Hellwig, Martin

Working Paper

The undesirability of randomized income taxation under decreasing risk aversion

Preprints of the Max Planck Institute for Research on Collective Goods, No. 2005,27

Provided in Cooperation with:
Max Planck Institute for Research on Collective Goods

Suggested Citation: Hellwig, Martin (2005) : The undesirability of randomized income taxation under decreasing risk aversion, Preprints of the Max Planck Institute for Research on Collective Goods, No. 2005,27, Max Planck Institute for Research on Collective Goods, Bonn

This Version is available at:
https://hdl.handle.net/10419/26884

Preprints of the
Max Planck Institute for
Research on Collective Goods
Bonn 2005/27

The Undesirability of Randomized Income Taxation under Decreasing Risk Aversion

Martin Hellwig

The Undesirability of Randomized Income Taxation under Decreasing Risk Aversion*

Martin F. Hellwig
Max Planck Institute for Research on Collective Goods
Kurt-Schumacher-Str. 10, D-53113 Bonn, Germany
hellwig@coll.mpg.de

December 27, 2005

Abstract

For the standard specification of the utilitarian optimal income tax problem with hidden characteristics, the paper shows that randomized tax schemes are undesirable if preferences exhibit a property of weakly decreasing risk aversion according to the multidimensional risk aversion concept of Hellwig (2004). The property of decreasing risk aversion also implies uniqueness of the optimal income tax schedule and continuity in cases where the type distribution has a continuous density.

Key Words: Optimal Income Taxation, Randomized Incentive Schemes, Nonincreasing Risk Aversion

JEL Classification: H21

1 Introduction

One of the most puzzling results in the theory of optimal income taxation is the finding that randomized schemes are sometimes superior to deterministic schemes. In the utilitarian approach to optimal income taxation with inequality aversion, randomization of outcomes for low-productivity agents can alleviate incentive constraints for high-productivity agents to such an extent that the induced welfare gains from additional scope for redistribution outweigh the immediate welfare losses from the randomization.[1]

*I am grateful to Felix Bierbrauer, Christoph Engel, Thomas Gaube, and Ivan Werning for helpful discussions and comments. The usual disclaimer applies.

[1]See Stiglitz (1982, 1987) or Brito *et al.* (1995). The possibility that randomization in tax schemes may be desirable had already been suggested by Weiss (1976), but he assumed

The profession seems to have treated this finding as a curiosum rather than something to be taken into account in the design of tax policy. However, I have been unable to find any systematic assessment of its relevance. Brito *et al.* (1995) give a necessary and sufficient condition for the availability of local improvements through randomization, but do not discuss its relevance.

This paper shows that, in the standard income taxation model of Mirrlees (1971, 1976), randomization is in fact *undesirable* if participants' preferences exhibit a property of *weakly decreasing risk aversion*. The condition of Brito *et al.* involves *increasing* risk aversion. To the extent that increasing risk aversion is implausible, this paper provides a justification for the profession's focus on deterministic rather than randomized schemes for optimal income taxation.

The underlying idea is straightforward: If welfare gains from the additional scope for redistribution from high-productivity to low-productivity agents are to outweigh the low-productivity agents' immediate welfare losses from randomization, then the additional redistribution must provide the low-productivity agents with sufficient compensation for the risk to which they are exposed. However, the scope for additional redistribution depends on the high-productivity agents' attitudes towards the risks that are generated by the randomization. If the risk premium that they are willing to pay to avoid this risk exceeds the risk premium that the low-productivity agents require to accept this risk, then the randomization of outcomes for low-productivity agents can make the latter better off. If the reverse inequality holds, randomization of outcomes for low-productivity agents is undesirable.[2]

The different risk premia that high-productivity and low-productivity agents attach to a given randomized outcome depend on these agents' respective degrees of risk aversion. If high-productivity agents exhibit more risk aversion than low-productivity agents, the risk premia that they attach to given randomized outcomes exceed the corresponding risk premia for low-productivity agents; if they exhibit less risk aversion, the reverse is true. Therefore randomized schemes are undesirable if participants' preferences exhibit a property of decreasing risk aversion.

However, defining the notion of decreasing risk aversion for the income tax problem is not a trivial matter. The usual Arrow-Pratt measure of risk aversion is defined for unidimensional outcomes only. Here, outcomes are two-dimensional: people care about the level of consumption that they get

linear tax schedules and indicated that randomization would *not* be desirable if lump sum taxation was admissible.

[2]For similar arguments in other settings with hidden characteristics, see Maskin and Riley (1984), Matthews and Moore (1987).

and about the level of output, or the labour input, that they have to provide. Moreover, people with different earning abilities have different preference orderings over consumption/output pairs. Kihlstrom and Mirman's (1974) extension of the Arrow-Pratt approach to multiple dimensions can therefore not be used; their approach requires that all utility functions induce the same preference ordering on the underlying outcome space.

This paper relies on a new concept of decreasing risk aversion for multidimensional outcomes, which can be used even when the utility functions that are compared may induce different preference orderings on the underlying outcome space. According to this concept, for any good i, one agent is said to exhibit *i-premium-specific greater risk aversion* than another if, for every nondegenerate multidimensional lottery, he is willing to give up a greater number of units of good i in order to get rid of the randomness in the lottery. For details, the reader is referred to Hellwig (2004). Here, I use this concept by imposing an assumption of "weakly decreasing consumption-specific risk aversion". This means that, for any random pair of consumption and output levels, the risk premium in terms of units of the consumption good that an agent with high earning ability is willing to give up in order to eliminate the randomness is no greater than the risk premium that an agent with low earning ability is willing to give up. The concept is sensitive to the specification of the good in which the risk premium is paid. However, under the usual assumptions of the optimal-income tax model, the assumption that "consumption-specific risk aversion" is decreasing in productivity is strictly weaker than an assumption that "leisure specific" or "output specific" risk aversion is decreasing in productivity. Any of these assumptions is sufficient to ensure that randomization in income taxation is undesirable.

Simple though the idea of the paper is, the argument is somewhat involved. Because incentive compatibility conditions for randomized schemes are not well understood, a head-on analysis of optimal income taxation with randomization seems out of the question. Instead, I use an indirect approach and study the *relaxed income tax problem*, which is obtained if the requirement of incentive compatibility is weakened to *downward* incentive compatibility, i.e., the requirement that no person should have an incentive to pretend to have a lower productivity than he or she actually has. This indirect approach was pioneered by Matthews and Moore (1987) for a monopoly problem. In Hellwig (2005), a version of it is applied to the optimal income tax problem without the additional assumption of weakly decreasing consumption-specific risk aversion.

For the relaxed income tax problem, it is easy to show that randomization is undesirable if preferences exhibit decreasing consumption-specific

risk aversion. Under this assumption, one can also show that the relaxed income tax problem is equivalent to the optimal income tax problem in the sense that, up to modifications on a null set, any solution to one problem is also a solution to the other and *vice versa*. Together, these results imply that randomization is undesirable in the optimal income tax problem itself.

The key to the analysis lies in the finding that, if preferences exhibit weakly decreasing consumption-specific risk aversion, then the relaxed income tax problem can be formulated as a convex problem in the sense that the objective function is concave and the constraint set convex. This convexity property also implies that, up to modifications on null sets, the solution to the relaxed income tax problem is unique. Because the relaxed income tax problem and the optimal income tax problem are equivalent, the solution to the latter is also unique up to modifications on null sets.

In the following, Section 2 formulates the optimal income tax problem with randomization. Section 3 introduces and discusses the assumption of weakly decreasing consumption-specific risk aversion. Section 4 states the result that, under this assumption, randomization is undesirable and the solution to the optimal income tax problem is unique. Sections 5 - 7 provide the proof of this result. Supplementary arguments are given in the Appendices.

2 The Generalized Income Tax Problem

Following Mirrlees (1971, 1976), I study a large economy with one produced good and labour. Each agent in the economy is characterized by a productivity parameter n. An agent with productivity parameter n who consumes c units of the produced good and who supplies the labour needed to produce y units of output obtains the payoff $u(c, y, n)$. The leading example in the literature is the specification

$$u(c, y, n) = U(c, \frac{y}{n}), \tag{2.1}$$

which has the interpretation that n is labour productivity (the wage rate) and $\frac{y}{n}$ is the number of hours the person needs to work to produce the output y or to obtain the labour income y.

The productivity parameter n of any one person is the realization of a nondegenerate random variable $\tilde{n}$ with probability distribution F, with compact support $N \subset \Re_+$. Typically then, an individual's consumption level c and output provision level y will depend on his productivity parameter n. To allow for the possibility of randomization, I also admit the possibility that

c and y depend on the realization ω of a random variable $\tilde{\omega}$, with values in $[0,1]$. This random variable is assumed to be independent of $\tilde{n}$ so the joint distribution of $\tilde{n}$ and $\tilde{\omega}$ takes the form $F \times \nu$.

The randomization indicator ω, as well as the productivity parameter n, vary across agents. The probability distribution $F \times \nu$ of the random pair $(\tilde{n}, \tilde{\omega})$ is assumed to be the same for all agents. Moreover, by a large-numbers effect, the cross-section distribution of the pair (n, ω) in the population is assumed to be $F \times \nu$, almost surely.

An *allocation* is defined as a pair of functions, $(c(\cdot,\cdot), y(\cdot,\cdot))$, which specify for each $(n,\omega) \in N \times [0,1]$ a consumption level $c(n,\omega)$ and an output provision level $y(n,\omega)$. An allocation is said to be *deterministic* if there exist functions $\bar{c}(\cdot), \bar{y}(\cdot)$, defined on N, such that $(c(n,\omega), y(n,\omega)) = (\bar{c}(n), \bar{y}(n))$ for almost all $n \in N$ and $\omega \in [0,1]$. An allocation is *feasible* if

$$\int_N \int_0^1 c(n,\omega) d\nu(\omega) dF(n) \leq \int_N \int_0^1 y(n,\omega) d\nu(\omega) dF(n), \tag{2.2}$$

so that aggregate consumption does not exceed aggregate production. The allocation is *incentive-compatible* if

$$\int_0^1 u(c(n,\omega), y(n,\omega), n) d\nu(\omega) \geq \int_0^1 u(c(\hat{n},\omega), y(\hat{n},\omega), n) d\nu(\omega) \tag{2.3}$$

for all n and $\hat{n}$ in N, so that nobody has an incentive to claim that his productivity parameter is $\hat{n}$ when in fact it is n. Incentive compatibility is imposed because the value of each individual's productivity parameter is assumed to be the individual's private information. The individual's randomization indicator is not known to anybody at all, including the individual himself. When he considers what value of the productivity parameter to communicate to the system, his assessment of an alternative announcement $\hat{n}$ is determined by taking expectations of $u(c(\hat{n},\tilde{\omega}), y(\hat{n},\tilde{\omega}), n)$ with respect to the distribution ν of the random variable $\tilde{\omega}$.

Allocations are assessed according to the utilitarian welfare functional

$$\int_N \int_0^1 u(c(n,\omega), y(n,\omega), n) d\nu(\omega) dF(n). \tag{2.4}$$

The utilitarian welfare maximization problem is to maximize (2.4) over the set of feasible and incentive-compatible allocations.

I refer to this problem as the *generalized income tax problem*. By the *taxation principle* of Hammond (1979) and Guesnerie (1995), the problem of

maximizing (2.4) over the set of feasible and incentive-compatible *deterministic* allocations is equivalent to the problem of maximizing (2.4) by choosing an income tax schedule $T(\cdot)$ and then letting each person choose an output provision level y and a consumption level $c = y - T(y)$. The problem of maximizing (2.4) over the set of feasible and incentive-compatible *deterministic* allocations is therefore referred to as the *optimal income tax problem.* The generalized income tax problem is the same as the optimal income tax problem, except that allocations are *not* required to be deterministic.

A systematic analysis of the optimal income tax problem is provided in Hellwig (2005). In that paper, standard results about optimal income taxation are derived from the following assumptions, which will also be imposed here.

RMQ **Regularity, Monotonicity, and Quasiconcavity:** The utility function $u : \Re^3_+ \to \Re$ is twice continuously differentiable as well as increasing in c, decreasing in y, nondecreasing in n, and strictly quasiconcave in c and y.

PEP **Positivity of Efficient Production:** For all $n \in \Re_+$ and all v in the range of $u(\cdot, \cdot, n)$, there exists a unique pair $(c^*(n, v), y^*(n, v))$ which solves the problem of minimizing the difference $c - y$, subject to the constraint that $u(c, y, n) \geq v$; moreover, $y^*(n, v) > 0$.

SSCC **Strict Single-Crossing Condition:** The utility function satisfies

$$\frac{\partial}{\partial n}\left[\frac{u_c(c, y, n)}{|u_y(c, y, n)|}\right] > 0 \tag{2.5}$$

for all $(c, y, n) \in \Re^2_{++} \times \Re_+$.

To formulate the last assumption, I need some additional terminology: For any $n \in N$, the consumption/output pair (c, y) is said to be *efficient* for n if

$$(c, y) = (c^*(n, u(c, y, n)), y^*(n, u(c, y, n))), \tag{2.6}$$

where $c^*(\cdot, \cdot)$ and $y^*(\cdot, \cdot))$ are given by PEP; the pair (c, y) is said to be *distorted downward from efficiency* if

$$(c, y) << (c^*(n, u(c, y, n)), y^*(n, u(c, y, n))). \tag{2.7}$$

Further, I denote the minimum of N as n^0 and the maximum of N as n^1.

DR **Desirability of Redistribution: (a)** The utility function is concave in y and satisfies $u_{yn}(c,y,n) > 0$ for all $(c,y,n) \in \Re^3_+$. **(b)** For any $(c,y,n) \in \Re^2_+ \times [n^0, n^1)$, there exists $\varepsilon > 0$ such that $n + \varepsilon \in N$, and, for all $n' \in (n, n+\varepsilon]$ and all $(c',y') \in \Re^2_+$ satisfying $c' \geq c$ and $u(c',y',n') \geq u(c,y,n')$, the inequality

$$\left|u_y(c',y',n')\right| < |u_y(c,y,n)| \tag{2.8}$$

holds if (c,y) is efficient for n and (c',y') is efficient or distorted downward from efficiency for n'.

Assumptions RMQ, PEP, and SSCC are standard. PEP means that, for every n and and every attainable utility level v, the pair $(c^*(n,v), y^*(n,v))$, which minimizes the net resource requirement $c-y$ of providing a person with productivity parameter n with the utility level v, is well defined and unique; moreover, the efficient output level $y^*(n,v)$ is strictly positive. SSCC reflects the notion that the tradeoff between consumption and leisure results in a higher level of consumption and a lower level of leisure (more output provision) when the productivity parameter is higher. For utility functions satisfying RMQ, it is well known that SSCC implies the Milgrom-Shannon *Strict Single-Crossing Property*, namely, for all $(c,y), (\hat{c},\hat{y})$ in $\Re^2_+$ and all n in $\Re_+$, $(c,y) << (\hat{c},\hat{y})$ and $u(\hat{c},\hat{y},n) \geq u(c,y,n)$ imply $u(\hat{c},\hat{y},n') > u(c,y,n')$ for all $n' > n$.[3]

Condition DR provides the rationale for utilitarian redistribution. If an allocation is such that, for some n and n' in N, people with productivity parameter n receive the outcome pair (c,y) and people with productivity parameter n' receive the outcome pair (c',y'), where (c,y) and (c',y') satisfy the inequality (2.8), then people with productivity parameter n' have a lower marginal disutility from providing additional output than people with productivity parameter n. The sum of these people's utilities is therefore increased if output requirements are redistributed from people with productivity parameter n to people with productivity parameter n'. If incentive considerations are not standing in the way, the utilitarian welfare maximizer would like to redistribute output requirements so as to allow the people with productivity parameter n to produce less and to require the people with productivity parameter n' to produce more.

Part (a) of condition DR postulates this desirability of redistributing output requirements whenever $c = c'$ and $y \geq y'$. Part (b) postulates the

[3] On the relation between SSCC and the Milgrom-Shannon single-crossing property, see Milgrom and Shannon (1994), and Edlin and Shannon (1998).

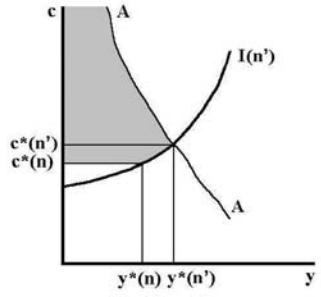

Figure 1: The set of outcome pairs (c', y') for n' for which part (b) of DR postulates (2.8).

desirability of redistributing output requirements whenever (c, y) is efficient for n, (c', y') is efficient or distorted downward from efficiency for n', and $n' > n$ is close to n; for an illustration, see Figure 1, which, for a given pair $(c, y) = (c^*(n), y^*(n))$, which is efficient for n, exhbibits the set of pairs (c', y') to which part (b) of condition DR applies as the area to the left of the line $A - A$ of outcomes that are efficient for type n' and above the horizontal through $(c^*(n), y^*(n))$.

For the deterministic optimal income tax problem, Hellwig (2005) shows that, in combination with RMQ, PEP, and SSCC, these postulates are sufficient to guarantee that, if the support N of the distribution F is a finite set or if N is an interval and F has a density that is continuous on N, then, in any optimal allocation, consumption/output pairs are distorted downward from efficiency for all values of the productivity parameter other than the maximum and, possibly, the minimum of N; in other words, the optimal marginal income tax is strictly positive at all income levels other than the highest and, possibly, the lowest in the population. Hellwig (2005) also shows that condition DR is weaker than the conditions that have previously been used to establish the posivity of the the optimal marginal income tax. In particular, for utility functions taking the special form (2.1), DR is strictly weaker than the familiar requirements that u be strictly concave and that leisure be a noninferior good.

In the remainder of the paper, I impose conditions RMQ, PEP, SSCC, and DR without further mention.

3 Decreasing Risk Aversion

In addition to RMQ, PEP, SSCC, and DR, I will impose the following new assumption which concerns risk attitudes and their dependence on the hidden characteristic n.

WDCRA **Weakly decreasing consumption-specific risk aversion:** For any n, any $\hat{n} < n$, any pair of nondegenerate nonnegative-valued random variables $(\tilde{c}, \tilde{y})$, and any $\pi \in \Re$,

$$Eu(\tilde{c}, \tilde{y}, \hat{n}) = u(E\tilde{c} - \pi, E\tilde{y}, \hat{n}) \tag{3.1}$$

implies[4]

$$\pi > 0 \tag{3.2}$$

and

$$Eu(\tilde{c}, \tilde{y}, n) \geq u(E\tilde{c} - \pi, E\tilde{y}, n). \tag{3.3}$$

Thus, if a person with productivity parameter $\hat{n}$ is indifferent between a random pair $(\tilde{c}, \tilde{y})$ and a nonrandom pair $(\bar{c}, \bar{y})$ with $\bar{c} = E\tilde{c} - \pi$ and $\bar{y} = E\tilde{y}$, then, under WDCRA, *no* person with productivity parameter $n > \hat{n}$ prefers the nonrandom pair $(\bar{c}, \bar{y})$ to the random pair $(\tilde{c}, \tilde{y})$. If the premium $\pi = E\tilde{c} - \bar{c}$ is enough to compensate a person with productivity parameter $\hat{n}$ for accepting the risk inherent in $(\tilde{c}, \tilde{y})$, then this premium is also enough to compensate a person with a higher productivity parameter for accepting this risk.

The assumption of weakly decreasing consumption-specific risk aversion bears some similarity to the assumption of nonincreasing absolute risk aversion in models with unidimensional outcomes. Both assumptions assert that individuals who are better off are better able to bear risk and are less willing to pay a risk premium to get rid of risk. In the traditional unidimensional setting, the monotonicity condition relates risk aversion to consumption (or wealth); here it relates risk aversion to earning ability. However, the heuristic justification is similar.

The following characterization of weakly decreasing consumption-specific risk aversion will be useful in the subsequent analysis.

Theorem 3.1 *Under RMQ and WDCRA, the following statements are equivalent:*

[4]Positivity of risk premia presumes that u is strictly concave in c and y. This is *not* implied by DR.

(a) u exhibits weakly decreasing consumption-specific risk aversion;

(b) u is strictly concave in c and y, and, for any n and any $\hat{n} < n$, there exists a concave function $\varphi^n_{\hat{n}}(\cdot,\cdot)$ such that for all $(c,y) \in \Re^2_+$,

$$u(c,y,\hat{n}) = \varphi^n_{\hat{n}}(u(c,y,n),y); \tag{3.4}$$

(c) the matrix

$$A(c,y|n) = -\frac{1}{u_c(c,y,n)} \begin{pmatrix} u_{cc}(c,y,n) & u_{cy}(c,y,n) \\ u_{yc}(c,y,n) & u_{yy}(c,y,n) \end{pmatrix} \tag{3.5}$$

is positive definite and weakly decreasing in n in the sense that, for every $(c,y) \in \Re^2_+$, any n and any $\hat{n} < n$, the matrix

$$B(c,y|n,\hat{n}) = A(c,y|n) - A(c,y|\hat{n}) \tag{3.6}$$

is negative semidefinite;

(d) if $c(\cdot,\cdot,\cdot)$ is defined so that for any y and n, $c(\cdot,y,n)$ is the inverse of the section $u(\cdot,y,n)$ of u that is determined by y and n, then, for any n, the function $(v,y) \to u_n(c(v,y,n),y,n)$ is convex.

The equivalence of statements (c) and (d) in Theorem 3.1 has been observed by Mirrlees (1986, p. 1231 f.) in an analysis designed to establish the sufficiency of first-order conditions for the optimal-income tax problem, however, without reference to the interpretation of (3.5) as a local measure of risk aversion. The equivalence of statements (a) - (c) in Theorem 3.1 is a special case of Theorem 2.1 in Hellwig (2004), which provides a characterization of i-premium-specific weakly decreasing risk aversion in a general model with m outcome dimensions. The characterization extends the one given by Pratt (1964) for nonincreasing risk aversion in the unidimensional case. In particular, the equivalence of statements (a) and (c) in the theorem relates the global property of consumption-specific risk aversion being weakly decreasing in n to a monotonicity property of a local measure of the relative curvature of the section $u(\cdot,\cdot,n)$ of u that is determined by n.

The reader may wonder why I work with a concept of decreasing risk aversion that presumes that risk premia are paid in units of the consumption good. Why not allow for risk premia that are paid by providing additional output? An assumption of weakly decreasing output-specific risk aversion would require that for any n, any $\hat{n} < n$, any pair of nonnegative-valued random variables $(\tilde{c},\tilde{y})$, and any $\pi \in \Re$,

$$Eu(\tilde{c},\tilde{y},\hat{n}) = u(E\tilde{c}, E\tilde{y} + \pi, \hat{n}) \tag{3.7}$$

implies

$$\pi > 0 \tag{3.8}$$

and

$$Eu(\tilde{c}, \tilde{y}, n) \geq u(E\tilde{c}, E\tilde{y} + \pi, n). \tag{3.9}$$

However, under SSCC, this monotonicity condition on output-specific risk premia is actually *stronger* than WDCRA.

Remark 3.2 *Assume RMQ and SSCC. Then, if u exhibits weakly decreasing output-specific risk aversion, u also exhibits weakly decreasing consumption-specific risk aversion.*

Proof. By Theorem 2.1 in Hellwig (2004), u exhibits weakly decreasing output-specific risk aversion if and only if u is strictly concave in c and y, and the matrix

$$\frac{u_c(c, y, n)}{|u_y c, y, n)|} A(c, y|n) - \frac{u_c(c, y, \hat{n})}{|u_y c, y, \hat{n})|} A(c, y|\hat{n}) \tag{3.10}$$

is negative semidefinite whenever $\hat{n} < n$. (3.10) can be rewritten as

$$\left[\frac{u_c(c, y, n)}{|u_y c, y, n)|} - \frac{u_c(c, y, \hat{n})}{|u_y c, y, \hat{n})|}\right] A(c, y|n) + \frac{u_c(c, y, \hat{n})}{|u_y c, y, \hat{n})|} B(c, y|n, \hat{n}), \tag{3.11}$$

where $B(c, y|n, \hat{n})$ is given by (3.6). Strict concavity of u implies that the the matrix $A(c, y|n)$ is positive definite. By SSCC, it follows that the first term in (3.11) is positive definite. Negative semidefiniteness of (3.10) and (3.11) therefore implies the matrix $B(c, y|n, \hat{n})$ is also negative semidefinite. The remark follows from the equivalence of statements (a) and (c) in Theorem 3.1. ■

Theorem 3.1 also implies that if the utility function u satisfies WDCRA, as well as RMQ, then, for any two types n and $\hat{n} < n$, the matrix

$$\begin{aligned} &\frac{u_c(c(\hat{n}), y(\hat{n}), n)}{|u_c(c(\hat{n}), y(\hat{n}), n) + u_y(c(\hat{n}), y(\hat{n}), n)|} A(c(\hat{n}), y(\hat{n})|n) \\ &- \frac{u_c(c(\hat{n}), y(\hat{n}), \hat{n})}{|u_c(c(\hat{n}), y(\hat{n}), \hat{n}) + u_y(c(\hat{n}), y(\hat{n}), \hat{n})|} A(c(\hat{n}), y(\hat{n})|\hat{n}) \end{aligned}$$

is negative semidefinite, if $u_c(c(\hat{n}), y(\hat{n}), \hat{n}) + u_y(c(\hat{n}), y(\hat{n}), \hat{n}) > 0$. Under WDCRA and RMQ, therefore, the utility function u violates the condition that Theorem (III) of Brito et al. (1995) identified as being necessary and

sufficient for randomization to provide a local improvement over the best deterministic allocation in a model involving exactly two types.[5] Their condition is thus incompatible with WDCRA. In the following, I will go further and show that, under WDCRA, there is no possibility at all for improving on the best deterministic allocation, whether through a large, non-local change or through a small, local change in the allocation.

4 The Main Theorem

The generalized income tax problem, as well as the optimal income tax problem, have multiple solutions because a modification of an allocation on a null set has no effects on the value of the welfare functional (2.4) or on the validity of the feasibility constraint (2.2); some such modifications are possible without upsetting incentive compatibility. In order to avoid having to account for this trivial multiplicity of solutions, I will say that two allocations are *equivalent* if, up to modifications on a null set in terms of the measure $F \times \nu$, they are the same. If all solutions to an income tax problem are equivalent, the problem is said to have an *essentially unique* solution. Two income tax problems are said to be *equivalent* if, for any solution to one problem, there exists an equivalent solution to the other, and *vice versa*.

With this terminology, the main results of the paper are stated as follows.

Theorem 4.1 *Under WDCRA, the solution to the generalized income tax problem is essentially unique and deterministic.*

Corollary 4.2 *Under WDCRA, the generalized income tax problem and the optimal income tax problem are equivalent. The solution to the optimal income tax problem is essentially unique.*

If the utility function exhibits weakly decreasing consumption-specific risk aversion, then, in the generalized income tax problem, it is undesirable to have $(c(n,\omega), y(n,\omega))$ depend in any substantial way on ω. The rationale for this finding has been sketched in the introduction: Conditions DR and SSCC imply a desire to redistribute consumption from people with high values of the productivity parameter to people with low values of the productivity parameter. Limits to this redistribution arise from the constraint that high-productivity people should not have an incentive to claim that,

[5] The statement of their result only says "if", but the surrounding text and the proof make clear that they mean "if and only if".

in fact, they have low productivity. Weakly decreasing risk aversion implies that this incentive constraint is relaxed rather than tightened if a random allocation is replaced by a nonrandom one with the same expected output levels and the same expected payoffs. By RMQ, the feasibility constraint is also unaffected by this change, and the level of welfare is actually increased.

If one tries to put this reasoning into a formal proof, one encounters the difficulty that most of the argument relies on insights about the deterministic optimal income tax problem. These insights have been gained by an analysis relying on the necessary and sufficient conditions for incentive compatibility in the deterministic context. For allocations involving randomization, no such characterization of incentive compatibility in terms of analytically tractable necessary and sufficient conditions is available. Therefore a proof of Theorem 4.1 by a direct characterization of the conditions for optimality is not available.

Given this difficulty, I use an indirect approach. Instead of the optimal income tax problem and the generalized income tax problem, I study the - deterministic and generalized - *relaxed* income tax problems, which result if incentive compatibility constraints are replaced by *downward* incentive compatibility constraints. I will show that, under WDCRA, the solution to the generalized relaxed income tax problem is unique up to modifications on null sets, and that it is deterministic. The reason is, essentially, that the objective function of the generalized relaxed income tax problem is strictly concave, and the constraint set is convex. An immediate consequence is that the generalized relaxed income tax problem and the deterministic relaxed income tax problem are equivalent.

Thereafter, I will show that, under WDCRA, the deterministic relaxed income tax problem and the optimal income tax problem are equivalent. The argument is first developed under the additional assumption that the support of the type distribution is finite. For this case, Hellwig (2005) shows that, under RMQ, PEP, SSCC, and DR, the optimal income tax problem is equivalent to the deterministic *weakly* relaxed income tax problem, which is defined as the problem of maximizing (2.4) over the set of deterministic allocations that satisfy feasibility, downward incentive compatibility, *and* the requirement that consumption be nondecreasing in n.[6] To make use of this finding, I will show that, if the support of the type distribution is finite, then, under the additional assumption of weakly decreasing consumption-specific risk aversion, the solution to the deterministic relaxed income tax

[6] The result in Hellwig (2005) also covers the case of a type distribution with a continuous density function.

problem automatically satisfies the monotonicity requirement on consumption. The deterministic relaxed income tax problem and the deterministic weakly relaxed income tax problem are thus equivalent; by the result in Hellwig (2005), it follows that the deterministic relaxed income tax problem and the optimal income tax problem are also equivalent.

The final step of the analysis shows that the equivalence of the deterministic relaxed and the optimal income tax problems extends from type distributions with finite supports to arbitrary type distributions. The argument is based on standard continuity considerations. As a by-product of these considerations, one obtains the following continuity property of optimal allocations.

Theorem 4.3 *Assume WDCRA, let* $\{F^k\}$ *be a sequence of type distributions with uniformly bounded supports, and suppose that the distributions* F^k *converge setwise to a distribution* F*, i.e., suppose that* $F^k(B)$ *converges to* $F(B)$ *for every measurable set* B*. For any* k*, let* $(\bar{c}^k(\cdot), \bar{y}^k(\cdot))$ *be the solution to the optimal income tax problem when the type distribution is* F^k*, and let* $(\bar{c}(\cdot), \bar{y}(\cdot))$ *be the solution when the type distribution is* F*. Then*

$$\lim_{k\to\infty} (\bar{c}^k(n), \bar{y}^k(n)) = (\bar{c}(n), \bar{y}(n)) \tag{4.1}$$

for F*-almost every* n*; (4.1) holds, in particular, for every* n *that is either a continuity point of* $(\bar{c}(\cdot), \bar{y}(\cdot))$ *or a mass point of* F*.*

For ease of exposition, I present the three parts of the argument in separate sections. Throughout the analysis, I impose WDCRA, as well as the other assumptions, without further mention.

5 The generalized relaxed income tax problem

The *generalized relaxed income tax problem* is to maximize (2.4) over the set of allocations that are feasible and *downward incentive-compatible* in the sense that (2.3) holds for all n and $\hat{n}$ in N such that $\hat{n} < n$. For this problem, one obtains:

Proposition 5.1 *Under WDCRA, the solution to the generalized relaxed income tax problem is essentially unique and deterministic.*

The proof of Proposition 5.1 is divided into several lemmas. The first lemma shows that, under WDCRA, a randomized allocation cannot do bet-

ter than the best deterministic allocation. Indeed, for any feasible and downward incentive-compatible allocation that involves randomization, there exists a deterministic allocation that generates the same welfare and is also feasible and downward incentive-compatible.

Lemma 5.2 *Let $(c(\cdot,\cdot), y(\cdot,\cdot))$ be any feasible and downward incentive-compatible allocation, and let $(\bar{c}(\cdot), \bar{y}(\cdot))$ be such that*

$$\bar{y}(n) = \int_0^1 y(n,\omega)d\nu(\omega), \tag{5.1}$$

and

$$u(\bar{c}(n), \bar{y}(n), n) = \int_0^1 u(c(n,\omega), y(n,\omega), n)d\nu(\omega) \tag{5.2}$$

for all n. Then the deterministic allocation that corresponds to the pair $(\bar{c}(\cdot), \bar{y}(\cdot))$ is also feasible and downward incentive-compatible and generates the same welfare as the original allocation $(c(\cdot,\cdot), y(\cdot,\cdot))$. Moreover, unless the original allocation itself is deterministic, the deterministic allocation satisfies the feasibility constraint with slack, and one has

$$\int_N [\bar{y}(n) - \bar{c}(n)]dF(n) > 0. \tag{5.3}$$

Proof. Let $(\bar{c}(\cdot,\cdot), \bar{y}(\cdot,\cdot))$ be such that $(\bar{c}(n,\omega), \bar{y}(n,\omega)) = (\bar{c}(n), \bar{y}(n))$ for almost all $n \in N$ and $\omega \in [0,1]$. From (5.2) and (2.4), one immediately sees that the allocations $(c(\cdot,\cdot), y(\cdot,\cdot))$ and $(\bar{c}(\cdot,\cdot), \bar{y}(\cdot,\cdot))$ generate the same welfare. Under WDCRA, (5.2) also yields

$$u(\bar{c}(n), \bar{y}(n), n') \leq \int_0^1 u(c(n,\omega), y(n,\omega), n')d\nu(\omega) \tag{5.4}$$

for any n and any $n' > n$. Upon combining (5.4) with the downward incentive compatibility condition

$$\int_0^1 u(c(n,\omega), y(n,\omega), n')d\nu(\omega) \leq \int_0^1 u(c(n',\omega), y(n',\omega), n')d\nu(\omega) \tag{5.5}$$

for the original allocation, one finds that

$$u(\bar{c}(n), \bar{y}(n), n') \leq \int_0^1 u(c(n',\omega), y(n',\omega), n')d\nu(\omega) = u(\bar{c}(n'), \bar{y}(n'), n') \tag{5.6}$$

for any n and any $n' > n$. This shows that the allocation $(\bar{c}(\cdot,\cdot), \bar{y}(\cdot,\cdot))$ is also downward incentive-compatible.

Because u is concave, (5.2) and (5.1) imply

$$\bar{c}(n) \leq \int_0^1 c(n,\omega)d\nu(\omega) \tag{5.7}$$

for all n. In combination with (5.1) and the feasibility of the original allocation, (5.7) implies

$$\int_N \bar{c}(n)dF(n) \leq \int_N \bar{y}(n)dF(n). \tag{5.8}$$

This shows that the allocation $(\bar{c}(\cdot,\cdot), \bar{y}(\cdot,\cdot))$ is also feasible. Indeed, because u is strictly concave, the inequalities in (5.7) and (5.8) are strict unless the original allocation $(c(\cdot,\cdot), y(\cdot,\cdot))$ is equivalent to $(\bar{c}(\cdot,\cdot), \bar{y}(\cdot,\cdot))$. ■

Given this lemma, I turn to the *deterministic relaxed income tax problem*. This is the problem of maximizing (2.4) over the set of deterministic, feasible, and downward incentive-compatible allocations. It is equivalent to the problem of choosing a payoff function $v(\cdot)$ and an output provision function $\bar{y}(\cdot)$ so as to maximize the welfare functional

$$\int_N v(n)dF(n) \tag{5.9}$$

subject to the feasibility condition

$$\int_N [\bar{y}(n) - c(v(n), \bar{y}(n), n)]dF(n) \geq 0 \tag{5.10}$$

and the downward incentive compatibility condition that

$$v(n) \geq u(c(v(\hat{n}), \bar{y}(\hat{n}), \hat{n}), \bar{y}(\hat{n}), n), \tag{5.11}$$

for all n and $\hat{n}$ in N such that $\hat{n} < n$. For any y and n, the function $c(\cdot, y, n)$ in (5.10) and (5.11) is the inverse of the section $u(\cdot, y, n)$ of u that is determined by y and n. A pair $(v(\cdot), \bar{y}(\cdot))$ that satisfies the constraints of this problem is said to be *admissible.*

The following lemma shows that WDCRA turns the generalized relaxed income tax problem into a convex problem.[7]

[7] I am grateful to Ivan Werning for suggesting that the assumption which rules out randomization should also turn the optimal income tax problem into a convex problem. For models with a continuum of types, Lemma 5.3 is closely related to the result of Mirrlees (1986) that if the curvature property of statement (c) in Theorem 3.1 is satisfied, then $u_n(c(v,y,n),y,n)$ is convex in v and y.

Lemma 5.3 *The set of admissible pairs $(v(\cdot), \bar{y}(\cdot))$ for the deterministic relaxed income tax problem is convex. Moreover, for any two admissible pairs $(v_i(\cdot), \bar{y}_i(\cdot)), i = 1, 2$, and any $\delta \in (0, 1)$, the convex combination*

$$(v_\delta(\cdot), \bar{y}_\delta(\cdot)) := \delta(v_1(\cdot), y_1(\cdot)) + (1 - \delta)(v_2(\cdot), y_2(\cdot)) \tag{5.12}$$

satisfies (5.10) with strict inequality unless $(v_1(n), \bar{y}_1(n)) = (v_2(n), \bar{y}_2(n))$ for F-almost all $n \in N$.

Proof. I use condition WDCRA to reformulate of the downward incentive compatibility constraint (5.11): For any n and any $\hat{n} < n$, let $\varphi^n_{\hat{n}}(\cdot, \cdot)$ be given by (3.4). By RMQ, $\varphi^n_{\hat{n}}(\cdot, \cdot)$ is increasing in the first argument. (5.11) is therefore equivalent to the requirement that

$$\begin{aligned} \varphi^n_{\hat{n}}(v(n), \bar{y}(\hat{n})) &\geq \varphi^n_{\hat{n}}(u(c(v(\hat{n}), \bar{y}(\hat{n}), \hat{n}), \bar{y}(\hat{n}), n), \bar{y}(\hat{n})) \\ &= u(c(v(\hat{n}), \bar{y}(\hat{n}), \hat{n}), \bar{y}(\hat{n}), \hat{n}). \end{aligned} \tag{5.13}$$

By the definition of $c(v(\hat{n}), \bar{y}(\hat{n}), \hat{n})$, the right-hand side of (5.13) is just $v(\hat{n})$, so (5.11) is equivalent to the inequality

$$\varphi^n_{\hat{n}}(v(n), \bar{y}(\hat{n})) \geq v(\hat{n}). \tag{5.14}$$

Condition WDCRA and Theorem 3.1 imply that, for any n and any $\hat{n} < n$, the function $\varphi^n_{\hat{n}}(\cdot, \cdot)$ is concave. The set of pairs $(v(\cdot), \bar{y}(\cdot))$ which satisfy (5.14) for all n and all $\hat{n} < n$ is therefore convex.

Turning to feasibility, by RMQ and WDCRA, the function $c(\cdot, \cdot, n)$ is strictly convex. For any admissible pairs $(v_i(\cdot), \bar{y}_i(\cdot)), i = 1, 2$, any $\delta \in (0, 1)$, and $(v^A_\delta(n), \bar{y}^A_\delta(n))$ given by (5.12), one therefore has

$$\begin{aligned} &\int_N [y_\delta(n) - c(v_\delta(n), \bar{y}_\delta(n), n)] dF(n) \\ \geq\ & \delta \int_N [\bar{y}_1(n) - c(v_1(n), \bar{y}_1(n), n)] dF(n) \\ &+ (1 - \delta) \int_N [\bar{y}_2(n) - c(v_2(n), \bar{y}_2(n), n)] dF(n). \end{aligned} \tag{5.15}$$

Thus, $(v_\delta(\cdot), \bar{y}_\delta(\cdot))$ satisfies (5.10) if $(v_1(\cdot), \bar{y}_1(\cdot))$ and $(v_2(\cdot), \bar{y}_2(\cdot))$ do. Moreover, unless $(v_1(n), \bar{y}_1(n)) = (v_2(n), \bar{y}_2(n))$ for F-almost all $n \in N$, the inequality in (5.15) is strict, and $(v_\delta(\cdot), \bar{y}_\delta(\cdot))$ satisfies (5.10) with slack. ■

The following lemma shows that an allocation that involves slack in the feasibility constraint is always dominated. Any slack can be used to raise the consumption of each type. Moreover, this can be done without upsetting downward incentive compatibility.

Lemma 5.4 *Let $(v(\cdot), y(\cdot))$ be admissible for the deterministic relaxed income tax problem and suppose that $(v(\cdot), y(\cdot))$ satisfies the feasibility condition (5.10) with strict inequality. Then there exists another admissible pair $(\bar{v}(\cdot), \bar{y}(\cdot))$ that satisfies $\bar{v}(n) > v(n)$ for all $n \in N$.*

Proof. Given the pair $(v(\cdot), y(\cdot))$, let v_0 be such that $v_0 > v(n_1)$, where $n_1 = \max N$ is the highest productivity type in the population. Consider the pair $(\hat{v}(\cdot), \hat{y}(\cdot))$, where $\hat{y}(n) = 0$ for all n, and $\hat{v}(\cdot)$ is the solution to the differential equation

$$\hat{v}'(n) = u_n(c(\hat{v}(n), 0, n), 0, n) \tag{5.16}$$

with initial condition $\hat{v}(n_0) = v_0$, where $n_0 = \min N$ is the lowest productivity type in the population. Because $\hat{v}(n_0) = v_0 > v(n_1)$ and $\hat{v}(\cdot)$ and $v(\cdot)$ are both nondecreasing, one has $\hat{v}(n) > v(n)$ for all $n \in N$. Moreover, because $\hat{y}(\cdot)$ is nondecreasing and $\hat{v}(\cdot)$ satisfies (5.16), an argument of Mirrlees (1976) implies that the allocation $(c(\hat{v}(\cdot), 0, \cdot), \hat{y}(\cdot))$ is incentive-compatible, hence downward incentive-compatible. Thus $(v(\cdot), y(\cdot))$ and $(\hat{v}(\cdot), \hat{y}(\cdot))$ both satisfy (5.14) for all n and all $\hat{n} < n$. Because the function $\varphi^n_{\hat{n}}(\cdot, \cdot)$ is concave, it follows that, for any $\delta \in (0, 1)$, the pair $\delta(\hat{v}(\cdot), \hat{y}(\cdot)) + (1-\delta)(v(\cdot), y(\cdot)$ also satisfies (5.14) for all n and all $\hat{n} < n$. Moreover, if $\delta > 0$ is sufficiently small, the pair

$$(\bar{v}(\cdot), \bar{y}(\cdot)) := \delta(\hat{v}(\cdot), \hat{y}(\cdot)) + (1-\delta)(v(\cdot), y(\cdot))$$

satisfies the feasibility condition (5.10); for such δ, the additional consumption under the allocation

$$(c(\delta(\hat{v}(\cdot), \hat{y}(\cdot)) + (1-\delta)(v(\cdot), y(\cdot)), \cdot), \delta\hat{y}(\cdot) + (1-\delta)y(\cdot))$$

just takes up the slack left by the allocation $(v(\cdot), y(\cdot))$. ■

Lemmas 5.2 and 5.4 together imply that the generalized relaxed income tax problem cannot have a nondeterministic solution. Lemmas 5.3 and 5.4 together imply that, up to modifications on null sets, any two solutions to the deterministic relaxed income tax problem must be the same. Proposition 5.1 follows immediately.

To get from Proposition 5.1 to Theorem 4.1, I need to show that the generalized relaxed income tax problem and the generalized income tax problem are equivalent. This is done in the following two sections. The argument makes repeated use of the following elementary lemma:

Lemma 5.5 *If the solution to the problem of maximizing (2.4) over a constraint set $\bar{\mathcal{C}}$ is essentially unique and if this solution belongs to a subset $\mathcal{C}$ of $\bar{\mathcal{C}}$, then the problem of maximizing (2.4) over $\mathcal{C}$ and the problem of maximizing (2.4) over $\bar{\mathcal{C}}$ are equivalent.*

Proof. Trivially, any solution to the problem of maximizing (2.4) over the set $\bar{\mathcal{C}}$ must also be a solution to the problem of maximizing (2.4) over the smaller set $\mathcal{C}$. Because any other solution to the problem of maximizing (2.4) over $\mathcal{C}$ must yield the same value of the objective function, it follows that any such solution is also a solution to the problem of maximizing (2.4) over $\bar{\mathcal{C}}$. If the solution to the latter problem is essentially unique, it follows that the solution to the former problem is also essentially unique. The lemma follows immediately. ■

The constraint sets $\mathcal{C}_R, \mathcal{C}_G, \mathcal{C}_{DR}, \mathcal{C}_O$ of the generalized relaxed, generalized, deterministic relaxed, and optimal income tax problems satisfy the inclusion relations

$$\mathcal{C}_O \subset \mathcal{C}_G \subset \mathcal{C}_R \tag{5.17}$$

and

$$\mathcal{C}_O \subset \mathcal{C}_{DR} \subset \mathcal{C}_R. \tag{5.18}$$

Proposition 5.1 implies that, for $\bar{\mathcal{C}} = \mathcal{C}_R$ and $\mathcal{C} = \mathcal{C}_{DR}$, the premises of Lemma 5.5 are always fulfilled. The generalized relaxed income tax problem and the deterministic relaxed income tax problem are therefore equivalent.

For $\bar{\mathcal{C}} = \mathcal{C}_R$ and $\mathcal{C} = \mathcal{C}_O$, the premises of Lemma 5.5 are also fulfilled if the solution to the generalized relaxed income tax problem is incentive-compatible. In this case, obviously, the premises of Lemma 5.5 are also fulfilled if one sets $\bar{\mathcal{C}} = \mathcal{C}_R$ and $\mathcal{C} = \mathcal{C}_G$. These considerations yield:

Lemma 5.6 *If the solution to the deterministic relaxed income tax problem is incentive-compatible, then the generalized relaxed, generalized, deterministic relaxed, and optimal income tax problems are all equivalent.*

The problem thus is to show that the solution to the deterministic relaxed income tax problem is incentive-compatible.

6 The Case of Finitely Many Types

This section proves the desired result under the additional assumption that the support of the type distribution is finite. Without loss of generality, I

write $N = \{n_1, n_2, ..., n_m\}$ with $n_1 < ... < n_m$. A deterministic allocation is identified with a sequence $\{(\bar{c}_i, \bar{y}_i)\}_{i=1}^m$ such that $(\bar{c}_i, \bar{y}_i) = (\bar{c}(n_i), \bar{y}(n_i))$ for $i = 1, ..., m$. The associated utility levels are $v_i = u(\bar{c}_i, \bar{y}_i, n_i)$, the associated efficient consumption/output pairs are $(c_i^*, y_i^*) = (c^*(v_i, n_i), y^*(v_i, n_i))$. The following proposition characterizes the solution to the generalized relaxed income tax problem.

Proposition 6.1 *If $N = \{n_1, n_2, ..., n_m\}$, with $n_1 < ... < n_m$, the solution $\{(\bar{c}_i, \bar{y}_i)\}_{i=1}^m$ to the deterministic relaxed income tax problem satisfies the following:*

a: There is no distortion at the top, i.e.,

$$(\bar{c}_m, \bar{y}_m) = (c_m^*, y_m^*). \tag{6.1}$$

b: Below the top, i.e., for $i = 1, ..., m-1$, consumption/output pairs are distorted downwards from efficiency, i.e.,

$$(\bar{c}_i, \bar{y}_i) << (c_i^*, y_i^*); \tag{6.2}$$

the adjacent downward incentive compatibility constraint for type n_{i+1} is strictly binding, i.e.,

$$u(\bar{c}_{i+1}, \bar{y}_{i+1}, n_{i+1}) = u(\bar{c}_i, \bar{y}_i, n_{i+1}). \tag{6.3}$$

c: Consumption and required output levels are nondecreasing and co-monotonic on N, i.e., for any i and $j > i$, $\bar{y}_j > \bar{y}_i$ if $\bar{c}_j > \bar{c}_i$, and $\bar{y}_j = \bar{y}_i$ if $\bar{c}_j = \bar{c}_i$.

d: The difference $\bar{y}_i - \bar{c}_i$ is co-monotonic with $\bar{y}_i$ and $\bar{c}_i$. Moreover, there exists $\hat{\imath}$ such that $\bar{y}_i - \bar{c}_i > 0$ for $i > \hat{\imath}$ and $\bar{y}_i - \bar{c}_i < 0$ for $i < \hat{\imath}$; in particular, $\bar{c}_1 > \bar{y}_1 \geq 0$.

By SSCC and the inequality $(\bar{c}_i, \bar{y}_i) \leq (\bar{c}_{i+1}, \bar{y}_{i+1})$, (6.3) implies that one also has

$$u(\bar{c}_{i+1}, \bar{y}_{i+1}, n_k) \leq u(\bar{c}_i, \bar{y}_i, n_k) \tag{6.4}$$

for all $k < i + 1$. Upon applying this inequality repeatedly, with fixed k and $i = k, k+1, ..., j$, one infers that

$$u(\bar{c}_j, \bar{y}_j, n_k) \leq u(\bar{c}_k, \bar{y}_k, n_k) \tag{6.5}$$

whenever $j > k$. The solution to the generalized relaxed income tax problem is thus upward, as well as downward, incentive-compatible. This observation yields:

Corollary 6.2 *If the support of the type distribution is finite, the solution to the deterministic relaxed income tax problem is incentive-compatible.*

By Lemma 5.6, one further obtains:

Corollary 6.3 *If the support of the type distribution is finite, the generalized relaxed, deterministic relaxed, generalized, and optimal income tax problems are all equivalent.*

The proof of Proposition 6.1 relies on a related result in Hellwig (2005). For the case where $N = \{n_1, n_2, ..., n_m\}$, with $n_1 < ... < n_m$, Theorem 3.1 in Hellwig (2005) shows that an allocation $\{(\bar{c}_i, \bar{y}_i)\}_{i=1}^m$ satisfies statements (a) - (d) of Proposition 6.1 if it is a solution to the deterministic *weakly relaxed income tax problem* of maximizing (2.4) over the set $\mathcal{C}_{WR}$ of allocations in $\mathcal{C}_{DR}$ that satisfy the consumption monotonicity condition

$$\bar{c}_i \leq \bar{c}_{i+1} \tag{6.6}$$

for $i = 1, ..., m-1$. Since, obviously,

$$\mathcal{C}_{WR} \subset \mathcal{C}_{DR} \subset \mathcal{C}_R, \tag{6.7}$$

Lemma 5.5 and Proposition 5.1 imply that the generalized relaxed income tax problem is equivalent to the deterministic weakly relaxed income tax problem if the solution to the deterministic relaxed income tax problem satisfies (6.6) for all i. The problem of proving Proposition 6.1 is thus reduced to the problem of proving that, if the support of the type distribution is finite, then the solution to the deterministic relaxed income tax problem satisfies the monotonicity requirement for consumption.

The following two lemmas will show that, under WDCRA, this is indeed the case. To understand the basic idea, it is useful to look at Figure 2. In this figure, $(\bar{c}_k, \bar{y}_k)$ and $(\bar{c}_{k+1}, \bar{y}_{k+1}) << (\bar{c}_k, \bar{y}_k)$ are outcomes for types k and $k+1$ that violate consumption monotonicity. $I_1^k, I_2^k, I_1^{k+1}, I_2^{k+1}$ are the indifference curves of types k and $k+1$ through these points. One immediately sees that the pairs $(\bar{c}_k, \bar{y}_k)$ and $(\bar{c}_{k+1}, \bar{y}_{k+1})$ satisfy downward incentive compatibility, but *not* upward incentive compatibility. However, it seems counterintutitive that a utilitarian welfare maximizer should want to assign $(\bar{c}_k, \bar{y}_k)$ to type k and $(\bar{c}_{k+1}, \bar{y}_{k+1})$ to type $k+1$. Given that the distance between the indifference curves I_1^k and I_2^k is rather greater than the distance between the indifference curves I_1^{k+1} and I_2^{k+1}, one suspects that type k would have more to gain than type $k+1$ from receiving $(\bar{c}_{k+1}, \bar{y}_{k+1})$ rather than $(\bar{c}_k, \bar{y}_k)$.

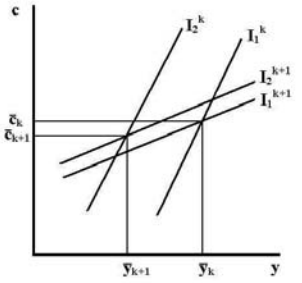

Figure 2: Decreasing Utility Differences for $(\bar{c}_{k+1}, \bar{y}_{k+1}) << (\bar{c}_k, \bar{y}_k)$.

The argument used to confirm this intuition proceeds in two steps. First, Lemma 6.4 shows that, under SSCC, the difference between the utility levels provided by outcome pairs such as $(\bar{c}_k, \bar{y}_k)$ and $(\bar{c}_{k+1}, \bar{y}_{k+1})$ is indeed greater for type k than for type $k+1$, i.e., the distances between the indifference curves I_1^k and I_2^k and the indifference curves I_1^{k+1} and I_2^{k+1} in Figure 2 provide the right impression about differences in cardinal utility levels. Thereafter, Lemma 6.5 uses this finding to show that, in the constellation of Figure 2, either the assigment of $(\bar{c}_k, \bar{y}_k)$ to type k or the assigment of $(\bar{c}_{k+1}, \bar{y}_{k+1})$ to type $k+1$ must be suboptimal.

The property of *monotone utility differences*, which is the subject of Lemma 6.4, can be understood as a strengthening of the Milgrom-Shannon strict single-crossing property. Whereas the single-crossing property is equivalent to the statement that, for any consumption/output pairs (c, y),$(\hat{c}, \hat{y})$ with $(c, y) << (\hat{c}, \hat{y})$ and any n, $u(c, y, n) \geq u(\hat{c}, \hat{y}, n)$ implies that any type $n' < n$ has a strict preference for (c, y) over $(\hat{c}, \hat{y})$, the following result asserts that, in this constellation, the difference $u(c, y, n') - u(\hat{c}, \hat{y}, n')$ in utility levels for $n' < n$ is actually greater than the difference $u(c, y, n) - u(\hat{c}, \hat{y}, n)$ in utility levels for n.

Lemma 6.4 *Let* (c, y),$(\hat{c}, \hat{y})$ *in* $\Re_+^2$ *and* $n \in \Re_+$ *be such that* $(c, y) << (\hat{c}, \hat{y})$ *and* $u(c, y, n) \geq u(\hat{c}, \hat{y}, n)$. *Then*

$$u(c, y, n') - u(\hat{c}, \hat{y}, n') > u(c, y, n) - u(\hat{c}, \hat{y}, n) \tag{6.8}$$

for all $n' < n$.

Proof. Let (c, y) and $(\hat{c}, \hat{y})$ in $\Re^2_+$ be such that $(c, y) << (\hat{c}, \hat{y})$. For any $n'' \in \Re_+$, one can write

$$\begin{aligned} & u(\hat{c}, \hat{y}, n'') - u(c, y, n'') \\ = & \int_0^1 \left[u_c(c'(s), y'(s), n'') \frac{dc'}{ds} + u_y(c'(s), y'(s), n'') \frac{dy'}{ds} \right] ds, \end{aligned} \quad (6.9)$$

where $(c'(\cdot), y'(\cdot))$ is any pair of differentiable functions such that $(c(0), y(0)) = (c, y)$, $(c(1), y(1)) = (\hat{c}, \hat{y})$. Therefore

$$\frac{d}{dn''}[u(\hat{c}, \hat{y}, n'') - u(c, y, n'')] = \int_0^1 \left[u_{cn} \frac{dc'}{ds} + u_{yn} \frac{dy'}{ds} \right] ds; \quad (6.10)$$

in this equation, as in (6.9), the derivatives of u are evaluated at $(c'(s), y'(s), n'')$. By a rearrangement of terms, (6.10) can be rewritten as

$$\begin{aligned} \frac{d}{dn''}[u(\hat{c}, \hat{y}, n'') - u(c, y, n'')] = & \int_0^1 \left(\frac{u_{cn}}{u_c} - \frac{u_{yn}}{u_y} \right) u_c \frac{dc'}{ds} ds \\ & + \int_0^1 \frac{u_{yn}}{u_y} \frac{du(c'(s), y'(s), n'')}{ds} ds \end{aligned} \quad (6.11)$$

If $u(c, y, n'') \geq u(\hat{c}, \hat{y}, n'')$, the functions $c'(\cdot)$ and $y'(\cdot)$ can be chosen so that

$$\frac{dc'}{ds} > 0 \text{ and } \frac{du(c'(s), y'(s), n'')}{ds} \leq 0.$$

In this case, the strict single-crossing condition implies that the first integral on the right-hand side of (6.11) is strictly positive. Moreover, the inequalities $u_y < 0$ and $u_{yn} > 0$, which are obtained from RMQ and DR, imply that the second integral on the right-hand side of (6.11) is nonnegative. Thus, $u(c, y, n'') \geq u(\hat{c}, \hat{y}, n'')$ implies

$$\frac{d}{dn''}[u(\hat{c}, \hat{y}, n'') - u(c, y, n'')] > 0. \quad (6.12)$$

By SSCC, $(c, y) << (\hat{c}, \hat{y})$ and $u(c, y, n) \geq u(\hat{c}, \hat{y}, n)$ imply $u(c, y, n'') > u(\hat{c}, \hat{y}, n'')$ for all $n'' < n$. In this case, (6.12) must hold for all $n'' \leq n$, and one obtains

$$\begin{aligned} & u(c, y, n) - u(\hat{c}, \hat{y}, n) - \left[u(c, y, n') - u(\hat{c}, \hat{y}, n') \right] \\ = & -\int_{n'}^{n} \frac{d}{dn''}[u(\hat{c}, \hat{y}, n'') - u(c, y, n'')] dn'' < 0 \end{aligned}$$

for all $n' < n$. ■

Lemma 6.5 *If $N = \{n_1, n_2, ..., n_m\}$, with $n_1 < ... < n_m$, the solution $\{(\bar{c}_i, \bar{y}_i)\}_{i=1}^m$ to the deterministic relaxed income tax problem satisfies*

$$\bar{c}_i \leq \bar{c}_{i+1} \tag{6.13}$$

for $i = 1, ..., m-1$.

Proof. Suppose that the lemma is false. Then there exists an index i such that $\bar{c}_i > \bar{c}_{i+1}$. Let k be the smallest such index, so that

$$\bar{c}_i \leq \bar{c}_k \tag{6.14}$$

for all $i < k$. By downward incentive compatibility, one has

$$u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_{k+1}) \geq u(\bar{c}_k, \bar{y}_k, n_{k+1}). \tag{6.15}$$

By (6.15) and the inequality $\bar{c}_k > \bar{c}_{k+1}$, one also has $\bar{y}_k > \bar{y}_{k+1}$. By SSCC, it follows that

$$u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_k) > u(\bar{c}_k, \bar{y}_k, n_k); \tag{6.16}$$

indeed, Lemma 6.4, yields

$$u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_k) - u(\bar{c}_k, \bar{y}_k, n_k) > u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_{k+1}) - u(\bar{c}_k, \bar{y}_k, n_{k+1}). \tag{6.17}$$

To prove the lemma, I will show that (6.17) is incompatible with the assumption that $\{(\bar{c}_i, \bar{y}_i)\}_{i=1}^m$ is a solution to the deterministic relaxed income tax problem.

For $i = 1, ..., m$, let $v_i = u(\bar{c}_i, \bar{y}_i, n_i)$. If $\{(\bar{c}_i, \bar{y}_i)\}_{i=1}^m$ is a solution to the deterministic relaxed income tax problem, then, by the argument in the proof of Lemma 5.3, the sequence $\{(v_i, \bar{y}_i)\}_{i=1}^m$ must maximize

$$\sum_{i=1}^{m} v_i F(\{n_i\}) \tag{6.18}$$

under the constraints that

$$\sum_{i=1}^{m} c(v_i, \bar{y}_i, n_i) F(\{n_i\}) \leq \sum_{i=1 \bar{y} i}^{m} \bar{y}_i F(\{n_i\}) \tag{6.19}$$

and

$$\varphi_{n_i}^{n_j}(v_j, \bar{y}_i) \geq v_i \tag{6.20}$$

for $i = 1, ..., m-1$, and $j = i+1, ..., m$. This maximization problem has a concave objective function and, by Lemma 5.3, a convex constraint set.

By Theorem 1, p. 217, of Luenberger (1969), it follows that there exist nonnegative multipliers $\lambda, \mu_i^j, i = 1, ..., m-1, j = i+1, ..., m$ such that $\{(v_i, \bar{y}_i)\}_{i=1}^m$ is also a *global* maximizer of the Lagrangian expression

$$\mathcal{L} = \sum_{i=1}^{m} v_i f_i + \lambda \sum_{i=1}^{m} (y_i - c(v_i, \bar{y}_i, n_i))\ f_i + \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \mu_i^j (\varphi_{n_i}^{n_j}(v_j, \bar{y}_i) - v_i), \quad (6.21)$$

where, for $i = 1, ..., m$, $f_i := F(\{n_i\})$.

In particular, the value of $\mathcal{L}$ is not increased if the pair $(v_{k+1}, \bar{y}_{k+1})$ for type n_{k+1} is replaced by the pair

$$(\hat{v}_{k+1}, \hat{y}_{k+1}) = (u(\bar{c}_k, \bar{y}_k, n_{k+1}), \bar{y}_k). \quad (6.22)$$

One therefore has

$$\begin{aligned} & v_{k+1} f_{k+1} + \lambda(\bar{y}_{k+1} - \bar{c}_{k+1})\ f_{k+1} \\ + \sum_{i<k+1} \mu_i^{k+1} (\varphi_{n_i}^{n_{k+1}}(v_{k+1}, \bar{y}_i) - v_i) + \sum_{j>k+1} \mu_{k+1}^j (\varphi_{n_{k+1}}^{n_j}(v_j, \bar{y}_{k+1}) - v_{k+1}) & \\ \geq \hat{v}_{k+1} f_{k+1} + \lambda(\hat{y}_{k+1} - c(\hat{v}_{k+1}, \hat{y}_{k+1}, n_{k+1})\ f_{k+1} & \quad (6.23) \\ + \sum_{i<k+1} \mu_i^{k+1} (\varphi_{n_i}^{n_{k+1}}(\hat{v}_{k+1}, \bar{y}_i) - v_i) + \sum_{j>k+1} \mu_{k+1}^j (\varphi_{n_{k+1}}^{n_j}(v_j, \hat{y}_{k+1}) - \hat{v}_{k+1}). & \end{aligned}$$

I claim that, for $(\hat{v}_{k+1}, \hat{y}_{k+1}) = (u(\bar{c}_k, \bar{y}_k, n_{k+1}), \bar{y}_k)$, (6.23) implies

$$v_{k+1} f_{k+1} + \lambda(\bar{y}_{k+1} - \bar{c}_{k+1})\ f_{k+1} \geq u(\bar{c}_k, \bar{y}_k, n_{k+1}) f_{k+1} + \lambda(\bar{y}_k - \bar{c}_k)\ f_{k+1}. \quad (6.24)$$

To prove this claim, I first observe that, by Theorem 1, p. 217, of Luenberger (1969), the sequence $\{(v_i, \bar{y}_i)\}_{i=1}^m$ and the multipliers μ_i^j must also satisfy the complementary-slackness condition

$$\mu_i^j (\varphi_{n_i}^{n_j}(v_j, \bar{y}_i) - v_i) = 0 \quad (6.25)$$

for $i = 1, ..., m-1$, and $j = i+1, ..., m$. On the left-hand side of (6.23) therefore, the terms that relate to the incentive constraints (6.20) all vanish. As for the right-hand side, I claim that replacing the pair $(v_{k+1}, \bar{y}_{k+1})$ for type n_{k+1} by the pair $(u(\bar{c}_k, \bar{y}_k, n_{k+1}), \bar{y}_k)$ does not upset any downward incentive compatibility constraint, i.e., that one has $\varphi_{n_i}^{n_{k+1}}(u(\bar{c}_k, \bar{y}_k, n_{k+1}), \bar{y}_i) \geq v_i$ for all $i \leq k$ and $\varphi_{n_{k+1}}^{n_j}(v_j, \bar{y}_k) \geq u(\bar{c}_k, \bar{y}_k, n_{k+1})$ for all $j > k+1$. Because the multipliers μ_i^j are nonnegative, it follows that, for $(\hat{v}_{k+1}, \hat{y}_{k+1}) = (u(\bar{c}_k, \bar{y}_k, n_{k+1}), \bar{y}_k)$, the terms on the right-hand side of (6.23) that relate to

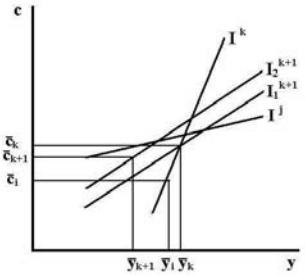

Figure 3: Preservation of Downward Incentive Compatibility when $(\bar{c}_k, \bar{y}_k)$ is Assigned to Type n_{k+1}, or $(\bar{c}_{k+1}, \bar{y}_{k+1})$ to Type n_k

the incentive constraints (6.20) are all nonnegative. Together with (6.25), this yields (6.24).

To see that replacing $(v_{k+1}, \bar{y}_{k+1})$ for type n_{k+1} by $(u(\bar{c}_k, \bar{y}_k, n_{k+1}), \bar{y}_k)$ does not upset downward incentive compatibility, consider Figure 3, which is the same as Figure 2, except that it also exhibits an indifference curve I^j for some type $j > k+1$ and an outcome pair $(\bar{c}_i, \bar{y}_i)$ for some type $i < k$, taking account of the fact that, by (6.14), $\bar{c}_i \leq \bar{c}_k$. Replacing the pair $(v_{k+1}, \bar{y}_{k+1})$ for type n_{k+1} by the pair $(u(\bar{c}_k, \bar{y}_k, n_{k+1}), \bar{y}_k)$ is the same as giving type n_{k+1} the outcome $(\bar{c}_k, \bar{y}_k)$ rather than $(\bar{c}_{k+1}, \bar{y}_{k+1})$. If $\bar{c}_i \leq \bar{c}_k$ and $\bar{y}_i \geq \bar{y}_k$, then, by RMQ, one also has $u(\bar{c}_k, \bar{y}_k, n_{k+1}) \geq u(\bar{c}_i, \bar{y}_i, n_{k+1})$. If $\bar{c}_i \leq \bar{c}_k$ and $\bar{y}_i < \bar{y}_k$, then, by SSCC and the downward incentive compatibility condition $u(\bar{c}_k, \bar{y}_k, n_k) \geq u(\bar{c}_i, \bar{y}_i, n_k)$, one also has $u(\bar{c}_k, \bar{y}_k, n_{k+1}) \geq u(\bar{c}_i, \bar{y}_i, n_{k+1})$. Thus, in either case, if $\bar{y}_i \geq \bar{y}_k$ and if $\bar{y}_i < \bar{y}_k$, (6.14) implies that $u(\bar{c}_k, \bar{y}_k, n_{k+1}) \geq u(\bar{c}_i, \bar{y}_i, n_{k+1})$ and, hence, that that $\varphi_{n_i}^{n_{k+1}}(u(\bar{c}_k, \bar{y}_k, n_{k+1}), \bar{y}_i) \geq v_i$. For $j > k+1$, the downward incentive compatibility of the the original allocation implies $u(\bar{c}_j, \bar{y}_j, n_j) \geq u(\bar{c}_k, \bar{y}_k, n_j)$. By the definition of $\varphi_{n_{k+1}}^{n_j}$, it follows that $\varphi_{n_{k+1}}^{n_j}(v_j, \bar{y}_k) \geq u(\bar{c}_k, \bar{y}_k, n_{k+1})$. Thus, $(v_{k+1}, \bar{y}_{k+1})$ for type n_{k+1} by $(u(\bar{c}_k, \bar{y}_k, n_{k+1}), \bar{y}_k)$ does not upset any of the constraints (6.20). (6.24).

The value of the Lagrangian $\mathcal{L}$ is not increased either if the pair $(v_k, \bar{y}_k)$ for type n_k is replaced by the pair

$$(\hat{v}_k, \hat{y}_k) = (u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_k), \bar{y}_{k+1}). \tag{6.26}$$

By the same argument as before, it follows that

$$v_k f_k + \lambda(\bar{y}_k - \bar{c}_k)\ f_k \geq u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_k) f_k + \lambda(\bar{y}_{k+1} - \bar{c}_{k+1})\ f_k. \tag{6.27}$$

If one now eliminates f_k from (6.27) and f_{k+1} from (6.24), and if one adds the resulting inequalities, one obtains

$$v_k + v_{k+1} \geq u(\bar{c}_k, \bar{y}_k, n_{k+1}) + u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_k). \tag{6.28}$$

By the definitions of v_k, v_{k+1}, (6.28) is equivalent to the inequality

$$u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_{k+1}) - u(\bar{c}_k, \bar{y}_k, n_{k+1}) \geq u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_k) - u(\bar{c}_k, \bar{y}_k, n_k), \tag{6.29}$$

which is incompatible with (6.17). The assumption that the lemma is false has thus led to a contradiction, which proves the lemma. ■

Lemmas 5.5 - 6.5 imply that, if WDCRA is imposed in addition to the other assumptions, then the deterministic relaxed income tax problem is equivalent to the deterministic weakly relaxed income tax problem. The solutions to the deterministic weakly relaxed income tax problem are characterized by Theorem 3.1 in Hellwig (2005). With WDCRA, this characterization must also be valid for the solution to the deterministic relaxed income tax problem. This completes the proof of Proposition 6.1.

7 Continuity Properties of Optimal Allocations

To deal with arbitrary type distributions, I will use Proposition 6.1 and Corollary 6.2 to establish the following result:

Proposition 7.1 *For any type distribution with compact support, a solution to the deterministic relaxed income tax problem satisfies:*

(a) Consumption/output pairs are never distorted upwards from efficiency, i.e.,

$$(\bar{c}(n), \bar{y}(n)) \leq (c^*(v(n), n), y^*(v(n), n)) \tag{7.1}$$

for all $n \in N$.

(b) The functions $\bar{c}(\cdot)$ and $\bar{y}(\cdot)$ are nondecreasing and co-monotonic.

(c) The allocation $(\bar{c}(\cdot), \bar{y}(\cdot))$ is incentive-compatible.

By Lemma 5.6, Proposition 7.1 has the immediate

Corollary 7.2 *For any type distribution with compact support, the generalized relaxed, deterministic relaxed, generalized, and optimal income tax problems are all equivalent.*

The proof of Proposition 7.1 is based on continuity considerations. The key observation is that any type distribution can be suitably approximated by a sequence of type distributions with finite supports. I will show that, along such an approximating sequence, the solutions to the generalized relaxed income tax problems must converge to a limit allocation $(\bar{c}(\cdot), \bar{y}(\cdot))$ and that, as in Berge's (1959) maximum theorem, the limit $(\bar{c}(\cdot), \bar{y}(\cdot))$ is the solution to the generalized relaxed income tax problem in the limit.

In talking about the convergence of the solutions $(\bar{c}^k(\cdot), \bar{y}^k(\cdot))$ to the deterministic relaxed income tax problems for type distributions $F^k, k = 1, 2, ...,$ I must take account of the fact that, strictly speaking, any solution $(\bar{c}^k(\cdot), \bar{y}^k(\cdot))$ is only defined on the support N^k of the type distribution F^k. The solutions to the deterministic relaxed income tax problems for different type distributions can thus have different domains. The domain of any such allocation $(\bar{c}^k(\cdot), \bar{y}^k(\cdot))$ can, however, be enlarged by writing

$$(\bar{c}^k(n), \bar{y}^k(n)) = (\bar{c}^k(\hat{n}^k(n)), \bar{y}^k(\hat{n}^k(n))), \tag{7.2}$$

with the understanding that, for any $n \in \Re_+$, $\hat{n}^k(n) \in N^k$ is closest to n, i.e., one has $|n' - n| \geq \left|\hat{n}^k(n) - n\right|$ for all $n' \in N^k$.

With this extended notion of an allocation, one easily obtains:

Lemma 7.3 *Let $\{F^k\}$ be a sequence of type distributions with finite supports that converges weakly to a distribution F, and suppose that the supports N^k of the distributions F^k are uniformly bounded. For $k = 1, 2, ...,$ let $(\bar{c}^k(\cdot), \bar{y}^k(\cdot))$ be the solution to the deterministic relaxed income tax problem with type distribution F^k. Then there exists an allocation $(\bar{c}(\cdot), \bar{y}(\cdot))$, which is nondecreasing, such that, for some subsequence $\{(\bar{c}^{k'}(\cdot), \bar{y}^{k'}(\cdot))\}$ of the sequence $\{(\bar{c}^k(\cdot), \bar{y}^k(\cdot))\}$, one has*

$$\lim_{k' \to \infty} (\bar{c}^{k'}(n), \bar{y}^{k'}(n)) = (\bar{c}(n), \bar{y}(n)) \tag{7.3}$$

for every mass point of the distribution F and

$$\lim_{k' \to \infty} (\bar{c}^{k'}(n^{k'}), \bar{y}^{k'}(n^{k'})) = (\bar{c}(n), \bar{y}(n)) \tag{7.4}$$

for F-almost every n that is not a mass point of F and every sequence $\{n^{k'}\}$ that converges to n.

Proof. I first show that the outcome pairs $(\bar{c}^k(n), \bar{y}^k(n))$ are uniformly bounded. Statement (c) of Proposition 6.1 implies that, for any k and any $n \in N^k$, one has

$$(0,0) \leq (\bar{c}^k(n), \bar{y}^k(n)) \leq (\bar{c}^k(\bar{n}^k), \bar{y}^k(\bar{n}^k)),$$

where $\bar{n}^k$ is the largest element of N^k. Therefore, it suffices to show that the outcome pairs $(\bar{c}^k(\bar{n}^k), \bar{y}^k(\bar{n}^k))$ are uniformaly bounded. Statement (a) of Proposition 6.1 also implies that, for any k,

$$(\bar{c}^k(\bar{n}^k), \bar{y}^k(\bar{n}^k)) = (c^*(\bar{v}^k, \bar{n}^k), y^*(\bar{v}^k, \bar{n}^k)),$$

where $\bar{v}^k = u(\bar{c}^k(\bar{n}^k), \bar{y}^k(\bar{n}^k), \bar{n}^k)$. By Berge's (1959) maximum theorem, the function pair $(c^*(\cdot,\cdot), y^*(\cdot,\cdot))$ is continuous. Therefore, it suffices to show that the pairs $(\bar{v}^k, \bar{n}^k)$ all lie in a compact set. By assumption, there exist $\underline{n}$ and $\bar{n}$ such that $N^k \subset [\underline{n},\bar{n}]$ for all k. As for $\bar{v}^k$, I compare this utility level to the utility levels $v_{LF}(n) = \max_y u(y,y,n)$, $n \in N^k$, which are obtained under laissez-faire. By RMQ, the function $v_{LF}(\cdot)$ is increasing. Statement (d) of Proposition 6.1 implies that, for any k, one has $\bar{v}^k < v_{LF}(\bar{n}^k)$, hence $\bar{v}^k < v_{LF}(\bar{n})$. Because the welfare generated by the allocation $(\bar{c}^k(\cdot), \bar{y}^k(\cdot))$ is higher than the welfare generated by the laissez-faire allocation, one must also have $u(\bar{c}^k(n), \bar{y}^k(n), n) \geq v_{LF}(n)$ for some $n \in N^k$. By the downward incentive compatibility of the allocation $(\bar{c}^k(\cdot), \bar{y}^k(\cdot))$, it follows that $\bar{v}^k \geq v_{LF}(n)$ for some $n \in N^k$, hence $\bar{v}^k \geq v_{LF}(\underline{n})$. Thus, $\bar{v}^k \in [v_{LF}(\underline{n}), v_{LF}(\bar{n})]$ for all k.

Let N^* be a countable dense subset of the interval $[\underline{n},\bar{n}]$ such that $\bar{n}$, as well as the mass points of the distribution F, if any, belong to N^*. Given that the outcome pairs $(\bar{c}^k(n), \bar{y}^k(n))$ are uniformly bounded, for any $n \in N^*$, there exists a subsequence $\{(\bar{c}^{k'}(n), \bar{y}^{k'}(n))\}$ of the sequence $\{(\bar{c}^k(n), \bar{y}^k(n))\}$ that converges to a limit $(\bar{c}(n), \bar{y}(n))$. Indeed, by a standard diagonalization argument, the sequence $\{(\bar{c}^k(\cdot), \bar{y}^k(\cdot))\}$ has a subsequence $\{(\bar{c}^{k\prime}(\cdot), \bar{y}^{k'}(\cdot))\}$ such that (7.3) holds for all $n \in N^*$. In particular, (7.3) holds for every mass point of F.

Because the functions $(\bar{c}^k(\cdot), \bar{y}^k(\cdot)))$ are all nondecreasing, one has

$$(\bar{c}(n), \bar{y}(n)) \leq (\bar{c}(n'), \bar{y}(n'))$$

for all n and all $n' > n$ in N^*. Given this monotonicity of $(\bar{c}(\cdot), \bar{y}(\cdot))$ on N^*, the limit

$$(\bar{c}(n), \bar{y}(n)) := \lim_{\substack{n' \downarrow n \\ n' \in N^*}} (\bar{c}(n'), \bar{y}(n')) \tag{7.5}$$

is well defined for all $n \in [\underline{n},\bar{n}]$. The allocation $(\bar{c}(\cdot), \bar{y}(\cdot))$, which is thus defined, is obviously nondecreasing. By construction, it also satisfies (7.4) for every n at which it is continuous and for every sequence $\{n^{k'}\}$ that converges to n. Being monotonic, $(\bar{c}(\cdot), \bar{y}(\cdot))$ has at most countably many discontinuity points. Under the measure F, therefore, the set of discontinuity points of $(\bar{c}(\cdot), \bar{y}(\cdot))$ that are not mass points of F has measure zero. ■

Lemma 7.4 *The limit allocation in Lemma 7.3 satisfies statements (a) - (c) of Proposition 7.1.*

Proof. Suppose that the allocation $(\bar{c}(\cdot), \bar{y}(\cdot))$ is not downward incentive-compatible. Then, there exist n and $\hat{n} < n$ in the support of F such that

$$u(\bar{c}(n), \bar{y}(n), n) < u(\bar{c}(\hat{n}), \bar{y}(\hat{n}), n).$$

By the definition of $(\bar{c}(\cdot), \bar{y}(\cdot))$, it follows that there exist n' and $\hat{n}' < n'$, which belong to the set N^* in the preceding proof, such that

$$u(\bar{c}(n'), \bar{y}(n'), n') < u(\bar{c}(\hat{n}'), \bar{y}(\hat{n}'), n').$$

Again, by the definition of $(\bar{c}(\cdot), \bar{y}(\cdot))$, it follows that, for any sufficiently large k', one has

$$u(\bar{c}^{k'}(n'), \bar{y}^{k'}(n'), n') < u(\bar{c}^{k'}(\hat{n}'), \bar{y}^{k'}(\hat{n}'), n'),$$

contrary to the assumption that the allocation $(\bar{c}^{k'}(\cdot), \bar{y}^{k'}(\cdot))$ is downward incentive-compatible. The assumption that the allocation $(\bar{c}(\cdot), \bar{y}(\cdot))$ is not downward incentive-compatible thus leads to a contradiction and must be false.

Exactly the same argument shows that, if the allocation $(\bar{c}(\cdot), \bar{y}(\cdot))$ is not upward incentive-compatible, then there exist n' and $\hat{n}' < n'$ in the set N^* such that, for any sufficiently large k', one has

$$u(\bar{c}^{k'}(n'), \bar{y}^{k'}(n'), n') > u(\bar{c}^{k'}(\hat{n}'), \bar{y}^{k'}(\hat{n}'), n'),$$

contrary to the finding in Proposition 6.1 that the allocation $(\bar{c}^{k'}(\cdot), \bar{y}^{k'}(\cdot))$ is upward, as well as downward, incentive-compatible. The assumption that the allocation $(\bar{c}(\cdot), \bar{y}(\cdot))$ is not upward incentive-compatible thus also leads to a contradiction and must be false. This proves that $(\bar{c}(\cdot), \bar{y}(\cdot))$ satisfies statement (c) of Proposition 7.1.

Statement (b) of Proposition 7.1 is also satisfied because the allocation $(\bar{c}(\cdot), \bar{y}(\cdot))$ is a limit of nondecreasing allocations, and, by standard arguments, incentive compatibility implies the co-monotonicity of $\bar{c}(\cdot)$ and $\bar{y}(\cdot)$. As for statement (a), this follows directly from Proposition 6.1 and the specification of $(\bar{c}(\cdot), \bar{y}(\cdot))$. ■

The next two lemmas rely on the concept of setwise convergence of distributions, which is stronger than weak convergence. As defined in Royden (1968), p.269, a sequence of distributions $\{F^k\}$ converges *setwise* to a distribution F if $F^k(B)$ converges to $F(B)$ for every measurable set B. Some technical results about setwise convergence are given in Appendix A.

Lemma 7.5 *If the sequence of type distributions $\{F^k\}$in Lemma 7.3 converges setwise to the distribution F, then the limit allocation $(\bar{c}(\cdot), \bar{y}(\cdot))$ in Lemma 7.3 is feasible for the type distribution F. Moreover,*

$$\int u(\bar{c}(n), \bar{y}(n), n) dF(n) = \lim_{k \to \infty} \int u(\bar{c}^k(n), \bar{y}^k(n), n) dF^k(n). \tag{7.6}$$

Proof. By Lemmas 7.3 and A.3, the assumptions of the lemma imply that

$$\lim_{k' \to \infty} \int f(\bar{c}^{k'}(n), \bar{y}^{k'}(n), n) dF^k(n) = \int f(\bar{c}(n), \bar{y}(n), n) dF(n) \tag{7.7}$$

for every continuous function $f : \Re^3_+ \to \Re$. (7.6) follows immediately. (7.7) also implies

$$\int [\bar{y}(n) - \bar{c}(n)] dF(n) = \lim_{k' \to \infty} \int [\bar{y}^{k'}(n) - \bar{c}^{k'}(n)] dF^k(n).$$

Given that $(\bar{c}^{k'}(\cdot), \bar{y}^{k'}(\cdot))$ is feasible when the type distribution is $F^{k'}$, it follows that

$$\int [\bar{y}(n) - \bar{c}(n)] dF(n) \geq 0.$$

This proves that the allocation $(\bar{c}(\cdot), \bar{y}(\cdot))$ is feasible. ■

Lemma 7.6 *If the sequence of type distributions $\{F^k\}$ in Lemma 7.3 converges setwise to the distribution F, then the limit allocation $(\bar{c}(\cdot), \bar{y}(\cdot))$ is a solution to the deterministic relaxed income tax problem for the type distribution F.*

Proof. Suppose that the lemma is false. Then there exists $\varepsilon > 0$, and there exists a deterministic allocation $(\hat{c}(\cdot), \hat{y}(\cdot))$ that is downward incentive-compatible, as well as feasible for F, such that

$$\int u(\hat{c}(n), \hat{y}(n), n) dF(n) \geq \int u(\bar{c}(n), \bar{y}(n), n) dF(n) + \varepsilon.$$

Then Lemma A.4 and (7.6) imply that, for any sufficiently large k', there exists an allocation $(\hat{c}^{k'}(\cdot), \hat{y}^{k'}(\cdot))$ that is downward incentive compatible, as well as feasible for $F^{k'}$, such that

$$\int u(\hat{c}^{k'}(n), \hat{y}^{k'}(n), n) dF^{k'}(n) \geq \int u(\bar{c}^{k'}(n), \bar{y}^{k'}(n), n) dF^{k'}(n) + \frac{\varepsilon}{3},$$

contrary to the assumption that $(\bar{c}^{k'}(\cdot), \bar{y}^{k'}(\cdot))$ is a solution to the deterministic relaxed income tax problem when the type distribution is $F^{k'}$. The assumption that $(\bar{c}(\cdot), \bar{y}(\cdot))$ is not a solution to the deterministic relaxed income tax problem for the type distribution F has thus led to a contradiction and must be false. ■

Proposition 7.1 now follows from Lemmas 7.3 - 7.6, Proposition 5.1, and the finding, established in Lemma A.2 in the appendix, that, in the topology of setwise convergence, any distribution on $\Re$ can be approximated by a sequence of distributions with finite supports.

Theorem 4.1 and Corollary 4.2 follow from Propositions 5.1 and Corollary 7.2. Theorem 4.3 follows from the observation that the monotonicity and boundedness properties of optimal allocation that are established in Proposition 7.1 are sufficient to establish the conclusions of Lemma 7.3, and therefore also of Lemmas 7.4 - 7.6, even when the supports of the distributions F^k are *not* finite. Moreover, Proposition 5.1 implies that, up to modificiations on null sets, all convergent subsequences have the same limit.

8 Concluding Remarks

I conclude with two remarks on additional questions. First, the finding that, under WDCRA, the solution to the optimal income tax problem is essentially unique, and depends continuously on the type distribution, leads on to the question whether, under WDCRA, this solution itself is continuous, i.e., whether the allocation $(\bar{c}(\cdot), \bar{y}(\cdot))$ is (equivalent to) a pair of continuous functions. Such a continuity property would imply that the optimal marginal income tax schedule is continuous. This is a much stronger property than

the continuity of the optimal income tax schedule itself that was established by Weibull (1989).

The answer to this question is mixed: If the type distribution F has a continuous density, then, under WDCRA, the solution to the optimal income tax problem is indeed given by a pair of continuous functions. This follows from the finding of Mirrlees (1986, p. 1231 f.) that, in the optimal-control formulation of the optimal income tax problem with a continuous type distribution, condition WDCRA ensures that, in the application of Pontryagin's maximum principle, for each $n \in N$, the maximand is quasi-concave in the control variables. For each $n \in N$, therefore, the maximand has a unique local and global maximum, and the maximizing controls depend continuously on the data of the problem at n, in particular, on the value $f(n)$ of the density of F. If the dependence of $f(n)$ on n is continuous, this implies that the optimal allocation is continuous.

However, if the density f is *not* continuous, the very same argument indicates that the optimal allocation will be discontinuous at any discontinuity point of F. In this case, the optimal marginal income tax will also be discontinuous at the corresponding income level. The reason is that a jump in the density $f(n)$ affects the equity-efficiency tradeoff. For instance, if one has $f(\hat{n}+) > f(\hat{n}-)$ for some $\hat{n} \in N$, then downward distortions from efficiency at values of n just above $\hat{n}$ concern relatively more people and therefore are costlier than downward distortions at values of n just below $\hat{n}$; yet, because, for n just above $\hat{n}$ and for n just below $\hat{n}$, the fraction of people in the population whose productivity parameters are lower is almost the same, the distributive implications of these distortions are also almost the same. The discontinuous change in the equity-efficiency tradeoff should lead to an optimal allocation which satisfies $(\bar{c}(\hat{n}+), \bar{y}(\hat{n}+)) >> (\bar{c}(\hat{n}-), \bar{y}(\hat{n}-))$. If the distribution F is not continuous, and $\hat{n}$ is a mass point of F, the same considerations suggest that one should actually have $(\bar{c}(\hat{n}+), \bar{y}(\hat{n}+)) = (\bar{c}(\hat{n}), \bar{y}(\hat{n})) >> (\bar{c}(\hat{n}-), \bar{y}(\hat{n}-))$.

A second remark concerns the role of condition WDCRA in establishing the equivalence of the deterministic relaxed income tax problem and the optimal income tax problem. I do not know whether WDCRA is really necessary for this. However, without WDCRA, I have been unable to establish that a solution to the deterministic relaxed income tax problem satisfies consumption monotonicity. This is why Hellwig (2005) works with the weakly generalized relaxed income tax problem, imposing consumption monotonicity as a constraint rather than deriving it as an implication of optimality.

However, the role of WDCRA in establishing consumption monotonicity is strangely technical. In the proof of Lemma 6.5, the replacement of the outcome pair $(\bar{c}_{k+1}, \bar{y}_{k+1})$ by the pair $(\bar{c}_k, \bar{y}_k)$ for type n_{k+1} is a large, i.e., more than marginal, change, as is the replacement of $(\bar{c}_k, \bar{y}_k)$ by $(\bar{c}_{k+1}, \bar{y}_{k+1})$ for type n_k. For an evaluation of such large changes in terms of the Lagrangian $\mathcal{L}$ to be relevant for the underlying constrained optimization, the objective function and constraints of the underlying problem must exhibit appropriate curvature properties. In the present setting, as shown in Lemma 5.3, condition WDCRA guarantees that these curvature properties are satisfied. However, this argument does not convey any intuition as to why the imposition of WDCRA should be germane to establishing the consumption monotonicity of any solution to the deterministic relaxed income tax problem or the equivalence of the deterministic relaxed and the optimal income tax problems.

An alternative, non-Lagrangian proof of consumption monotonicity is given in Appendix B. The argument there relies on local, rather non-local, comparisons. However, the local comparisons that are being considered move out of the space of deterministic allocations and involve lotteries. Given a deterministic, feasible, and downward incentive-compatible allocation that violates consumption monotonicity, the argument considers moving to a new allocation, under which type n_{k+1} has a small probability of receiving $(\bar{c}_k, \bar{y}_k)$ rather than $(\bar{c}_{k+1}, \bar{y}_{k+1})$, and type n_k has a small probability of receiving $(\bar{c}_{k+1}, \bar{y}_{k+1})$ rather than $(\bar{c}_k, \bar{y}_k)$. The probabilities are calibrated so that feasibility is preserved. The property of monotone utility differences, which was established in Lemma 6.4, implies that such a change raises welfare without upsetting any of the constaints of the generalized relaxed income tax problem. Condition WDCRA then enters the argument through Lemmas 5.2 and 5.4, showing that the allocation with lotteries in turn is dominated by another deterministic allocation.

This argument indicates that, if, in the absence of WDCRA, a solution to the deterministic relaxed income tax problem violates consumption monotonicity, then randomization must be desirable. The same forces that induce the breakdown of the equivalence of the deterministic relaxed and optimal income tax problems also make for the desirability of a randomized, rather than a deterministic, allocation.

A Appendix: Some Results on Setwise Convergence

In this appendix, I prove the technical results about setwise convergence of distributions that are used in the proof of Proposition 7.1.

Lemma A.1 *The sequence $\{F^k\}$ of distributions on $\Re$ converges setwise to the distribution F if and only if $\{F^k\}$ converges weakly to F and $\{F^k(\{n\})\}$ converges to $F(\{n\})$ for every atom n of F.*

Proof. The "only if" part of the lemma follows immediately from the definitions and from Theorem 2.1, p. 11, in Billingsley (1968). To prove the "if" part of the lemma, suppose that $\{F^k\}$ converges weakly to F and $\{F^k(\{n\})\}$ converges to $F(\{n\})$ for every atom n of F. To establish setwise convergence, it suffices to show that $F^k([a,b])$ converges to $F([a,b])$ for all a and $b \geq a$. By Theorem 2.1, p. 11, in Billingsley (1968), the weak convergence of the sequence $\{F^k\}$ to F implies

$$\limsup F^k([a,b]) \leq F([a,b]) \tag{A.1}$$

and

$$\liminf F^k((a,b)) \geq F((a,b)). \tag{A.2}$$

If $b = a$, one has $[a,a] = \{a\}$. If $F(\{a\}) > 0$, one has

$$\lim F^k([a,a]) = F([a,a]) \tag{A.3}$$

by assumption. If $F(\{a\}) = 0$, (A.1) implies $\limsup F^k([a,a]) \leq 0$; given that $F^k([a,a]) \geq 0$ for all k, (A.3) holds in this case as well.

If $b > a$, (A.1) is equivalent to the inequality

$$\limsup[F^k(\{a\}) + F^k((a,b)) + F^k(\{b\})] \leq F([a,b]). \tag{A.4}$$

By (A.3), one has $\lim F^k(\{a\}) = F(\{a\})$ and $\lim F^k(\{b\}) = F(\{b\})$. (A.4) then yields

$$\limsup F^k((a,b)) \leq F((a,b)). \tag{A.5}$$

Upon combining (A.5) and (A.2), one obtains

$$\lim F^k((a,b)) = F((a,b)). \tag{A.6}$$

Upon combining (A.6) and (A.3), one also obtains

$$\lim F^k([a,b]) = F([a,b]). \tag{A.7}$$

■

Lemma A.2 *In the topology of setwise convergence, any distribution on $\Re$ can be approximated by a sequence of distributions with finite supports, i.e., the distributions with finite supports are dense in the set of all distributions.*

Proof. It is well known that the assertion of the lemma is true if the space of distributions on $\Re$ is endowed with the topology of weak convergence; see, e.g., Theorem 4, p.237, in Billingsley (1968). Let F be any distribution on $\Re$, and let $N^a = \{n_1^a, n_2^a, ...\}$ be the set of atoms of F. If $N^a = \varnothing$, the result just cited implies that there is a sequence $\{F^k\}$ of distributions with finite supports that converge weakly to F. Because F has no atoms, Lemma A.1 implies that the sequence $\{F^k\}$ actually converges setwise to F. If $N^a \neq \varnothing$, one can write

$$F = \delta F_c + (1 - \delta)F_a, \tag{A.8}$$

where F_c is an atomless distribution, F_a is a purely atomic distribution, and $\delta \in [0, 1]$. By the result in Billingsley (1968), there exists a sequence $\{G^k\}$ of distributions with finite supports that converges weakly to F_c. For any k, one can also define F_a^k so that, for $i = 1, ..., k$,

$$F_a^k(\{n_i^a\}) = \frac{F_a(\{n_i^a\})}{F_a(\{n_1^a, ..., n_k^a\})}.$$

If one sets

$$F^k = \delta G^k + (1 - \delta)F_a^k,$$

then, by Lemma A.1, the sequence distributions F^k converge setwise to F. Moreover, for any k, the support of F^k is finite. ■

Lemma A.3 *Let $\{F^k\}$ be a sequence of distributions on $\Re$ that converges setwise to a distribution F, and suppose that the supports N^k of the distributions F^k are uniformly bounded. Let $(\bar{c}^k(\cdot), \bar{y}^k(\cdot)), k = 1, 2, ...,$ and $(\bar{c}(\cdot), \bar{y}(\cdot))$ be such that (i) the pairs $(\bar{c}^k(n), \bar{y}^k(n))$, with $n \in N^k$, are uniformly bounded, (ii) $(\bar{c}^k(n), \bar{y}^k(n))$ converges to $(\bar{c}(n), \bar{y}(n))$ for every mass point n of F, and (iii) $(\bar{c}^k(n^k), \bar{y}^k(n^k))$ converges to $(\bar{c}(n), \bar{y}(n))$ for F-almost every n that is not a mass point of F and every sequence $\{n^k\}$ that converges to n. Then*

$$\lim_{k\to\infty} \int f(\bar{c}^k(n), \bar{y}^k(n), n)dF^k(n) = \int f(\bar{c}(n), \bar{y}(n), n)dF(n) \tag{A.9}$$

for every continuous function $f : \Re_+^3 \to \Re$.

Proof. Suppose first that the distribution F has atoms as well as a nonatomic part. Then F can be decomposed in the form (A.8), where, again, F_c is an atomless distribution, F_a is a purely atomic distribution, and $\delta \in (0,1)$. Let N_c and $N_a = \{n_1^a, n_2^a, ...\}$ be the supports of the distributions F_c and F_a. For any k, one can also write

$$F^k = \delta_k G_1^k + (1-\delta_k) G_2^k, \tag{A.10}$$

where $G_1^k(N_a) = 0$ and $G_2^k(N_a) = 1$.[8] Given (A.8) and (A.10), (A.9) is equivalent to the requirement that

$$\begin{aligned} &\lim_{k\to\infty} \delta_k \int f(\bar{c}^k(n), \bar{y}^k(n), n) dG_1^k(n) + \lim_{k\to\infty} (1-\delta_k) \sum_{n\in N_a} f(\bar{c}^k(n), \bar{y}^k(n), n) G_2^k(\{n\}) \\ &= \delta \int f(\bar{c}(n), \bar{y}(n), n) dF_c(n) + (1-\delta) \sum_{n\in N_a} f(\bar{c}(n), \bar{y}(n), n) F_a(\{n\}). \end{aligned} \tag{A.11}$$

The setwise convergence of the sequence $\{F^k\}$ to F implies that $F^k(N_a) = (1-\delta_k)G_2^k(N_a)$ converges to $F(N_a) = (1-\delta)F_a(N_a)$ as k goes out of bounds. Since $F_a(N_a) = G_2^k(N_a) = 1$, it follows that

$$\lim_{k\to\infty} \delta_k = \delta. \tag{A.12}$$

By (A.12) and Lemma A.1, the setwise convergence of F^k to F also implies that G_1^k converges weakly to F_c and that $G_2^k(\{n\})$ converges to $F_a(\{n\})$ for every $n \in N_a$.

Let $id(\cdot)$ be the identity function on $\Re$. By Theorem 5.5, p.34, of Billingsley (1968), the convergence property (ii) of the functions $(\bar{c}^k(\cdot), \bar{y}^k(\cdot))$ and the weak convergence of G_1^k to F_c imply that the distributions $G_1^k \circ (\bar{c}^k(\cdot), \bar{y}^k(\cdot), id(\cdot))^{-1}$ converge weakly to the distribution $F_c \circ (\bar{c}(\cdot), \bar{y}(\cdot), id(\cdot))^{-1}$. Therefore, one has

$$\lim_k \int f(\bar{c}^k(n), \bar{y}^k(n), n) dG_1^k(n) = \int f(\bar{c}(n), \bar{y}(n), n) dF_c(n) \tag{A.13}$$

for every function $f : \Re_+^3 \to \Re$ that is continuous and bounded on the union of the ranges of the functions $(\bar{c}^k(\cdot), \bar{y}^k(\cdot), id(\cdot))$, $k = 1, 2, ...$, and $(\bar{c}(\cdot), \bar{y}(\cdot), id(\cdot))$.

[8] Because the distribution F^k can have different atoms from F, the decomposition of F^k that is provided by (A.10) does *not* generally coincide with the decomposition of F^k into a continuous part and a purely atomic part.

Similarly, for every function $f : \Re^3_+ \to \Re$ that is continuous and bounded on the union of the ranges of the functions $(\bar{c}^k(\cdot), \bar{y}^k(\cdot), id(\cdot))$, $k = 1, 2, ...$, and $(\bar{c}(\cdot), \bar{y}(\cdot), id(\cdot))$, the convergence property (iii) of the functions $(\bar{c}^k(\cdot), \bar{y}^k(\cdot))$ and the convergence of $G_2^k(\{n\})$ to $F_a(\{n\})$ for every $n \in N_a$ imply that

$$\lim_{k\to\infty} f(\bar{c}^k(n), \bar{y}^k(n), n)G_2^k(\{n\}) = f(\bar{c}(n), \bar{y}(n), n)F_a(\{n\}) \tag{A.14}$$

for every $n \in N_a$, hence

$$\lim_{k\to\infty} \sum_{n\in N_a} f(\bar{c}^k(n), \bar{y}^k(n), n)G_2^k(\{n\}) = \sum_{n\in N_a} f(\bar{c}(n), \bar{y}(n), n)F_a(\{n\}). \tag{A.15}$$

Since, by assumption, the functions $(\bar{c}^k(\cdot), \bar{y}^k(\cdot), id(\cdot))$, $k = 1, 2, ...$, and $(\bar{c}(\cdot), \bar{y}(\cdot), id(\cdot))$ all take values in a compact subset of $\Re^3_+$, (A.13) and (A.15) in fact hold for every continuous $f : \Re^3_+ \to \Re$.

(A.11) now follows from (A.12), (A.13), and (A.15).

If the distribution F has no atoms, one has $\delta = 1$ in (A.8) and $\delta_k = 1$ in (A.10). (A.9) then is equivalent to (A.13), which follows by the same argument as before. If the distribution F is purely atomic, one has $\delta = 0$ in (A.8) and $\delta_k = 0$ in (A.10). (A.9) then is equivalent to (A.15), which also follows by the same argument as before. ■

Lemma A.4 *Let F be any type distribution with compact support. If the allocation $(\hat{c}(\cdot), \hat{y}(\cdot))$ is downward incentive-compatible and feasible when the type distribution is F, then, for any sequence of type distributions $\{F^k\}$ that converges setwise to F, there exists an associated sequence $\{(\hat{c}^k(\cdot), \hat{y}^k(\cdot))\}$ of allocations such that, for any sufficiently large k, $(\hat{c}^k(\cdot), \hat{y}^k(\cdot))$ is downward incentive-compatible and feasible when the type distribution is F^k, and, moreover,*

$$\lim_{k\to\infty} \int u(\hat{c}^k(n), \hat{y}^k(n), n)dF^k(n) = \int u(\hat{c}(n), \hat{y}(n), n)dF(n). \tag{A.16}$$

Proof. Consider the allocation under which everybody is assigned the constant pair $(0, \bar{y})$, where $\bar{y}$ is some strictly positive number. Clearly, this allocation is downward incentive-compatible, as well as feasible for any type distribution. For any n, let $\bar{v}(n) := u(0, \bar{y}, n)$ be the utility that type n gets from this allocation. Also, let $\hat{v}(n) := u(\hat{c}(n), \hat{y}(n), n)$ be the utility that type n gets from the allocation $(\hat{c}(\cdot), \hat{y}(\cdot))$. For any $\delta \in [0, 1]$ and any n, define

$$v_\delta(n) := \delta\bar{v}(n) + (1 - \delta)\hat{v}(n)$$

and

$$y_\delta(n) := \delta\bar{y} + (1-\delta)\hat{y}(n),$$

and consider the allocation

$$(c_\delta(\cdot), y_\delta(\cdot)) = (c(v_\delta(\cdot), y_\delta(\cdot), \cdot), y_\delta(\cdot)),$$

where, for any n, $c(v_\delta(n), y_\delta(n), n)$ is the level of consumption that is needed to provide type n with the utility $v_\delta(n)$ when the output requirement is $y_\delta(n)$.

For any $\delta \in (0,1]$, Lemma 5.3 implies that the allocation $(c_\delta(\cdot), y_\delta(\cdot))$ is downward incentive-compatible, as well as feasible, when the type distribution is F; indeed, one has

$$\int [y_\delta(n) - c_\delta(n)]dF(n) \geq \delta\bar{y} + (1-\delta)\int [\hat{y}(n) - \hat{c}(n)]dF(n) > 0. \quad \text{(A.17)}$$

By Proposition 18, p. 270, in Royden (1988), one also has

$$\lim_{k\to\infty} \int [y_\delta(n) - c_\delta(n)]dF^k(n) = \int [y_\delta(n) - c_\delta(n)]dF(n)$$

for any $\delta \in [0,1]$. For any $\delta \in (0,1]$ and any sufficiently large k, one therefore has

$$\int [y_\delta(n) - c_\delta(n)]dF^k(n) \geq 0. \quad \text{(A.18)}$$

For any k, let δ_k be the smallest δ for which (A.18) holds, and define

$$(\hat{c}^k(\cdot), \hat{y}^k(\cdot)) = (c_{\delta_k}(\cdot), y_{\delta_k}(\cdot)). \quad \text{(A.19)}$$

As k goes out of bounds, δ_k converges to zero, and the allocations $(\hat{c}^k(\cdot), \hat{y}^k(\cdot))$ converge pointwise to $(\hat{c}(\cdot), \hat{y}(\cdot))$. Therefore, (A.16) follows by another application of Proposition 18, p. 270, in Royden (1988). ■

B Appendix: An Alternative Proof of Consumption Monotonicity

As mentioned in the Concluding Remarks, this second appendix gives an alternative proof of the result that, with finitely many types, any solution to the deterministic relaxed income tax problem exhibits consumption monotonicity. As a preliminary result, the following lemma shows that if consumption monotonicity is violated, then the corresponding downward incentive constraint cannot be binding.

Lemma B.1 *If $N = \{n_1, n_2, ..., n_m\}$, with $n_1 < ... < n_m$, then, for $i = 1, ..., m-1$, the solution $\{(\bar{c}_i, \bar{y}_i)\}_{i=1}^m$ to the deterministic relaxed income tax problem satisfies either*

$$\bar{c}_i \leq \bar{c}_{i+1} \tag{B.1}$$

or

$$u(\bar{c}_{i+1}, \bar{y}_{i+1}, n_{i+1}) > u(\bar{c}_i, \bar{y}_i, n_{i+1}). \tag{B.2}$$

Proof. Suppose that the lemma is false. Then there exists an index k such that

$$\bar{c}_{k+1} < \bar{c}_k \tag{B.3}$$

and

$$u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_{k+1}) \not> u(\bar{c}_k, \bar{y}_k, n_{k+1}). \tag{B.4}$$

By downward incentive compatibility, (B.4) implies

$$u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_{k+1}) = u(\bar{c}_k, \bar{y}_k, n_{k+1}). \tag{B.5}$$

I claim that (B.3) and (B.5) imply that the outcome pair $(\bar{c}_{k+1}, \bar{y}_{k+1})$ is efficient for type n_{k+1}. To establish this claim, I note that, by SSCC, (B.3) and (B.5) imply

$$u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_j) < u(\bar{c}_k, \bar{y}_k, n_j) \tag{B.6}$$

for all $j > k+1$. Because downward incentive compatibility implies

$$u(\bar{c}_j, \bar{y}_j, n_j) \geq u(\bar{c}_k, \bar{y}_k, n_j), \tag{B.7}$$

it follows that

$$u(\bar{c}_j, \bar{y}_j, n_j) > u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_j) \tag{B.8}$$

for all $j > k+1$. Thus, none of the downward incentive constraints that relate types n_j, $j > k+1$, to type n_{k+1} are binding. If $(\bar{c}_{k+1}, \bar{y}_{k+1})$ were not efficient for type n_{k+1}, it would be possible to improve the allocation by moving $(\bar{c}_{k+1}, \bar{y}_{k+1})$ in a direction where type n_{k+1} obtains greater utility without any change in net resource use $\bar{y}_{k+1} - \bar{c}_{k+1}$.

By SSCC, (B.3) and (B.5) also imply

$$u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_k) > u(\bar{c}_k, \bar{y}_k, n_k). \tag{B.9}$$

Because $(\bar{c}_{k+1}, \bar{y}_{k+1})$ is efficient for type n_{k+1}, condition RMQ implies that

$$y - c < \bar{y}_{k+1} - \bar{c}_{k+1}$$

for any pair $(c, y) >> (\bar{c}_{k+1}, \bar{y}_{k+1})$ that lies on the indifference curve of type n_{k+1} through $(\bar{c}_{k+1}, \bar{y}_{k+1})$. In particular, therefore,

$$\bar{y}_k - \bar{c}_k < \bar{y}_{k+1} - \bar{c}_{k+1}. \tag{B.10}$$

However, (B.9) and (B.10) together imply that, if one replaces $(\bar{c}_k, \bar{y}_k)$ for type n_k by $(\bar{c}_{k+1}, \bar{y}_{k+1})$, one can raise the utility of this type, while saving on net resource use. Given that no downward incentive compatibility constraint is upset by this change, it follows that the allocation $\{(\bar{c}_i, \bar{y}_i)\}_{i=1}^m$ cannot be optimal. The assumption that the lemma is false has thus led to a contradiction. ■

On the basis of Lemma B.1, the following lemma establishes consumption monotonicity by showing that any allocation that violates consumption monotonicity is dominated by a suitable randomized allocation. Because Lemmas 5.2 and 5.4 imply that the randomized allocation in turn is dominated by another deterministic allocation, this implies that the allocation in question cannot be a solution to the deterministic relaxed income tax problem.

Lemma B.2 *If $N = \{n_1, n_2, ..., n_m\}$, with $n_1 < ... < n_m$, the solution $\{(\bar{c}_i, \bar{y}_i)\}_{i=1}^m$ to the deterministic relaxed income tax problem satisfies $\bar{c}_i \leq \bar{c}_{i+1}$ for $i = 1, ..., m-1$.*

Proof. Suppose that the lemma is false. Then, there exists an index i such that $\bar{c}_{i+1} < \bar{c}_i$. Let k be the smallest such index. Then

$$\bar{c}_i \leq \bar{c}_k \tag{B.11}$$

for $i < k$. By Lemma B.1 and downward incentive compatibility, one has

$$u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_{k+1}) > u(\bar{c}_k, \bar{y}_k, n_{k+1}). \tag{B.12}$$

I will show that the allocation $\{(\bar{c}_i, \bar{y}_i)\}_{i=1}^m$ is dominated by a randomized allocation, which is specified as follows: Any type n_i other than n_k and n_{k+1} gets $(\bar{c}_i, \bar{y}_i)$ with probability one. Type n_k gets $(\bar{c}_k, \bar{y}_k)$ with probability $1-\eta$ and $(\bar{c}_{k+1}, \bar{y}_{k+1})$ with probability η, where $\eta = \frac{\varepsilon}{f_k}$ for some small $\varepsilon > 0$. Type n_{k+1} gets $(\bar{c}_{k+1}, \bar{y}_{k+1})$ with probability $1 - \delta$ and $(\bar{c}_k, \bar{y}_k)$ with probability δ, where $\delta = \frac{\varepsilon}{f_{k+1}}$.

By construction, this randomized allocation satisfies

$$\begin{aligned}
&\int_N \int_0^1 (y(n,\omega) - c(n,\omega)) d\nu(\omega) dF(n) \\
= \quad &\sum_{i=1}^{m} (\bar{y}_i - \bar{c}_i) f_i + \frac{\varepsilon}{f_k}[\bar{y}_{k+1} - \bar{c}_{k+1} - (\bar{y}_k - \bar{c}_k)] f_k \\
&+ \frac{\varepsilon}{f_{k+1}}[\bar{y}_k - \bar{c}_k - (\bar{y}_{k+1} - \bar{c}_{k+1})] f_{k+1} = \sum_{i=1}^{m} (\bar{y}_i - \bar{c}_i) f_i,
\end{aligned}$$

so it inherits feasibility from $\{(\bar{c}_i, \bar{y}_i)\}_{i=1}^m$.

Similarly, all downward incentive compatibility conditions that do not involve n_k or n_{k+1} are inherited from$\{(\bar{c}_i, \bar{y}_i)\}_{i=1}^m$. For types n_j with $j > k+1$, the downward incentive compatibility of the allocation $\{(\bar{c}_i, \bar{y}_i)\}_{i=1}^m$ implies

$$u(\bar{c}_j, \bar{y}_j, n_j) \geq \max[u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_j), u(\bar{c}_k, \bar{y}_k, n_j)],$$

hence

$$u(\bar{c}_j, \bar{y}_j, n_j) \geq (1-\eta) u(\bar{c}_k, \bar{y}_k, n_j) + \eta u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_j) \tag{B.13}$$

and

$$u(\bar{c}_j, \bar{y}_j, n_j) \geq (1-\delta) u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_j) + \delta u(\bar{c}_k, \bar{y}_k, n_j). \tag{B.14}$$

Thus, $(\bar{c}_j, \bar{y}_j)$ is downward incentive-compatible for n_j relative to the lotteries assigned to n_k and n_{k+1}.

Turning to type n_k, I note that, by SSCC, (B.12) and $(c_{k+1}, y_{k+1}) << (c_k, y_k)$ yield

$$u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_k) > u(\bar{c}_k, \bar{y}_k, n_k), \tag{B.15}$$

i.e., type n_k prefers the bundle $(\bar{c}_{k+1}, \bar{y}_{k+1})$ to his own bundle $(\bar{c}_k, \bar{y}_k)$. Therefore, he also prefers the lottery over $(\bar{c}_{k+1}, \bar{y}_{k+1})$ and $(\bar{c}_k, \bar{y}_k)$ to the bundle $(\bar{c}_k, \bar{y}_k)$. Given that the downward incentive-compatible allocation $\{(\bar{c}_i, \bar{y}_i)\}_{i=1}^m$ satisfies

$$u(\bar{c}_k, \bar{y}_k, n_k) \geq u(\bar{c}_i, \bar{y}_i, n_k) \tag{B.16}$$

for all $i < k$, it follows that he also prefers the lottery over $(\bar{c}_{k+1}, \bar{y}_{k+1})$, and $(\bar{c}_k, \bar{y}_k)$ to any of the bundles $(\bar{c}_i, \bar{y}_i)$ for $i < k$.

As for type n_{k+1}, the strict inequality in (B.12) implies that, for $\delta = \frac{\varepsilon}{f_{k+1}}$ and $\eta = \frac{\varepsilon}{f_k}$ close to zero, he prefers the lottery that is assigned to him to

the lottery that is assigned to type n_k and to the bundle $(\bar{c}_k, \bar{y}_k)$. By SSCC, (B.11) and (B.16) imply

$$u(\bar{c}_k, \bar{y}_k, n_{k+1}) \geq u(\bar{c}_i, \bar{y}_i, n_{k+1}), \tag{B.17}$$

so that he also prefers the lottery to $(\bar{c}_i, \bar{y}_i)$ for any $i < k$. This completes the proof that the randomized allocation is downward incentive-compatible, as well as feasible.

The welfare that the randomized allocation generates is computed as:

$$\begin{aligned}\sum_{i=1}^{m} u(\bar{c}_i, \bar{y}_i, n_i) f_i + \frac{\varepsilon}{f_k}[u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_k) - u(\bar{c}_k, \bar{y}_k, n_k)] f_k \\ - \frac{\varepsilon}{f_{k+1}}[(u(\bar{c}_k, \bar{y}_k, n_{k+1}) - u(\bar{c}_{k+1}, \bar{y}_{k+1}, n_{k+1})] f_{k+1}.\end{aligned} \tag{B.18}$$

By Lemma 6.4, this welfare level is strictly greater than the welfare that is generated by the original allocation $\{(\bar{c}_i, \bar{y}_i)\}_{i=1}^{m}$.

Given that the randomized allocation dominates the allocation $\{(\bar{c}_i, \bar{y}_i)\}_{i=1}^{m}$, Lemmas 5.2 and 5.4 imply that there also exists a deterministic allocation which dominates the allocation $\{(\bar{c}_i, \bar{y}_i)\}_{i=1}^{m}$. The assumption that the solution to the deterministic relaxed income tax problem satisfies $\bar{c}_{i+1} < \bar{c}_i$ for some i has thus led to a contradiction and must be false. ■

References

[1] Arrow, K.J. (1965), *Aspects of the Theory of Risk-Bearing*, Yrjö Jahnsson Foundation, Helsinki.

[2] Berge, C. (1959), *Espaces Topologiques,* Dunod, Paris.

[3] Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, New York.

[4] Brito, D. L., J.H. Hamilton, S.M. Slutsky, and J.E. Stiglitz (1995), Randomization in Optimal Income Tax Schedules, *Journal of Public Economics* 56, 189 - 223.

[5] Guesnerie, R. (1995), *A Contribution to the Pure Theory of Taxation*, Econometric Society Monograph, Cambridge University Press, Cambridge, U.K.

[6] Hammond, P. (1979), Straightforward Individual Incentive Compatibility in Large Economies, *Review of Economic Studies* 46, 263 - 282.

[7] Hellwig, M.F. (2004), Risk Aversion in the Small and the Large with Multidimensional Outcomes, Preprint 2004/6, Max Planck Institute for Research on Collective Goods, Bonn.

[8] Hellwig, M.F. (2005), A Contribution to the Theory of Optimal Income Taxation, Preprint 2005/23, Max Planck Institute for Research on Collective Goods, Bonn.

[9] Kihlstrom, R.E., and L.J. Mirman (1974), Risk Aversion with Many Commodities, *Journal of Economic Theory 8, 361 - 388.*

[10] Luenberger, D.M. (1969), *Optimization by Vector Space Methods*, Wiley, New York.

[11] Maskin, E., and J. Riley (1984), Optimal Auctions with Risk Averse Buyers, *Econometrica* 52, 1473 - 1518.

[12] Matthews, S., and J. Moore (1987), Monopoly Provision of Quality and Warranties: An Exploration in the Theory of Multidimensional Screening, *Econometrica* 55, 441 - 467.

[13] Mirrlees, J.M. (1971), An Exploration in the Theory of Optimum Income Taxation, *Review of Economic Studies* 38, 175 - 208.

[14] Mirrlees, J.M. (1976), Optimal Tax Theory: A Synthesis, *Journal of Public Economics* 6, 327 - 358.

[15] Mirrlees, J.M. (1986), The Theory of Optimal Taxation, in: K.J. Arrow and M.D. Intriligator (eds.), *Handbook of Mathematical Economics*, North Holland, Elsevier Publishers, Amsterdam *et al.,* Vol. III, Ch. 24, 1198 - 1249.

[16] Pratt, J.W. (1964), Risk Aversion in the Small and the Large, *Econometrica* 32, 122 - 136.

[17] Royden, H.L. (1988), *Real Analysis*, Third Edition, Macmillan, New York and London.

[18] Stiglitz, J.E. (1982), Self-Selection and Pareto Efficient Taxation, *Journal of Public Economics* 17, 213 - 240.

[19] Stiglitz, J.E. (1987), Pareto Efficient and Optimal Taxation and the New New Welfare Economics, in: A. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics*, North Holland, Elsevier Publishers, Amsterdam *et al.*, 991 - 1042.

[20] Weibull, J. (1989), A Note on the Continuity of Incentive Schedules, *Journal of Public Economics* 39, 239 - 243.

[21] Weiss, L.W. (1976), The Desirability of Cheating Incentives and Randomness in the Optimal Income Tax, *Journal of Political Economy* 84, 1342 - 1352.