningful answers to questions such as: Are the costs of additional inefficiency too great to warrant improving the distribution of welfare, or is the opposite true? We showed how to operationalise just such a question in section 5, inferring a distribution of productivity types for the US economy based on data from the 2009 wave of the PSID, and using this to simulate our key cost terms over the entire range of the income distribution, given an estimated parametric marginal tax schedule. The results strongly implied that the US tax system is giving too little weight to efficiency concerns relative to equity, in the sense that the marginal costs of tolerating information rents under the induced consumption-income allocation are far greater than the marginal efficiency costs associated with reducing them – at least for a utilitarian policymaker facing agents with concave utility functions of a conventional form.

These results are certainly not definitive, though they do require an answer if the existing tax system is to be defended. Two obvious criticisms are, first, that they rely on the static version of the Mirrlees model, and, second, that they are sensitive to the precise specification of individual and social preferences. The first of these is important because the static model assumes the tax system is the only means individuals have for providing themselves with consumption insurance – a factor that may bias policy preferences in favour of more distribution than would be optimal were precautionary savings behaviour appropriately incorporated. Note, for instance, that if incomes had a large stochastic component then those with high earnings one year would not necessarily consume much more than those whose earnings were substantially lower, and thus differences in the marginal cost of providing utility to each of them might not be that substantial.

These issues should be surmountable by considering an appropriate dynamic incarnation of the optimal policy problem, and it will be interesting to see how the results generalise to such a setting. The second criticism certainly has some force. In practice economists can only hope to recover ordinal preference maps, and any objective criterion that relies on the curvature properties of the direct utility function or of a social preference function is essentially contestable. But

if the Mirrleesian problem is to be posed at all then some sort of interpersonal utility comparison is unavoidable: one cannot set policy without an objective. Among the set of possible objectives our main focus has been on a utilitarian criterion, combined with isoelastic preferences at the individual level that are separable between consumption and labour supply. Within this domain our conclusions do seem robust to most plausible parameterisations of the ordinal preference map. There are, of course, limits. A utilitarian faced with no curvature in the consumption utility schedule will clearly not accept any productive inefficiencies, since the marginal cost of information rents to this policymaker is always zero. But to justify efficiency being the dominating concern certainly requires a non-conventional parameterisation for social or individual preferences, and in itself this is an interesting result.

References

- [1] Atkinson, A.B., and J.E. Stiglitz (1980), *Lectures on Public Economics* (Second Ed.), McGraw-Hill.
- [2] Attanasio, O.P., J. Banks and S. Tanner (2002), ‘Asset Holding and Consumption Volatility’, *Journal of Political Economy*, 110, 771–792.
- [3] Barsky, R.B., F.T. Juster, M.S. Kimball and M.D. Shapiro (1997), ‘Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study’, *Quarterly Journal of Economics*, 112, 537–579.
- [4] Blundell, R., and A. Shephard (2012), ‘Employment, Hours of Work and the Optimal Taxation of Low-Income Families’, *Review of Economic Studies*, 79, 481–510.
- [5] Bolton, P., and M. Dewatripont (2005), *Contract Theory*, MIT Press.
- [6] Brendon, C.F. (2012), ‘Efficiency-equity Trade-offs for Dynamic Mirrleesian Tax Problems’, Manuscript (available at <http://mwpweb.eu/CharlesBrendon>)
- [7] Chetty, R. (2006), ‘A New Method of Estimating Risk Aversion’, *American Economic Review*, 96, 1821–1834.
- [8] Chetty, R. (2012), ‘Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply’, *Econometrica*, 80, 969–1018.
- [9] Chetty, R., A. Guren, D. Manoli and A. Weber (2011), ‘Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins’, *American Economic Review*, 101, pages 471–475.

- [10] Diamond, P.A. (1998), ‘Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates’, *American Economic Review*, 88, 83–95.
- [11] Diamond, P.A., and E. Saez (2011), ‘The Case for a Progressive Tax: From Basic Research to Policy Recommendations’, *Journal of Economic Perspectives*, 25, 165–190.
- [12] Golosov, M., N.R. Kocherlakota and A. Tsyvinski (2003), ‘Optimal Indirect and Capital Taxation’, *Review of Economic Studies*, 70, 569–587.
- [13] Golosov, M., A. Tsyvinski and I. Werning (2006), ‘The New Dynamic Public Finance: A User’s Guide’, *NBER Macroeconomics Annual*, 21, 317–363.
- [14] Gouveia, M., and R.P. Strauss (1994), ‘Effective Federal Tax Functions: An Exploratory Empirical Analysis’, *National Tax Journal*, 47, 317–340.
- [15] Greenwood, J., Z. Hercowitz and G. Huffman (1988), ‘Investment, Capacity Utilization, and the Real Business Cycles’, *American Economic Review*, 78, 402–417.
- [16] King, R.G., C.I. Plosser and S.T. Rebelo (1988), ‘Production, Growth and Business Cycles I: The Basic Neoclassical Model’, *Journal of Monetary Economics*, 21, 195–232.
- [17] Kocherlakota, N.R. (1996), ‘The Equity Premium: It’s Still a Puzzle’, *Journal of Economic Literature*, 34, 42–71.
- [18] Kocherlakota, N.R. (2010), *The New Dynamic Public Finance*, Princeton, Princeton University Press.
- [19] Laffont, J.-J., and D. Martimort (2002), *The Theory of Incentives: The Principal-Agent Model*, Princeton University Press.
- [20] Ljungqvist, L., and T.J. Sargent (2012), *Recursive Macroeconomic Theory*, Third Edition, MIT Press
- [21] Lucas, R.E. Jr, and N.L. Stokey (1983), ‘Optimal fiscal and Monetary Policy in an Economy without Capital’, *Journal of Monetary Economics*, 12, 55–93.
- [22] Mankiw, N.G., and M.C. Weinzierl (2010), ‘The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution’, *American Economic Journal: Economic Policy*, 2, 155–76.
- [23] Milgrom, P., and I. Segal (2002), ‘Envelope Theorems for Arbitrary Choice Sets’, *Econometrica*, 70, 583–601.
- [24] Mirrlees, J.A. (1971), ‘An Exploration in the Theory of Optimal Income Taxation’, *Review of Economic Studies*, 38, 175–208.

- [25] Mirrlees, J.A. (ed.) (2010), *Dimensions of Tax Design: The Mirrlees Review*, Institute for Fiscal Studies, Oxford University Press.
- [26] Mirrlees, J.A. (ed.) (2011), *Tax By Design: The Mirrlees Review*, Institute for Fiscal Studies, Oxford University Press.
- [27] Piketty, T., and E. Saez (2013), ‘Optimal Labor Income Taxation’, manuscript, forthcoming in Handbook of Public Economics, vol. 5.
- [28] Roberts, K. (2000), ‘A Reconsideration of the Optimal Income Tax’, in Hammond, P.J., and G.D. Myles (eds.), *Incentives and Organization: Papers in Honour of Sir James Mirrlees*, Oxford University Press.
- [29] Rochet, J.-C., and Choné, P. (1998), ‘Ironing, Sweeping, and Multi-dimensional Screening’, *Econometrica*, 66, 783–826.
- [30] Sadka, E. (1976), ‘On Income Distribution, Incentive Effects and Optimal Income Taxation’, *Review of Economic Studies*, 43, 261–268.
- [31] Saez, E. (2001), ‘Using Elasticities to Derive Optimal Income Tax Rates’, *Review of Economic Studies*, 68, 205–229.
- [32] Saez, E. (2002), ‘Optimal Income Transfer Programs: Intensive Versus Extensive Labour Supply Responses’, *Quarterly Journal of Economics*, 117, 1039–1073.
- [33] Saez, E., and S. Stantcheva (2013), ‘Generalized Social Marginal Welfare Weights for Optimal Tax Theory’, *NBER Working Paper 18835*.
- [34] Seade, J.K. (1977) ‘On the Shape of Optimal Tax Schedules’, *Journal of Public Economics*, 7, 203–236.
- [35] Weinzierl, M.C. (2012), ‘Why Do We Redistribute So Much But Tag So Little? The Principle of Equal Sacrifice and Optimal Taxation’, *Harvard Business School Working Paper 12-64*.
- [36] Werning, I. (2007), ‘Pareto Efficient Income Taxation’, MIT Working Paper.

7 Appendix

7.1 Preferences and single crossing

We demonstrate here the claim in Section 2.1 that the single crossing property is implied whenever agents have common preferences over consumption and labour supply, whilst differing in their ability to convert labour supply into output according to θ . Specifically, suppose that preferences over consumption and labour supply for all agents are described by a common utility function $\tilde{u} : \mathbb{R}_+^2 \rightarrow \mathbb{R}$, where \tilde{u} is C^2 , increasing in its first argument (consumption)

and decreasing in its second (labour supply), and describes normal demands for both consumption and leisure. Labour supply, in turn, can be converted into output according to a C^2 function $l : \mathbb{R}_+ \times \Theta \rightarrow \mathbb{R}_+$. $l(y; \theta)$ thus gives the number of hours that it takes an individual whose type is θ to produce y units of output. We assume $l_y > 0$, $l_\theta < 0$ and $l_{y\theta} \leq 0$. The latter two restrictions imply that higher values of θ are unambiguously associated with higher productive efficiency. Note that this setup nests the case in which θ is a linear productivity parameter, which means $l(y; \theta) = \frac{y}{\theta}$.

We can then define the function $u : \mathbb{R}_+^2 \times \Theta \rightarrow \mathbb{R}$ by:

$$u(c, y; \theta) := \tilde{u}(c, l(y; \theta)) \quad (32)$$

Given the restrictions that we have imposed on the \tilde{u} and l functions, the single crossing property follows:

Lemma 5 $u(c, y; \theta)$ defined in (32) satisfies single crossing in θ . That is, for any distinct pair of allocations (c', y') and (c'', y'') such that $(c', y') < (c'', y'')$ (in the product order sense) and $\theta' < \theta''$, if $u(c'', y''; \theta') \geq u(c' y'; \theta')$ then $u(c'', y''; \theta'') > u(c' y'; \theta'')$.

Proof. Geometrically it is easy to see that this will be true provided the derivative condition (1) is satisfied at all allocations – that is, indifference curves in output-consumption space are uniformly ‘flattening’ in θ . We have:

$$\begin{aligned} \frac{d}{d\theta} \left(-\frac{u_y(c, y; \theta)}{u_c(c, y; \theta)} \right) &= -\frac{u_{y\theta}(c, y; \theta)}{u_c(c, y; \theta)} + \frac{u_y(c, y; \theta)}{u_c(c, y; \theta)} \frac{u_{c\theta}(c, y; \theta)}{u_c(c, y; \theta)} \\ &= -\frac{u_y(c, y; \theta)}{u_c(c, y; \theta)} \left[\frac{u_{y\theta}(c, y; \theta)}{u_y(c, y; \theta)} - \frac{u_{c\theta}(c, y; \theta)}{u_c(c, y; \theta)} \right] \end{aligned}$$

These derivatives satisfy:

$$\begin{aligned} u_c(c, y; \theta) &= \tilde{u}_c(c, l(y; \theta)) \\ u_y(c, y; \theta) &= \tilde{u}_l(c, l(y; \theta)) l_y(y; \theta) \\ u_{c\theta}(c, y; \theta) &= \tilde{u}_{cl}(c, l(y; \theta)) l_\theta(y; \theta) \\ u_{y\theta}(c, y; \theta) &= \tilde{u}_{ll}(c, l(y; \theta)) l_y(y; \theta) l_\theta(y; \theta) + \tilde{u}_l(c, l(y; \theta)) l_{y\theta}(y; \theta) \end{aligned}$$

Plugging into the previous expression gives:

$$\begin{aligned} \frac{d}{d\theta} \left(-\frac{u_y(c, y; \theta)}{u_c(c, y; \theta)} \right) &= -\frac{\tilde{u}_l(c, l(y; \theta)) l_y(y; \theta)}{\tilde{u}_c(c, l(y; \theta))} \left[\frac{\tilde{u}_{ll}(c, l(y; \theta)) l_\theta(y; \theta)}{\tilde{u}_l(c, l(y; \theta))} \right. \\ &\quad \left. + \frac{l_{y\theta}(y; \theta)}{l_y(y; \theta)} - \frac{\tilde{u}_{cl}(c, l(y; \theta)) l_\theta(y; \theta)}{\tilde{u}_c(c, l(y; \theta))} \right] \\ &= \underbrace{-\frac{\tilde{u}_l(c, l(y; \theta))}{\tilde{u}_c(c, l(y; \theta))}}_{>0} \underbrace{l_{y\theta}(y; \theta)}_{\leq 0} \\ &\quad - \underbrace{\frac{\tilde{u}_l(c, l(y; \theta)) l_y(y; \theta)}{\tilde{u}_c(c, l(y; \theta))}}_{>0} \underbrace{l_\theta(y; \theta)}_{< 0} \underbrace{\left[\frac{\tilde{u}_{ll}(c, l(y; \theta))}{\tilde{u}_l(c, l(y; \theta))} - \frac{\tilde{u}_{cl}(c, l(y; \theta))}{\tilde{u}_c(c, l(y; \theta))} \right]}_{>0} \end{aligned}$$

where the sign of the last term follows from normality in consumption, which implies increases in labour supply must raise the value of $-\frac{\tilde{u}_l}{u_c}$. This confirms condition (1), completing the proof. ■

7.2 Proof of Proposition 1

We first establish that an incentive-feasible allocation implies increasing c and y schedules together with condition (4). Increasingness follows almost immediately from single crossing. Suppose $(c(\theta'), y(\theta')) > (c(\theta''), y(\theta''))$ for some $\theta'' > \theta'$. θ' must weakly prefer $(c(\theta'), y(\theta'))$ to $(c(\theta''), y(\theta''))$ at an incentive-feasible allocation, but then single crossing implies θ'' must have a strict preference for $(c(\theta'), y(\theta'))$, violating incentive compatibility.

To obtain (4), note that for all $\theta < \bar{\theta}$ we can find $\varepsilon > 0$ such that for all $\delta \in [0, \varepsilon]$, $\theta + \delta \in \Theta$, and thus by incentive compatibility we have:

$$u(c(\theta), y(\theta); \theta) \geq u(c(\theta + \delta), y(\theta + \delta); \theta) \quad (33)$$

and

$$u(c(\theta + \delta), y(\theta + \delta); \theta + \delta) \geq u(c(\theta), y(\theta); \theta + \delta) \quad (34)$$

Rearranging and taking limits we have:

$$\lim_{\delta \rightarrow 0} \left[\frac{u(c(\theta + \delta), y(\theta + \delta); \theta) - u(c(\theta), y(\theta); \theta)}{\delta} \right] \leq 0 \quad (35)$$

$$\lim_{\delta \rightarrow 0} \left[\frac{u(c(\theta + \delta), y(\theta + \delta); \theta + \delta) - u(c(\theta), y(\theta); \theta + \delta)}{\delta} \right] \geq 0 \quad (36)$$

But by the assumed continuity properties of u these two objects must coincide. Thus the partial right-derivative of utility with respect to type report at an incentive-feasible allocation exists, and is equal to zero.²⁵ An identical argument establishes that the corresponding left-derivative exists and is equal to zero for all $\theta > \underline{\theta}$. We can summarise this by:

$$\frac{\partial}{\partial \sigma} [u(c(\sigma), y(\sigma); \theta)]|_{\sigma=\theta} = 0$$

Totally differentiating utility with respect to type gives:

$$\begin{aligned} \frac{d}{d\theta} [u(c(\theta), y(\theta); \theta)] &= \frac{\partial}{\partial \sigma} [u(c(\sigma), y(\sigma); \theta)]|_{\sigma=\theta} \\ &\quad + \frac{\partial}{\partial \theta} [u(c(\sigma), y(\sigma); \theta)]|_{\sigma=\theta} \end{aligned} \quad (37)$$

Condition (4) follows immediately.

We then need that a pair of increasing allocation schedules satisfying (4) will be globally incentive compatible. Suppose otherwise. Then there must exist a pair of allocations $(c(\theta'), y(\theta'))$ and $(c(\theta''), y(\theta''))$ for some θ', θ'' such that:

$$u(c(\theta''), y(\theta''); \theta') > u(c(\theta'), y(\theta'); \theta')$$

²⁵Notice that this does not require continuity in the allocation schedules $c(\cdot)$ and $y(\cdot)$.

Suppose first that $\theta' > \theta''$. We have:

$$\begin{aligned}
u(c(\theta'), y(\theta'); \theta') &= u(c(\theta''), y(\theta''); \theta'') & (38) \\
&+ \int_{\theta''}^{\theta'} u_\theta(c(\theta''), y(\theta''); \sigma) d\sigma \\
&+ \int_{\theta''}^{\theta'} \frac{d}{d\sigma} [u(c(\sigma), y(\sigma); \theta')] d\sigma \\
&= u(c(\theta''), y(\theta''); \theta') & (39) \\
&+ \int_{\theta''}^{\theta'} \frac{d}{d\sigma} [u(c(\sigma), y(\sigma); \theta')] d\sigma
\end{aligned}$$

From the envelope condition we know:

$$\left. \frac{d}{d\sigma} [u(c(\sigma), y(\sigma); \tilde{\theta})] \right|_{\tilde{\theta}=\sigma} = 0$$

Since $\theta' \geq \sigma$ for all σ in the integral in (39) it follows by single crossing that each term under the integral must be weakly positive. But clearly this implies:

$$u(c(\theta'), y(\theta'); \theta') \geq u(c(\theta''), y(\theta''); \theta')$$

contradicting the earlier supposition. A symmetric argument can be applied when $\theta'' > \theta'$, completing the proof.

7.3 Proof of Proposition 2

We prove the result by a perturbation argument, constructing a parametric class of direct revelation mechanisms that is relaxed incentive-compatible, and taking derivatives with respect to the main parameter in the neighbourhood of an optimum, noting that an allocation which is relaxed incentive-compatible and delivers the same value to the policymaker as the constrained-optimal allocation for the relaxed problem cannot generate a resource surplus.

Let $u^*(\sigma; \theta)$ denote the utility obtained by an agent whose type is θ and who reports σ when the allocation $c^* : \Theta \rightarrow \mathbb{R}_{++}$, $y^* : \Theta \rightarrow \mathbb{R}_{++}$ is constrained-optimal for the relaxed problem. Note the assumption of interiority in the c^* and y^* schedules here. We are interested in changing the consumption allocation to $c^* + \delta^c$ and the output allocation to $y^* + \delta^y$, where $\delta^c : \Theta \times I \rightarrow \mathbb{R}$ and $\delta^y : \Theta \times I \rightarrow \mathbb{R}$ are perturbation schedules defining a changed allocation for each type in Θ , given the value of some parameter $\Delta \in I$, where I is a convex, open interval of \mathbb{R} that includes 0. We normalise $\delta^c(\theta, 0) = \delta^y(\theta, 0) = 0$. For any $(\theta, \Delta) \in \Theta \times I$ these perturbation schedules to c and y can be used to define a utility perturbation $\delta^u : \Theta \times I \rightarrow \mathbb{R}$:

$$\delta^u(\theta, \Delta) := u(c^*(\theta) + \delta^c(\theta, \Delta), y^*(\theta) + \delta^y(\theta, \Delta); \theta) - u^*(\theta; \theta) \quad (40)$$

To obtain an optimality statement we need the perturbed allocation to be relaxed incentive-feasible. This means it must continue to satisfy the envelope

condition (4), which will be the case provided δ^u is C^1 and together with δ^c and δ^y satisfies:

$$\begin{aligned} \delta_\theta^u(\theta, \Delta) &= u_\theta(c^*(\theta) + \delta^c(\theta, \Delta), y^*(\theta) + \delta^y(\theta, \Delta); \theta) \\ &\quad - u_\theta(c^*(\theta), y^*(\theta); \theta) \end{aligned} \quad (41)$$

The following Lemma then motivates a focus on the object δ^u :

Lemma 6 *Suppose that a function $\delta^u : \Theta \times \mathbb{R} \rightarrow \mathbb{R}$ is C^1 in both of its arguments with $\delta^u(\theta, 0) = 0$ for all $\theta \in \Theta$ and satisfies:*

$$\begin{aligned} &\int_{\Theta} G(u(c^*(\theta), y^*(\theta); \theta) + \delta^u(\theta, \Delta), \theta) dF(\theta) \\ &= \int_{\Theta} G(u(c^*(\theta), y^*(\theta); \theta), \theta) dF(\theta) \end{aligned} \quad (42)$$

for all $(\theta, \Delta) \in \Theta \times I$ where I is an open subset of \mathbb{R} , $0 \in I$. Suppose further that the constrained-optimal allocation for the relaxed problem, (c^*, y^*) , is strictly interior. Then there exists a pair of functions $\delta^c : \Theta \times I \rightarrow \mathbb{R}$ and $\delta^y : \Theta \times I \rightarrow \mathbb{R}$ such that:

1. $\delta^c : \Theta \times I \rightarrow \mathbb{R}$ and $\delta^y : \Theta \times I \rightarrow \mathbb{R}$ are C^1 in Δ .

2. $\forall \theta \in \Theta$:

$$\delta^c(\theta, 0) = \delta^y(\theta, 0) = 0 \quad (43)$$

3. Equation (40) holds.

4. Equation (41) holds.

Proof. Condition 2 can be satisfied simply by ensuring $\delta^c(\theta, 0) = \delta^y(\theta, 0) = 0$. For $\Delta \neq 0$ condition (40) defines an indifference surface in consumption-output space on which the perturbed allocation must lie for any given θ if condition 3 is to hold. For condition 4 we additionally need to select an allocation on this surface for which u_θ satisfies (41). Consider the impact on u_θ of moving along the indifference surface, changing y by one unit at the margin. This is given by:

$$u_{c\theta} \frac{dc}{dy} + u_{y\theta}$$

where

$$\frac{dc}{dy} = -\frac{u_y}{u_c}$$

From the single crossing condition it is easy to show:

$$-u_{c\theta} \frac{u_y}{u_c} + u_{y\theta} > 0 \quad (44)$$

This implies movements along the indifference surface have a monotonic impact on u_θ . We have $\delta_\theta^u(\theta, 0) = 0$ and δ^u is C^1 , so by the interiority of the optimum

it must be possible to choose a unique point on the new indifference surface that satisfies (41) for any given Δ in an open neighbourhood of 0. This point defines $\delta^c(\theta, \Delta)$ and $\delta^y(\theta, \Delta)$. The continuity properties of u and δ^u clearly ensure that these schedules must be C^1 in Δ , so condition 1 is additionally satisfied. ■

This Lemma is useful because it allows us to proceed as follows: First, specify a C^1 schedule δ^u that satisfies (42) and has $\delta^u(\theta, 0) = 0$. Second, evaluate the marginal costs of applying this schedule as a utility perturbation within the set of relaxed incentive compatibility constraints, as Δ is varied in the neighbourhood of $\Delta = 0$. This marginal cost will be given by:

$$\int_{\Theta} [\delta_{\Delta}^c(\theta, 0) - \delta_{\Delta}^y(\theta, 0)] dF(\theta)$$

Because δ^u satisfies (42) this object must equal zero: if this were not true then a (positive or negative) movement in Δ away from zero would generate a resource surplus whilst holding constant the policymaker's objective, contradicting optimality.

We consider the following utility perturbation schedule:

$$\delta^u(\theta, \Delta) = \begin{cases} -\Delta & \text{if } \theta \leq \theta' \\ h(\theta, \Delta) & \text{if } \theta \in (\theta', \theta'') \\ \varepsilon(\Delta) & \text{if } \theta \geq \theta'' \end{cases}$$

for arbitrary values $\theta', \theta'' \in \Theta$, an arbitrary C^1 function $h : (\theta', \theta'') \times I \rightarrow \mathbb{R}$, with $h(\theta', \Delta) = -\Delta$, $h(\theta'', \Delta) = \varepsilon(\Delta)$, $h_{\theta}(\theta', \Delta) = h_{\theta}(\theta'', \Delta) = 0$, and $\varepsilon(\Delta)$ chosen to satisfy:

$$\begin{aligned} & \int_{\theta < \theta'} G(u(\theta, \theta) - \Delta, \theta) dF(\theta) & (45) \\ & + \int_{\theta \in (\theta', \theta'')} G(u(\theta, \theta) + h(\theta, \Delta), \theta) dF(\theta) \\ & + \int_{\theta > \theta''} G(u(\theta, \theta) + \varepsilon(\Delta), \theta) dF(\theta) \\ & = \int_{\Theta} G(u(\theta, \theta), \theta) dF(\theta) \end{aligned}$$

This clearly satisfies the requirements of Lemma 6, so will be implementable by consumption-output perturbations δ^c and δ^y for values of Δ close enough to 0.

We are interested in the marginal effects of moving Δ away from 0 on the policymaker's resources. By differentiating (40) and (41) with respect to Δ and using the cost definitions provided above we have:

$$\begin{aligned} & \int_{\Theta} [\delta_{\Delta}^c(\theta, 0) - \delta_{\Delta}^y(\theta, 0)] dF(\theta) & (46) \\ & = \int_{\Theta} [\delta_{\Delta}^u(\theta, 0) MC(\theta) - \delta_{\theta\Delta}^u(\theta, 0) DC(\theta)] dF(\theta) \\ & = 0 \end{aligned}$$

where the last line follows from the optimality of the original allocation for the relaxed problem.

For the chosen schedule we have:

$$\delta_{\Delta}^u(\theta, 0) = \begin{cases} -1 & \text{if } \theta \leq \theta' \\ h_{\Delta}(\theta, 0) & \text{if } \theta \in (\theta', \theta'') \\ \varepsilon'(0) & \text{if } \theta \geq \theta'' \end{cases}$$

and:

$$\delta_{\theta\Delta}^u(\theta, 0) = \begin{cases} 0 & \text{if } \theta \leq \theta' \\ h_{\theta\Delta}(\theta, 0) & \text{if } \theta \in (\theta', \theta'') \\ 0 & \text{if } \theta \geq \theta'' \end{cases}$$

Thus a necessary condition for an optimum is:

$$- \int_{\theta \leq \theta'} MC(\theta) dF(\theta) \quad (47)$$

$$+ \int_{\theta \in (\theta', \theta'')} [h_{\Delta}(\theta, 0) \cdot MC(\theta) - h_{\theta\Delta}(\theta, 0) \cdot DC(\theta)] dF(\theta)$$

$$+ \varepsilon'(0) \cdot \int_{\theta \geq \theta''} MC(\theta) dF(\theta)$$

$$= 0 \quad (48)$$

To obtain the required optimality statement we need to eliminate the object $\varepsilon'(0)$ and the arbitrary function h (together with its derivatives). The arguments are valid for all choices of $\theta' < \theta''$, so we can consider the limiting outcome as $\theta'' \rightarrow \theta'$. For each distinct choice of θ'' we will need a distinct h , and we now denote this by a superscript. We have:

$$\begin{aligned} & \lim_{\theta'' \rightarrow \theta'} \left[\int_{\theta'}^{\theta''} [h_{\Delta}^{\theta''}(\theta, 0) \cdot MC(\theta) - h_{\theta\Delta}^{\theta''}(\theta, 0) \cdot DC(\theta)] dF(\theta) \right] \\ &= \lim_{\theta'' \rightarrow \theta'} \left[- \int_{\theta'}^{\theta''} h_{\theta\Delta}^{\theta''}(\theta, 0) \cdot DC(\theta) \cdot f(\theta) d\theta \right] \\ &= -DC(\theta') \cdot f(\theta') \lim_{\theta'' \rightarrow \theta'} \left[\int_{\theta'}^{\theta''} h_{\theta\Delta}^{\theta''}(\theta, 0) d\theta \right] \\ &= -DC(\theta') \cdot f(\theta') \lim_{\theta'' \rightarrow \theta'} [h_{\Delta}^{\theta''}(\theta'', 0) - h_{\Delta}^{\theta''}(\theta', 0)] \\ &= -(1 + \varepsilon'(0)) \cdot DC(\theta') \cdot f(\theta') \end{aligned}$$

where $f : \Theta \rightarrow \mathbb{R}_+$ is the density function associated with F .

Taking the derivative of (45) with respect to Δ at $\Delta = 0$ and rearranging gives:

$$\varepsilon'(0) = \frac{\int_{\theta \leq \theta'} G'(u(\theta, \theta), \theta) dF(\theta) - \int_{\theta \in (\theta', \theta'')} G'(u(\theta, \theta), \theta) f_{\Delta}(\theta, 0) dF(\theta)}{\int_{\theta \geq \theta''} G'(u(\theta, \theta), \theta) d\mu(\theta)} \quad (49)$$

which in the limiting case of $\theta'' \rightarrow \theta'$ becomes:

$$\varepsilon'(0) = \frac{\int_{\theta \leq \theta'} G'(u(\theta, \theta), \theta) dF(\theta)}{\int_{\theta > \theta'} G'(u(\theta, \theta), \theta) dF(\theta)} \quad (50)$$

Putting these objects together, the necessary optimality condition becomes (at the limit):

$$\begin{aligned} & - \int_{\Theta} MC(\theta) dF(\theta) \\ & - \frac{\int_{\Theta} G'(u(\theta, \theta), \theta) dF(\theta)}{\int_{\theta > \theta'} G'(u(\theta, \theta), \theta) dF(\theta)} \cdot DC(\theta') \cdot \pi(\theta') \\ & + \frac{\int_{\Theta} G'(u(\theta, \theta), \theta) dF(\theta)}{\int_{\theta > \theta'} G'(u(\theta, \theta), \theta) dF(\theta)} \cdot \int_{\theta > \theta'} MC(\theta) dF(\theta) \\ & = 0 \end{aligned} \quad (51)$$

The result then follows from simple algebra.

7.4 Proof of Proposition 3

We take as given the logic of the proof of Proposition 2, only detailing the further arguments necessary.

‘Only if’ requires considering the three conditions of the proposition in turn. The first is straightforward: suppose there is a θ' for which:

$$E[MC(\theta) | \theta > \theta'] \cdot (1 - F(\theta')) - DC(\theta') \cdot f(\theta') < 0 \quad (52)$$

Then a marginal reduction in information rents at θ' with uniform utility increments for all higher types would generate strictly positive resources, raise the utility of all types above θ' , and keep constant the utility of those at θ' or lower. This is a Pareto improvement.

Next suppose non-decreasingness is violated. Then it must be possible to find $\theta' < \theta''$ such that:

$$\begin{aligned} & E[MC(\theta) | \theta > \theta'] \cdot (1 - F(\theta')) - DC(\theta') \cdot f(\theta') \\ & - E[MC(\theta) | \theta > \theta''] \cdot (1 - F(\theta'')) + DC(\theta'') \cdot f(\theta'') \\ & < 0 \end{aligned}$$

But this implies:

$$DC(\theta'') \cdot f(\theta'') - DC(\theta') \cdot f(\theta') + \int_{\theta'}^{\theta''} MC(\theta) f(\theta) d\theta < 0$$

which in turn means that it is possible to generate surplus resources via the Pareto improvement that gives uniform utility at the margin to all agents with types in (θ', θ'') and leaves others unaffected. This is a Pareto improvement.

Finally, suppose $E[MC(\theta)] < 0$. Then we can generate surplus resources by giving all agents a uniform utility increase, which is a Pareto improvement.

For the ‘if’ part we consider increasing the utility of all agents at the margin according to some C^1 profile $\delta^u(\theta)$. From equation (46) in the proof of Proposition 2 we have that the marginal cost to the policymaker of applying this perturbation will be:

$$\int_{\Theta} [\delta^u(\theta) MC(\theta) - \delta_{\theta}^u(\theta) DC(\theta)] f(\theta) d\theta \quad (53)$$

Integrating the first term by parts, we have:

$$\begin{aligned} & \int_{\Theta} [\delta^u(\theta) MC(\theta) - \delta_{\theta}^u(\theta) DC(\theta)] f(\theta) d\theta \quad (54) \\ &= \delta^u(\underline{\theta}) \left[\int_{\underline{\theta}}^{\bar{\theta}} MC(\tilde{\theta}) f(\tilde{\theta}) d\tilde{\theta} \right] \\ & \quad + \int_{\Theta} \delta_{\theta}^u(\theta) \left\{ \left[\int_{\theta}^{\bar{\theta}} MC(\tilde{\theta}) f(\tilde{\theta}) d\tilde{\theta} \right] - DC(\theta) f(\theta) \right\} d\theta \end{aligned}$$

We can then apply integration by parts again to the last line:

$$\begin{aligned} & \int_{\underline{\theta}}^{\bar{\theta}} \delta_{\theta}^u(\theta) \left\{ \left[\int_{\theta}^{\bar{\theta}} MC(\tilde{\theta}) f(\tilde{\theta}) d\tilde{\theta} \right] - DC(\theta) f(\theta) \right\} d\theta \quad (55) \\ &= -\delta^u(\bar{\theta}) DC(\bar{\theta}) f(\bar{\theta}) - \delta^u(\underline{\theta}) \left\{ \left[\int_{\underline{\theta}}^{\bar{\theta}} MC(\tilde{\theta}) f(\tilde{\theta}) d\tilde{\theta} \right] - DC(\underline{\theta}) f(\underline{\theta}) \right\} \\ & \quad - \int_{\Theta} \delta^u(\theta) \frac{d}{d\theta} \left(\left[\int_{\theta}^{\bar{\theta}} MC(\tilde{\theta}) f(\tilde{\theta}) d\tilde{\theta} \right] - DC(\theta) f(\theta) \right) d\theta \end{aligned}$$

The transversality condition implies $\delta^u(\bar{\theta}) DC(\bar{\theta}) f(\bar{\theta}) \leq 0$, whilst the term in curly brackets on the second line is non-positive by condition 1 in the proposition, applied for $\theta' = \underline{\theta}$. The final term is non-negative by the non-increasingness imposed in condition 2 of the proposition, and thus the final line of the overall expression for the cost of the perturbation in (54) must be positive. It follows that the total cost will be positive provided:

$$\left[\int_{\underline{\theta}}^{\bar{\theta}} MC(\tilde{\theta}) f(\tilde{\theta}) d\tilde{\theta} \right] \geq 0$$

This is condition 3 in the proposition.

7.5 Proof of Proposition 4

The proof follows by application of condition (22) and a contradiction.²⁶ For $\theta' > \tilde{\theta}$ the distribution function $F(\theta')$ must satisfy:

$$\begin{aligned} F(\theta') &= 1 - P(\theta > \theta') \\ &= 1 - P(\theta > \theta' | \theta > \tilde{\theta}) P(\theta > \tilde{\theta}) \\ &= 1 - e^{-\frac{1+\varepsilon}{1+\sigma\varepsilon}\alpha(\theta' - \tilde{\theta})} [1 - F(\tilde{\theta})] \end{aligned} \quad (56)$$

So that $f(\theta')$ is given by:

$$f(\theta') = \frac{1+\varepsilon}{1+\sigma\varepsilon} \alpha e^{-\frac{1+\varepsilon}{1+\sigma\varepsilon}\alpha(\theta' - \tilde{\theta})} [1 - F(\tilde{\theta})] \quad (57)$$

Substituting for the hazard ratio term and simplifying, condition (22) may then be written for $\theta' > \tilde{\theta}$ as:

$$\frac{\tau(\theta')}{1 - \tau(\theta')} = \left(\sigma + \frac{1}{\varepsilon}\right) \cdot \frac{1}{\alpha} \cdot \frac{E[c(\theta)^\sigma | \theta > \theta'] - g(\theta') E[c(\theta)^\sigma]}{c(\theta')^\sigma} \quad (58)$$

Writing the expectations operators in integral form, we have:

$$\begin{aligned} \frac{\tau(\theta')}{1 - \tau(\theta')} &= \left(\sigma + \frac{1}{\varepsilon}\right) \cdot \frac{1}{\alpha} \cdot \frac{1}{c(\theta')^\sigma} \\ &\cdot \left[[1 - g(\theta') (1 - F(\theta'))] \frac{1+\varepsilon}{1+\sigma\varepsilon} \alpha \int_{\theta'}^{\infty} c(\theta)^\sigma e^{-\frac{1+\varepsilon}{1+\sigma\varepsilon}\alpha(\theta - \theta')} d\theta \right. \\ &\quad \left. - g(\theta') \int_{\tilde{\theta}}^{\theta'} c(\theta)^\sigma f(\theta) d\theta \right] \end{aligned} \quad (59)$$

Suppose the marginal tax rate were to be bounded above strictly by some value $\tilde{\tau} \in (0, 1)$ as θ becomes large. It follows from consumer optimality that for all $\theta \in \Theta$ we have:

$$c(\theta)^\sigma e^{-\theta} (y(\theta) e^{-\theta})^{\frac{1}{\varepsilon}} > (1 - \tilde{\tau}) \quad (60)$$

From the (strict) bound $\tilde{\tau}$ places on the tax schedule and the fact that $c(\theta)$ and $y(\theta)$ will be unbounded in θ when $(1 - \tau)$ is bounded above zero we must additionally have for large enough θ :

$$c(\theta) > (1 - \tilde{\tau}) y(\theta)$$

Combining the last two objects gives:

$$c(\theta)^\sigma e^{-\theta} \left((1 - \tilde{\tau})^{-1} c(\theta) e^{-\theta} \right)^{\frac{1}{\varepsilon}} > (1 - \tilde{\tau}) \quad (61)$$

²⁶For the first of these we also need that the transversality restriction:

$$\lim_{\theta \rightarrow \infty} DC(\theta) f(\theta) < \infty$$

holds for all finite upper tax rates. This is straightforward to verify.

So:

$$\begin{aligned} c(\theta)^{\sigma+\frac{1}{\varepsilon}} &> (1-\tilde{\tau})^{1+\frac{1}{\varepsilon}} e^{\theta(1+\frac{1}{\varepsilon})} \\ c(\theta)^\sigma &> (1-\tilde{\tau})^{\sigma(\frac{1+\varepsilon}{1+\sigma\varepsilon})} e^{\theta\sigma(\frac{1+\varepsilon}{1+\sigma\varepsilon})} \end{aligned}$$

This gives us a bound on the first integral in (59):

$$\int_{\theta'}^{\infty} c(\theta)^\sigma e^{-\frac{1+\varepsilon}{1+\sigma\varepsilon}\alpha(\theta-\theta')} d\theta > (1-\tilde{\tau})^{\sigma(\frac{1+\varepsilon}{1+\sigma\varepsilon})} e^{\frac{1+\varepsilon}{1+\sigma\varepsilon}\alpha\theta'} \int_{\theta'}^{\infty} e^{-\frac{1+\varepsilon}{1+\sigma\varepsilon}(\alpha-\sigma)\theta} d\theta \quad (62)$$

But the integral on the right-hand side here is infinite whenever $\sigma > \alpha$. Since the second integral on the right-hand side of (59) is finite, and $[1-g(\theta')(1-F(\theta'))] > 0$ is true by assumption, we must have $\frac{\tau(\theta')}{1-\tau(\theta')} \rightarrow \infty$ as θ' gets large in order for the optimality condition to hold. But this contradicts taxes being bounded by $\tilde{\tau}$.