



**DEPARTMENT OF ECONOMICS**  
**DISCUSSION PAPER SERIES**

**APPLYING PERTURBATION ANALYSIS TO DYNAMIC  
OPTIMAL TAX PROBLEMS**

**Charles Brendon**

Number 581  
November 2011

Manor Road Building, Oxford OX1 3UQ

# Applying perturbation analysis to dynamic optimal tax problems\*

Charles Brendon  
University of Oxford

November 15, 2011

## Abstract

This paper shows how to derive a complete set of optimality conditions characterising the solution to a dynamic optimal income tax problem in the spirit of Mirrlees (1971), under the assumption that a ‘first-order’ approach to incentive compatibility is valid. The method relies on constructing perturbations to the consumption-output allocations of agents in a manner that preserves incentive compatibility for movements in both directions along the specified dimension. We are able to use it to generalise the ‘inverse Euler condition’ to cases in which preferences are non-separable between consumption and labour supply, and to prove a number of novel results about optimal income and savings tax wedges.

JEL Classification Codes: D82, E61, H21, H24

Keywords: New Dynamic Public Finance, first-order approach, non-separable preferences, inverse Euler condition

---

\*I would like to thank Ian Jewitt, Antonio Mele and Kevin Roberts for useful discussions relating to this work. All errors are mine.

# 1 Introduction

There is a growing interest among macroeconomists in dynamic optimal policy problems in the presence of asymmetric information. One such class of problems that has received particular attention is that of multi-period optimal tax analyses, based on the seminal works of Mirrlees (1971) and Diamond and Mirrlees (1978). Yet the complexity of the models in which this analysis is conducted has led to relatively few general analytical results emerging, of the kind that might confidently inform policy discussions.<sup>1</sup> Considerable progress has certainly been made under special assumptions regarding utility functions and skill distributions,<sup>2</sup> but in what ways the associated results generalise remains an open question. Indeed, the most clear (and most celebrated) analytical statement that *has* emerged – the so-called ‘inverse Euler condition’ – is itself particular to the quite strict requirement that consumption and labour supply should be separable in all agents’ preferences. In short, there is much theoretical work still to be done.

The aim of this paper is to contribute to that theoretical project. Working under the assumption that the ‘first-order approach’ is valid – so that the set of incentive compatibility constraints that binds at the optimum is known – we set out a novel perturbation method that is capable of providing a complete characterisation of that optimum. That is, we are able to obtain a set of distinct necessary optimality conditions exactly equal in number to the degrees of freedom available to the policymaker. In itself this result holds out the promise of substantially simplifying the numerical calculation of optimal tax schedules, which reduces to a ‘mechanical’ question of solving a given – albeit often very large – system of simultaneous equations (with need to use dynamic programming techniques). Perhaps more significantly, the conditions that we derive imply several important general results regarding the character of dynamic optimal tax schedules.

First, and of substantial theoretical interest, we are able to generalise the inverse Euler condition to situations in which an agent’s within-period consumption and labour levels are non-separable in utility. Second, using this result we are able to reach the general conclusion that optimal taxes should always deter savings (in a well-defined sense) when consumption and labour supply are either substitutes or separable from one another in preferences, but that this need not be true in the event that they are complements. Third, and again using our generalisation of the inverse Euler condition, we show that the long-run ‘immiseration’ results that are known to characterise dynamic Mirrlees economies with infinitely-lived dynasties under preference separability again generalise to the case of substitutes, but not of complements. The previous two results seem

---

<sup>1</sup>In the words of one prominent recent survey (Mankiw, Weinzierl and Yagan (2009)): “The theory of optimal taxation has yet to deliver clear guidance on a general system of history-dependent, coordinated labor and capital taxation ... Most of the recommendations of dynamic optimal tax theory are recent and complex.”

<sup>2</sup>Important recent contributions in this regard are Farhi and Werning (2010) and Golosov, Troshkin and Tsyvinski (2011).

closely related, and shed some light on the precise dynamics responsible for immiseration.

Fourth, we show that the set of ‘intratemporal’ optimality conditions characterising allocations in the simple case that skill distributions are iid is identical to the set of conditions that must hold in a static optimal income tax model, providing an important mapping from the traditional static literature to the more recent dynamic problems. But (fifth) when skills are Markov in a more general sense there is a reduction by one in the number of intratemporal optimality conditions – supplanted by an additional *intertemporal* condition, capturing the capacity of the policymaker to spread through time the distortions required to prevent more productive agents mimicking. This condition will generally imply higher marginal tax rates as time progresses for those whose productivity is low. Sixth, we show that it is never optimal in any time period to subsidise labour supply at the margin. Seventh (and finally) we show that effective marginal labour income tax rates will always be zero at the top of the skill distribution, in the event that this distribution has an upper support.

The key argument that lies behind all of these results is that if the first-order approach is valid then it is always possible to construct a set of perturbations to optimal (equilibrium) allocations such that *local* incentive compatibility constraints will continue to bind. That is, if we know that agent A is just envied by agent B in equilibrium (so that truth-telling is only weakly preferred to mimicking by the latter), we can construct simultaneous changes to the consumption and income levels of each such that the resulting increase in agent B’s utility from truth-telling is exactly the same as any increase in the utility he or she could obtain by mimicking agent A. If an allocation is optimal, such perturbations cannot be used to generate surplus resources, provided they additionally hold expected utility constant from the start of time.

Specifically, this approach requires that one should define perturbations to the optimal allocation that simultaneously satisfy three conditions: local incentive compatibility, reversibility, and welfare-neutrality. The first of these is particular to dynamic screening models: a perturbation to outcomes that changes the incentives for truthful reporting at the same time as it changes allocations will not generally be of use for our purposes, due to the discrete shifts in consumption and output patterns it could induce. We are interested, rather, in studying perturbations to *allocations* whilst holding constant agents’ *type reports* (exploiting the revelation principle to focus on a mechanism whereby agents report their idiosyncratic productivities directly).<sup>3</sup>

The second of the conditions is necessary if optimality conditions are to be stated with equality. It demands that if we can increase the consumption and output allocations of agents at a given time period along some marginal vector  $\Delta$ , then we can also increase them along the vector  $-\Delta$ . In a simple consumption-savings problem, this is the equivalent of noting that we must not

---

<sup>3</sup>The ‘static’ optimal income tax literature also makes use of perturbation analysis, but without exploiting direct revelation mechanisms: rather, the focus is directly on changes to the tax *schedule* subject to which all individuals choose. See, in particular, Roberts (2000) and Saez (2001).

be at a corner solution if we are to state the consumption Euler equation with equality.

The final requirement is that the perturbations should be welfare-neutral from the perspective of the policymaker in the initial time period. This is useful, since it means we can focus simply on whether any given perturbation raises surplus resources – in assessing whether it is to the advantage of the policymaker. Satisfying these three requirements for a broad class of perturbations – far broader than the intertemporal utility reallocations already applied in the literature when deriving the inverse Euler condition – is a non-trivial challenge, and establishing a general procedure for doing so forms the heart of the analysis in what follows.

## 1.1 Literature review

To date, two closely related methodological approaches to solving dynamic problems under asymmetric information have emerged in the macroeconomics literature. The first, and most widely-used, follows the foundational work on dynamic games by Abreu, Pearce and Stachetti (1990), considering directly the planner’s problem of maximising a given objective criterion subject to a series of lifetime utility constraints that must hold in each time period in equilibrium (preventing any incentive for agents to mis-report their private information). Examples include: Atkeson and Lucas (1992), investigating consumption allocations across agents subject to idiosyncratic taste shocks; Kocherlakota (1996), looking at consumption risk sharing when incomes are stochastic; Golosov, Troshkin and Tsyvinski (2011) in a dynamic tax setting; and numerous other papers besides. An important feature of these approaches is the reformulation of the policymaker’s problem to an equivalent recursive choice across current outcomes and a vector of discounted utility promises – the latter summarising the dynamic incentives that are being provided to ensure truthful reporting.

An important refinement to this method – particularly in the context of the present paper – has been provided by Kapička (2010) (extending the general work of Pavan, Segal and Toikka (2011)), who illustrated the potential of the ‘first-order approach’ to reduce the state-space required in dynamic Mirrlees models – particularly in the (realistic) event that agents’ productivities evolve according to non-iid processes. Specifically, Kapička demonstrates that one requires just two variables to summarise the policymaker’s past promises to an individual with a given history of productivity draws: a promised lifetime utility, and a value expressing how this utility changes at the margin as the agent’s type changes. This method substantially eases the computational burden associated with computing optimal allocations by the ‘primal’ (promised utilities) recursive technique, relative to existing methods valid under non-iid assumptions – notably that proposed by Fernandes and Phelan (2000). It has been adopted fruitfully by Farhi and Werning (2010) among others.

The second general approach, referred to as the ‘dual’ method by Messner, Pavoni and Sleet (2011), follows Marcet and Marimon (1998) in exploiting the evolution of *costates* associated with lifetime utility constraints, in order to aug-

ment the policymaker’s objective criterion in a manner that ensures incentive compatibility constraints are always satisfied. The problem is again set in a recursive form, but with no explicit choice over a set of future utilities; instead the Pareto weights that are placed on distinct agents’ utilities in the policy objective are increased exactly as necessary to ensure the resulting optimisation satisfies incentive compatibility. Following important work by Mele (2011), extending the work of Marcet and Marimon to repeated hidden action problems, this method has recently been applied to optimal dynamic tax policy by Sleet and Yeltekin (2010b). The latter authors have also provided an important general analysis applying the earlier theory to settings with private information (see Sleet and Yeltekin (2010a)).

Both of these methods arrive at solutions to the underlying problem through functional iteration on a Bellman-type operator. Whilst this has the advantage of quite widespread applicability, it necessitates numerical methods that may prevent the essential analytical character of the underlying solution from being completely clear. Rather than follow these papers in pursuing a variant upon the dynamic programming literature, here we instead develop a method more closely related to the calculus of variations. That is, we assume that an optimum has been found, and ask what properties that optimum would have to satisfy. This logic has already been applied in a relatively limited fashion by Kocherlakota (2005) and Golosov, Tsyvinski and Werning (2006), among others, to characterise intertemporal optimality in a dynamic Mirrleesian economy for which preferences between leisure and consumption are separable. It has been shown in particular that an ‘inverse Euler condition’ must hold in these circumstances, linking the marginal cost of providing consumption utility to a consumer in one time period to the expected value of the same marginal cost across distinct realisations for that consumer’s idiosyncratic productivity level in the next period. The marginal cost of providing consumption utility is the inverse of the marginal utility of consumption. The basic idea is that if an allocation is optimal the policymaker cannot transfer through time the provision of a utility to a consumer with a particular productivity history and raise a resource surplus.

Since the work of Thomas and Worrall (1990) (in the context of a repeated moral hazard model) the long-run implications of intertemporal optimality conditions of this form have been an important focus of study, particularly in the event that they define a bounded martingale sequence – which is generally the case when the real rate of interest is equal to consumers’ and the policymaker’s discount factor.<sup>4</sup> In this case long-run ‘immiseration’ results will generally follow by applying martingale convergence theorems, in the sense that the marginal utility of consumption for all agents will almost surely be unboundedly large as time progresses (its inverse will almost surely converge on zero). But these results rely heavily on the assumption of separability between consumption and

---

<sup>4</sup>Papers by Phelan (2006) and Farhi and Werning (2007) consider the implications of the social discount factor differing from individuals’, showing that the inverse Euler condition ceases to be a martingale in this event, so long as the real interest rate remains equal to the inverse of the household discount factor.

labour supply. In the event that this does not hold, the policymaker may find that an incentive-compatible perturbation requires changes to the labour supply of agents simultaneously to consumption changes, with the implication that the inverse of the marginal utility of consumption is no longer the marginal cost of (incentive-compatible) utility provision. We show that this is indeed the case in our optimal tax model.

## 2 Model setup

The basic framework that we use essentially follows the recent textbook treatment of Kocherlakota (2011), except that we allow for a general specification of preferences from the outset. An economy is populated by a large number of agents, modeled as a continuum with each agent indexed by a position on the unit interval. Each agent is the current manifestation of an infinitely-lived dynasty, and gains utility from that dynasty's consumption and leisure from the current period into the infinite future. Labour is the only factor of production and there are no firms – so agents can be thought of as directly choosing the level of output that they produce each period via their labour supply decision. Their preferences over output and consumption profiles from time  $t$  onwards are described by the function  $U_t$ :

$$U_t = E_t \sum_{s=0}^{\infty} \beta^s u(c_{t+s}, y_{t+s}; \theta_{t+s}) \quad (1)$$

where  $u : \mathbb{R}_+^3 \rightarrow \mathbb{R}$ .  $c_{t+s}$  and  $y_{t+s}$  are, respectively, the agent's consumption and output levels in period  $t + s$ ,  $\beta \in (0, 1)$  is the dynasty's time preference parameter, and  $\theta_{t+s}$  is an idiosyncratic productivity parameter that allows one to map from a level of output to a quantity of labour supply. The productivity parameter belongs to a set  $\Theta \subset \mathbb{R}$ , which is time- and history-invariant.<sup>5</sup> For the entirety of this paper we work under the assumption that  $\Theta$  contains a finite number of elements  $N$ , which turns out to provide the most straightforward setting to present the main arguments. We generalise to the (more conventional) assumption that  $\Theta$  is an interval of the real line in a companion paper (see Brendon (2011)). Though we assume an infinite horizon, none of the optimality conditions that we derive is dependent on this perspective.<sup>6</sup>

Expectations are taken across all stochastic variables relevant to the equilibrium evolution of the agent's utility (ultimately, drawings from  $\Theta$  at each future horizon). We analyse the model as if nature is responsible at the start of time for drawing a distinct element for each dynasty from the infinite-dimensional product space  $\Theta^\infty$ , say  $\theta^\infty$ , according to some probability measure on  $\Theta^\infty$ ,  $\pi_\Theta$ .

<sup>5</sup>The analysis is made simpler by assuming that  $\Theta$  itself does not depend on past draws. The probability of any one element of  $\Theta$  being drawn after a given history can always be made arbitrarily small, so this does not seem a particularly restrictive assumption.

<sup>6</sup>An interesting feature of our approach is that it provides a novel representation of the optimality requirements even in a 'static' optimal income tax model.

These draws are iid across dynasties. At the start of each time period agents are informed of their within-period productivity, so that at time  $t$  they are aware of their complete history of draws to date,  $\theta^t \in \Theta^t$ , where  $\theta^t$  is a  $t$ -length truncation of  $\theta^\infty$ . This purely idiosyncratic information is private knowledge to the agent, so policymakers must provide sufficient incentives to prevent mimicking in any tax system that is established.

To keep the problem regular we assume that the utility function is twice continuously differentiable in all of its arguments, with  $u_c > 0$ ,  $u_y < 0$ , and  $u_\theta > 0$ , and that the partial Hessian  $\begin{bmatrix} u_{cc} & u_{cy} \\ u_{cy} & u_{yy} \end{bmatrix}$  is negative definite for any given  $\theta$ . We additionally impose Inada conditions:  $\lim_{c \rightarrow \infty} u_c(c, y; \theta) = 0$  and  $\lim_{c \rightarrow 0} u_c(c, y; \theta) = \infty$  for all non-zero, finite  $(y, \theta)$  pairs, and  $\lim_{y \rightarrow \infty} u_y(c, y; \theta) = -\infty$  and  $\lim_{y \rightarrow 0} u_y(c, y; \theta) = 0$  for all non-zero, finite  $(c, \theta)$  pairs. These conditions will ensure an interior solution obtains at all finite horizons. To maintain the interpretation of the model as an optimal tax problem with unobservable labour supply we impose that marginal changes to  $\theta$  will reduce the marginal disutility associated with a unit of extra output when consumption and utility (and thus, implicitly, labour) are jointly held constant. This implies:

$$u_{y\theta} - u_{yy} \frac{u_\theta}{u_y} > 0 \quad (2)$$

Similarly, if consumption and utility are jointly held constant as  $\theta$  is changed then labour supply must implicitly also be being held fixed – and thus the marginal utility of consumption should likewise be constant. This is quite easily shown to imply the following:

$$u_{c\theta} - u_{cy} \frac{u_\theta}{u_y} = 0 \quad (3)$$

Finally, a variant upon the Spence-Mirrlees single-crossing condition is imposed, to ensure higher realisations of  $\theta$  can naturally be associated with higher productivity:

$$u(c'', y''; \theta'') - u(c', y'; \theta'') > u(c'', y''; \theta') - u(c', y'; \theta') \quad (4)$$

whenever  $c'' > c'$ ,  $y'' > y'$  and  $\theta'' > \theta'$ .

Note that this condition is slightly stronger than could be obtained simply by differentiating the expression for the slope of a within-period indifference curve in output-consumption space ( $\frac{dc}{dy} = -\frac{u_y}{u_c}$ ); although (4) *implies* that this indifference curve should be flattening in  $\theta$  (as seen by assuming one of the agents is indifferent between the two bundles), it also implies certain properties are associated with utility changes between bundles across which neither agent is indifferent, and it is useful to exploit these properties in what follows. Occasionally it is useful also to state the condition in terms of marginal rates of



substitution: if  $\theta'' > \theta'$  then condition (4) implies for all  $(c, y)$  pairs:

$$\frac{u_y(c, y; \theta'')}{u_c(c, y; \theta'')} < \frac{u_y(c, y; \theta')}{u_c(c, y; \theta')} \quad (5)$$

This follows directly from the fact that indifference curves in consumption-output space must be ‘flattening’ as  $\theta$  increases, provided (4) holds.

The policymaker’s role is to choose, at the start of the first time period, effective tax schedules for all future periods that will link an individual’s consumption to their output, conditional on their history of actions to date. The purpose of this choice is to maximise some social welfare function, defined across the set of possible equilibrium allocations. Individuals can be thought of as devising history-contingent action profiles to implement in each future time period, given the mechanism with which the policymaker presents them. Since the revelation principle will apply in this setting,<sup>7</sup> we may restrict policy choice to direct revelation mechanisms that deliver consumption and output bundles to individuals as functions of their current and past productivity reports – deferring a consideration of decentralisation schemes for subsequent work. In treating consumption as a choice variable of the policymaker in this way, we are implicitly assuming that there are no opportunities for the individuals to engage in ‘hidden’ saving – so that the policymaker could always behave as if taxing savings at a 100 per cent marginal rate, if this were necessary to prevent ‘unwanted’ consumption deferral.<sup>8</sup> We denote by  $\hat{\theta}_t^i : \Theta^t \rightarrow \Theta$  individual  $i$ ’s report at time  $t$  as a function of their actual productivity (where this function is measurable with respect to  $\theta^t$ ), by  $\hat{\theta}^{i,t} : \Theta^t \rightarrow \Theta^t$  the history of all such reports up to time  $t$ , and by  $\hat{\theta}^{i,\infty} : \Theta^\infty \rightarrow \Theta^\infty$  a complete sequence of reports. We occasionally refer to  $\hat{\theta}^{i,t}(\cdot)$  as the  $t$ -truncation of  $\hat{\theta}^{i,\infty}(\cdot)$ .

For the remainder of the paper we follow the majority of the literature and assume a utilitarian policy criterion, assessed from the perspective of the initial time period. This criterion has the advantage in a dynamic context of being the only objective that satisfies ‘welfarism’ at every horizon that is also time-consistent. That is to say, social welfare is assessed in each period as a function of individual lifetime utilities alone, and if two candidate policies deliver exactly the same outcomes between periods 1 and  $t$  then the relative preference of the policymaker between those two paths will be the same at time 1 as at time  $t$ . Whilst no claim is made that these normative features should be elevated above all others, they do arguably allow for the simplest treatment of the dynamic tax questions that are of chief interest to us.

<sup>7</sup>We seek a Bayes-Nash equilibrium of the game played between the policymaker and all individuals whose types may be drawn from  $\Theta^\infty$ . The revelation principle states that any such equilibrium can be supported by a direct revelation mechanism.

<sup>8</sup>Da Costa and Werning (2002) and Golosov and Tsyvinski (2006) consider economies with hidden savings opportunities; these substantially reduce the options available to the policymaker.

The policymaker's primal choice problem is, therefore:

$$\max_{\{c_t(\theta^\infty), y_t(\theta^\infty)\}_{t=1}^\infty} \int_{\Theta^\infty} \sum_{t=1}^\infty \beta^{t-1} u(c_t(\theta^\infty), y_t(\theta^\infty); \theta_t) d\pi_\Theta(\theta^\infty) \quad (6)$$

subject to  $c_t(\theta^\infty)$  and  $y_t(\theta^\infty)$  being measurable with respect to  $\theta^t$ , together with the incentive compatibility constraint:

$$\begin{aligned} & \int_{\Theta^\infty} \sum_{s=0}^\infty \beta^s u(c_{t+s}(\theta^\infty), y_{t+s}(\theta^\infty); \theta_{t+s}) d\pi_\Theta(\theta^\infty | \theta^t) \\ & \geq \int_{\Theta^\infty} \sum_{s=0}^\infty \beta^s u\left(c_{t+s}(\hat{\theta}^\infty(\theta^\infty)), y_{t+s}(\hat{\theta}^\infty(\theta^\infty)); \theta_{t+s}\right) d\pi_\Theta(\theta^\infty | \theta^t) \end{aligned} \quad (7)$$

which must hold for all  $t$ , all  $\theta^t$ , and all functions  $\hat{\theta}^\infty : \Theta^\infty \rightarrow \Theta^\infty$  whose  $s$ -truncations  $\hat{\theta}^s(\cdot)$  are measurable with respect to  $\theta^s$  for all  $s \geq 1$ ; and finally the resource constraint:

$$\int_{\Theta^\infty} [c_t(\theta^\infty) - y_t(\theta^\infty)] d\pi_\Theta(\theta^\infty) + A_{t+1} = R_t A_t \quad (8)$$

where  $A_t$  is the quantity of real bonds that the policymaker purchases for time  $t$ , each paying  $R_t$  units of real income. The initial value  $R_1 A_1$  is given. Dynamic solvency requires that  $\lim_{s \rightarrow \infty} \left[ \left( \prod_{r=1}^s R_{t+r}^{-1} \right) A_{t+s} \right] = 0$ .<sup>9</sup>

### 3 Full information benchmark

In a manner equivalent to Kapička (2010) and Broer, Kapička and Klein (2011), we will ultimately focus our attention on a relaxed version of the incentive compatibility constraint, arguing (in the context of a discrete number of types in  $\Theta$ ) that it is sufficient to impose a binding restriction to prevent agents with histories  $(\theta^{t-1}, \theta_t)$  mimicking those with histories  $(\theta^{t-1}, \theta'_t)$ , where  $\theta'_t = \max\{\theta \in \Theta : \theta < \theta_t\}$ . The basic reason for our making this assumption – that envy is always directed ‘downwards’ from one type to the next in equilibrium – is familiar from the analysis of static optimal tax models, and was articulated most clearly by Dasgupta (1982). To understand why it is likely to hold, it is useful to start by considering the character of optimal policy when the idiosyncratic productivity draws are common knowledge.

If the policymaker is aware of agents' types each period the incentive compatibility constraint (7) can be neglected, with lump-sum taxation used to implement a first-best. We summarise four important properties of this first-best

---

<sup>9</sup>In what follows we will sometimes suppress the explicit dependence of  $c_t$  and  $y_t$  upon  $\theta^\infty$ , as well as indexing these functions with individual-specific superscripts where this is most natural.

in the following list. The proofs of each statement are trivial, and hence omitted, with the exception of the fourth, which is provided in the appendix.

1. In the full information benchmark the optimal allocations  $c_t(\theta^\infty)$  and  $y_t(\theta^\infty)$  are measurable with respect to  $\theta_t$ .
2. The following conditions hold for all  $t \geq 1$  and all  $i \in [0, 1]$ :

$$u_c(c_t^i, y_t^i; \theta_t^i) = -u_y(c_t^i, y_t^i; \theta_t^i) \quad (9)$$

$$u_c(c_t^i, y_t^i; \theta_t^i) = \beta R_{t+1} \sum_{\theta_{t+1}^i \in \Theta} u_c(c_{t+1}^i, y_{t+1}^i; \theta_{t+1}^i) \pi_\Theta(\theta_{t+1}^i | \theta_t^i) \quad (10)$$

3. The following condition holds for all  $t \geq 1$  and all agents  $i, j \in [0, 1]$ :

$$u_c(c_t^i, y_t^i; \theta_t^i) = u_c(c_t^j, y_t^j; \theta_t^j) \quad (11)$$

In the event that consumption and labour are additively separable in the utility function we will additionally have  $u_{cy} = u_{c\theta} = 0$ , and this condition then implies equalised consumption across all agents (since  $u_{cc} < 0$ ).

4.  $\theta_t^i > \theta_t^j$  implies  $u(c_t^i, y_t^i; \theta_t^i) < u(c_t^j, y_t^j; \theta_t^j)$ , so long as leisure is a normal good when income is untaxed at the margin.

Summarising the main lessons of these four statements in turn, we know from the first that there is no incentive for the policymaker to introduce any form of history dependence in outcomes. The fact that a particular individual has been very productive in the past makes no difference to their current consumption-output bundle, independently of the contemporary productivity draw  $\theta_t$ . In this sense the first-best solution offers no scope for agents to claim credit for past accomplishments. The second statement implies that the optimal solution for a utilitarian policymaker involves zero marginal distortions on savings and labour supply, whilst the third points to equalised marginal consumption utility (and, thus, output disutility) across agents each period. Since agents who are more productive have, by definition, a higher marginal product for a *given* quantity of labour they will generally be required to work longer hours at the optimum. This is the logic behind the fourth condition – that utility is decreasing in type so long as leisure is a normal good.<sup>10</sup> This last result is key to understanding which incentive compatibility constraints will bind at the optimum: together with the fact that there is no history dependence in outcomes at the first-best, it strongly implies higher-type agents would mimick their lower-type peers if they had the capacity to do so – that is, in the event that the policymaker could only verify agents' output levels, and not their types.

---

<sup>10</sup>It is, indeed, a case of 'From each according to his means, to each according to his needs.'

## 4 The first-order approach to incentive compatibility

We now move to the constrained problem, in which the policymaker is forced to abide by incentive compatibility constraints – and hence will be prevented from providing higher-productivity types with a lower level of utility than their (lower-productivity) peers. As mentioned above, we retain a focus on the case in which  $\Theta$  contains a discrete, finite number of elements. To apply our perturbation method, we first need to be clearer on the set of constraints that will bind at the optimum.

For all periods  $t \geq 1$ , define  $\hat{\theta}_{m,t}^\infty : \Theta^\infty \rightarrow \Theta^\infty$  as the reporting strategy associated with truth-telling in all periods up to  $t$ , at which point the agent mimics a type one lower and follows an optimal reporting strategy thereafter:

$$\hat{\theta}_{m,t}^\infty(\theta'^\infty) = [\theta'_1, \theta'_2, \dots, \theta'_{t-1}, \theta''_t, \theta''_{t+1}, \dots]$$

where  $\theta''_t = \max\{\theta \in \Theta : \theta < \theta'_t\}$  and  $\{\theta''_{t+1}, \theta''_{t+2}, \dots\}$  are then optimal choices conditional upon prior reports. So long as the type distribution is Markov, outcomes for an agent with a given reporting history will in fact be independent of the veracity of that reporting history – so we are free to focus exclusively on ‘one-off’ deviations from the truth, with  $\{\theta''_{t+1}, \theta''_{t+2}, \dots\} = \{\theta'_{t+1}, \theta'_{t+2}, \dots\}$ . If  $\theta'_t = \min\{\theta \in \Theta\}$  then we normalise  $\hat{\theta}_{m,t}^\infty(\theta^\infty) = \theta^\infty$ . If incentive compatibility is said to hold ‘downwards’, the following is true:

$$\begin{aligned} & \int_{\Theta^\infty} \sum_{s=0}^{\infty} \beta^s u(c_{t+s}(\theta^\infty), y_{t+s}(\theta^\infty); \theta_{t+s}) d\pi_\Theta(\theta^\infty | \theta^t) \\ & \geq \int_{\Theta^\infty} \sum_{s=0}^{\infty} \beta^s u\left(c_{t+s}(\hat{\theta}_{m,t}^\infty(\theta^\infty)), y_{t+s}(\hat{\theta}_{m,t}^\infty(\theta^\infty)); \theta_{t+s}\right) d\pi_\Theta(\theta^\infty | \theta^t) \end{aligned} \quad (12)$$

So the agent with history  $\theta^t$  is just indifferent between reporting  $\theta_t$  truthfully and mimicking a type one lower, provided  $\theta_t$  is not itself the smallest element in  $\Theta$ . Again, for any Markovian productivity process it must be true that if (12) holds for agents whose past reports of  $\hat{\theta}^{t-1}$  were truthful, it must also hold for all agents with identical past reports and a true contemporary productivity draw equal to  $\theta_t$ .<sup>11</sup> We are interested in the conditions under which this restriction implies *global* incentive compatibility – that is, for an arbitrary reporting

---

<sup>11</sup>To see this, consider the ‘arbitrary’ reporting strategy  $\hat{\theta}_{a,t}^\infty(\theta^\infty) = [\theta^{t-1}, \theta_t, \theta_{t+1}, \dots]$  for some  $\theta^{t-1} \in \Theta^{t-1}$ . From time  $t$  onwards this will deliver the same utility as is received by a permanently truth-telling agent whose history of shocks prior to  $t$  was indeed  $\theta^{t-1}$ , and has current type  $\theta_t$  (assuming that productivity draws are Markov). But these two agents will also receive identical expected utilities to one another from mimicking a marginally lower type at  $t$ ; and for the prior truth-teller we know that the incentive compatibility constraint binds – implying it must also do so for the agent whose prior reports were arbitrary.

strategy at  $t$ ,  $\hat{\theta}_{a,t}^\infty : \Theta^\infty \rightarrow \Theta^\infty$ , defined by:

$$\hat{\theta}_{a,t}^\infty(\theta'^\infty) = [\theta'_1, \theta'_2, \dots, \theta'_{t-1}, \theta'_t, \theta''_{t+1}, \dots]$$

for any  $\theta''_t \in \Theta$ , with  $\{\theta''_{t+1}, \theta''_{t+2}, \dots\}$  chosen optimally thereafter, we want to know when it will be the case that equation (12) implies:

$$\begin{aligned} & \int_{\Theta^\infty} \sum_{s=0}^{\infty} \beta^s u(c_{t+s}(\theta^\infty), y_{t+s}(\theta^\infty); \theta_{t+s}) d\pi_\Theta(\theta^\infty | \theta^t) \\ & \geq \int_{\Theta^\infty} \sum_{s=0}^{\infty} \beta^s u\left(c_{t+s}\left(\hat{\theta}_{a,t}^\infty(\theta^\infty)\right), y_{t+s}\left(\hat{\theta}_{a,t}^\infty(\theta^\infty)\right); \theta_{t+s}\right) d\pi_\Theta(\theta^\infty | \theta^t) \end{aligned} \quad (13)$$

This problem lies at the heart of discussions on the applicability of the first-order approach in problems of this kind – an issue first considered by Mirrlees (1971), and studied in great depth in the context of dynamic models by Pavan, Segal and Toikka (2011). The first-order approach takes as its starting point the fact that under any incentive-compatible direct revelation mechanism no agent can induce an increase in their expected lifetime utility by changing their report. We can define the value function  $W(\hat{\theta}_t; \theta_t, \hat{\theta}^{t-1})$ , with  $W : \Theta \times \Theta \times \Theta^{t-1} \rightarrow \mathbb{R}$  specifying the maximum lifetime utility that could be expected for an agent whose past reports were  $\hat{\theta}^{t-1}$ , whose current productivity is  $\theta_t$  and whose current report is  $\hat{\theta}_t$ . Then the approach notes that for a given  $(\theta_t, \hat{\theta}^{t-1})$  pair this function must have a global maximum where  $\hat{\theta}_t = \theta_t$ . Thus instead of choosing directly from among the (difficult to characterise) set of allocations for which condition (13) is explicitly asserted for all admissible functions  $\hat{\theta}_{a,t}^\infty$ , one may instead choose simply from the set for which  $W(\cdot; \theta_t, \hat{\theta}^{t-1})$  is known to have a stationary point at  $\theta_t$ . In the case that a discrete number of types features in  $\Theta$  (rather than  $\Theta$  being a proper subset of the real line), it is not directly apparent what this implies: we cannot place a restriction on the derivative of  $W$  with respect to  $\hat{\theta}_t$  if there is no possibility of marginal changes to the agent's report. Yet we may invoke our earlier result that the first-best optimum involves decreasing utility in  $\theta$  to apply a 'first-order' approach in which choice is from the set of allocations such that the condition:

$$W(\theta_t; \theta_t, \hat{\theta}^{t-1}) \geq W(\theta'_t; \theta_t, \hat{\theta}^{t-1}) \quad (14)$$

is imposed for  $\theta'_t = \max\{\theta \in \Theta : \theta < \theta_t\}$ .<sup>12</sup> That is, consistent with the familiar logic of the Mirrlees model, incentive compatibility must be imposed 'downwards'. It should be stressed that in general condition (14) is not sufficient for  $\theta_t \in \arg\hat{\theta}_t \max W(\hat{\theta}_t; \theta_t, \hat{\theta}^{t-1})$  to hold, though it certainly is necessary; the validity of the approach needs to be checked carefully in any given case.

---

<sup>12</sup>No restriction is imposed in the event that  $\theta_t = \min\{\theta \in \Theta\}$ .

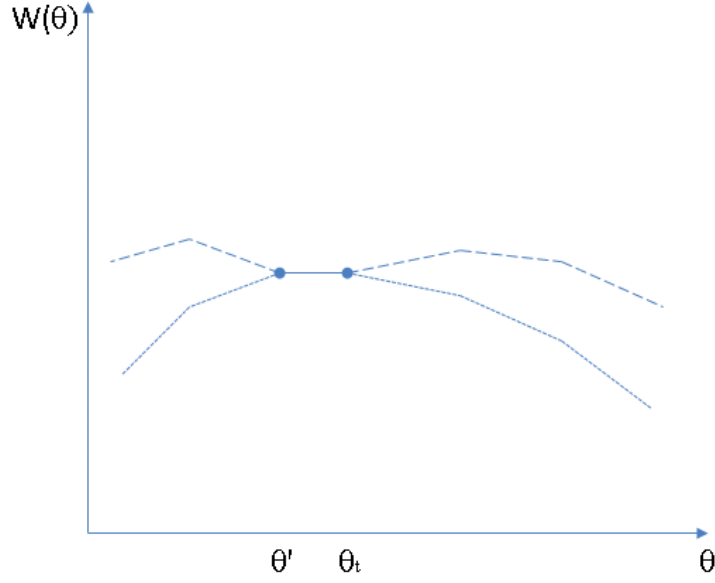


Figure 1: Local incentive compatibility need not imply global.

Graphically, the potential pitfalls of the approach are illustrated using Figure 1.

The vertical axis here denotes the value of  $W\left(\cdot; \theta_t, \hat{\theta}^{t-1}\right)$  for all given values of an agent's  $t$ -dated type report, which is mapped on the horizontal axis. To be sure that the incentive compatibility constraints are binding across all potential reports we would need to impose that this function is maximised at  $W\left(\theta_t; \theta_t, \hat{\theta}^{t-1}\right)$ . Since this is an onerous requirement, as noted, our 'first-order' approach instead asserts simply that condition (14) must hold. We assume it is binding, and represent this by the horizontal line linking the value of  $W\left(\cdot; \theta_t, \hat{\theta}^{t-1}\right)$  at the relevant arguments in Figure 1. But knowing that  $W\left(\cdot; \theta_t, \hat{\theta}^{t-1}\right)$  does not change between  $\theta'_t$  and  $\theta_t$  is clearly not the same as knowing that  $\theta_t \in \arg\max_{\hat{\theta}_t} W\left(\hat{\theta}_t; \theta_t, \hat{\theta}^{t-1}\right)$ . Whilst the rest of the value function certainly *may* be characterised by gradual and steady decay from the maximum, as in the case of the 'dotted' (lower) line in the figure, we equally cannot rule out the possibility of higher values being obtained elsewhere – as in the case of the higher 'dashed' line. The general point (which extends to cases in which  $\Theta$  is a continuum) is that the first-order approach admits a broader set of possible policies than the underlying incentive compatibility constraints, and unless one knows something about the properties of  $W\left(\cdot; \theta_t, \hat{\theta}^{t-1}\right)$  away from

$\theta_t$  and  $\theta'_t$  one can never be sure that a given case satisfying condition (14) will *additionally* satisfy the full constraint set.

For this reason the following result proves useful. The proof can be found in the appendix.

**Proposition 1** *Suppose the type set  $\Theta$  contains only a finite number of elements and that under a given policy strategy the value function  $W(\hat{\theta}_t; \theta_t, \hat{\theta}^{t-1})$  satisfies increasing differences in  $(\hat{\theta}_t, \theta_t)$ , so that the inequality*

$$W(\hat{\theta}_t''; \theta_t'', \hat{\theta}^{t-1}) - W(\hat{\theta}_t'; \theta_t'', \hat{\theta}^{t-1}) > W(\hat{\theta}_t''; \theta_t', \hat{\theta}^{t-1}) - W(\hat{\theta}_t'; \theta_t', \hat{\theta}^{t-1})$$

*holds for all  $(\hat{\theta}_t'', \hat{\theta}_t', \theta_t'', \theta_t') \in \Theta^4$  such that  $\hat{\theta}_t'' > \hat{\theta}_t'$  and  $\theta_t'' > \theta_t'$ . Then if condition (14) is known to hold with equality for all  $\theta_t \in \Theta \setminus \underline{\theta}$  and all histories  $\hat{\theta}^{t-1} \in \Theta^{t-1}$  it must also be that  $W(\theta_t; \theta_t, \hat{\theta}^{t-1}) > W(\theta_t'; \theta_t, \hat{\theta}^{t-1})$  holds for all  $\theta_t' \in \Theta \setminus (\theta_t', \theta_t)$  (where  $\theta_t' = \max\{\theta \in \Theta : \theta < \theta_t\}$  and  $\underline{\theta} = \min\{\theta \in \Theta\}$ ).*

This is a natural translation to our discrete-type setting of Theorem 5 in Kapička (2011).<sup>13</sup> Like that result, it is only an intermediate step in providing sufficiency conditions for the first-order approach, since the value function in any given setting will itself depend endogenously upon the chosen policy. But the fact that we have a sufficient condition for knowing incentive compatibility will bind for all non-truthful reporting strategies is useful in supporting the arguments that follow. It implies that a solution to the problem in which just condition (14) is imposed will also be a solution to the full problem (subject to the entire constraint set), provided the former solution exhibits the given increasing differences property. Moreover, combined with the single-crossing condition we have enough here to assert something much stronger about the iid case, which we present in the following Corollary:

**Corollary 2** *Suppose that agent-level productivities follow an iid process, and that the single-crossing condition (4) applies. Then provided a given policy strategy requires higher-type agents with a given history to produce higher output quantities than lower-type agents with the same history, and simultaneously provides them with higher consumption (in the period in which these productivities obtain), condition (14) holding with equality is sufficient for incentive compatibility*

**Proof.** When productivity shocks are iid, agents' future values (from  $t+1$  on) for a given report  $\hat{\theta}^t$  are identical in expectation from the perspective of time  $t$ , irrespective of their true types. Hence increasing differences will follow provided we have, under the given policy:

$$u(\hat{\theta}_t''; \theta_t'', \hat{\theta}^{t-1}) - u(\hat{\theta}_t'; \theta_t'', \hat{\theta}^{t-1}) > u(\hat{\theta}_t''; \theta_t', \hat{\theta}^{t-1}) - u(\hat{\theta}_t'; \theta_t', \hat{\theta}^{t-1})$$

<sup>13</sup>That theorem imposes that the derivative of  $W$  with respect to  $\theta_t$  should be increasing in  $\hat{\theta}_t$ .

where  $u(\hat{\theta}_t; \theta_t, \hat{\theta}^{t-1})$  is used to denote  $u(c_t(\hat{\theta}^t(\theta^\infty)), y_t(\hat{\theta}^t(\theta^\infty)); \theta_t)$  for  $\hat{\theta}^t(\theta^\infty) = (\hat{\theta}^{t-1}, \hat{\theta}_t)$ , for all  $(\hat{\theta}_t'', \hat{\theta}_t', \theta_t'', \theta_t') \in \Theta^4$  such that  $\hat{\theta}_t'' > \hat{\theta}_t'$  and  $\theta_t'' > \theta_t'$ . The result is then a direct implication of the single crossing condition, noting the assumption that output and consumption are increasing in type. ■

Whilst this result clearly still depends on the optimum having the particular property that output and consumption are increasing in type (for agents with a common reporting history), this is a very straightforward condition to check in any particular calculated example, and the analysis below suggests it will indeed hold under the optimal policy chosen from the set of direct revelation mechanisms satisfying condition (14).

In what follows we refer to the problem of policy choice from among the set of direct revelation mechanisms satisfying condition (14) as the ‘relaxed’ problem, in contrast with the ‘general’ problem that imposes  $\theta_t \in \arg_{\hat{\theta}_t} \max W(\hat{\theta}_t; \theta_t, \hat{\theta}^{t-1})$  directly for all  $(\theta_t, \hat{\theta}^{t-1}) \in \Theta^t$ .<sup>14</sup> Our focus will be on the properties of the solution to this relaxed problem, under the assumption that the solution to it coincides with the solution to the general problem. If this is the case, then we know that any other candidate policy that satisfies the constraint set of the relaxed problem is inferior from the policymaker’s perspective to the solution to the general problem. We exploit this fact in what follows, showing how to perturb allocations in such a way that the constraint set of the relaxed problem must remain satisfied – and hence allocations must be inferior to the general problem’s optimum.

## 5 Applying perturbation analysis

### 5.1 A diagrammatic primer

This section introduces the main focal point of our analysis: how one can apply local perturbations to optimal consumption and output allocations in order to obtain a set of conditions that the optimal tax system must satisfy. The analysis is easiest to understand with the aid of an indifference curve map, linking output on the horizontal axis and consumption on the vertical. To make the relevant ideas concrete, and allow us to illustrate some important intuition for the dynamic tax problem, Figure 2 shows such a mapping.

The two within-period indifference curves are drawn for arbitrary distinct types  $\theta_1$  and  $\theta_2$ , with  $\theta_2 > \theta_1$ . The diagram can be used to show intuitively why positive effective marginal income taxes are desirable at a constrained optimum. Recall that the first-best allocation involved consumption-output bundles for each agent such that  $u_c = -u_y$ . Diagrammatically this would correspond to a situation in which each agent’s bundle is such that their indifference curve has a slope of 1, as if there is no taxation of income at the margin. We may suppose for illustrative purposes that this is true of points A (for the agent of type  $\theta_1$ ) and

<sup>14</sup>We distinguish between ‘relaxed’ and ‘general’ constraint sets in analogous fashion.



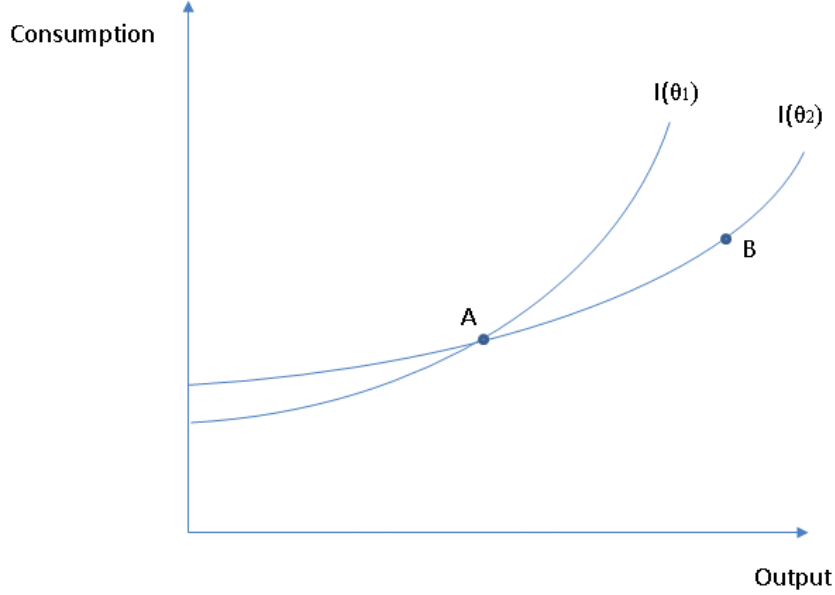


Figure 2: Graphical introduction to the policy trade-off

B (for type  $\theta_2$ ) in the diagram. We also know that at a first-best allocation the marginal utility of consumption would be equalised across agents (representable in the case of separable utility by equalised consumption across all agents), and that there would be no history dependence in allocations.

When incentive compatibility constraints must additionally be satisfied these conditions can no longer be satisfied simultaneously. Figure 2 shows a situation in which the policymaker has chosen to violate just one of them: the equality of marginal utility across agents. Type  $\theta_2$  consumes at B, and is entirely compensated within the present period for choosing not to mimic type  $\theta_1$  (which would involve consumption at A). Yet this does not correspond to a second-best allocation. As the analysis of static optimal income tax problems has shown, the policymaker can improve on this outcome by violating productive optimality for the lower-type agent. If type  $\theta_1$  is asked to produce at a point somewhere to the south-west of A along the curve  $I(\theta_1)$ , the ‘information rent’ that the higher-type agent can extract will be reduced, in the sense that the utility level type  $\theta_2$  could obtain by mimicking type  $\theta_1$  would fall, reducing the need for (wasteful) compensation – and thus freeing up resources to be redistributed to lower-type agents. Thus, perhaps counterintuitively, an equilibrium in which  $\theta_1$  is dissuaded from producing at the margin, via positive marginal income taxes, may be better for that agent than one in which there are zero marginal taxes.

Identical ‘second best’ logic may be applied to assert the desirability of spreading incentives through time. Rather than ensuring that the higher-type

agent is just indifferent within a period between truthful reporting and mimicking, it will be preferable for that agent's within-period utility to be reduced – generating a strictly positive marginal benefit when the associated resources are redistributed – and for their future utility instead to be raised in expectation by an offsetting amount. At least in the iid case, the latter distortion will initially come at zero marginal cost when one starts from a situation in which there is no history dependence – since any resources used to provide future incentives may be treated as if coming from the future allocations of agents whose utility profiles across future states are initially identical to that of the agent under study. Hence the theory of the second best applies, and it will be better to introduce dynamic distortions in addition.

## 5.2 Developing a perturbation approach

Our purpose is to make formal the intuition highlighted in the preceding discussion. The presumption throughout is that the policymaker is able to solve the ‘relaxed’ problem, in which equation (14) replaces the complete constraint set, and that the solution to this problem coincides with the solution to the general problem in which the full constraint set is imposed. Conditional upon a particular reporting history prior to the current period  $t$ ,  $\hat{\theta}^{t-1}$ , an agent's time- $t$  report-contingent consumption and output allocations under the optimal scheme can be described by an  $N \times 2$  matrix  $X_t^* \left( \hat{\theta}^{t-1} \right)$ , with each row in this matrix corresponding to a particular  $\hat{\theta}_t \in \Theta$ ,<sup>15</sup> and the columns listing, in turn, consumption and output levels for the given reported productivity draw. Our aim is to show how these allocations can be perturbed by the addition of one or more of a particular set of  $N \times 2$  matrices of continuously differentiable parametric functions, which in the generic case we denote by  $\Delta(\delta)$  (with  $\Delta : \mathbb{R} \rightarrow \mathbb{R}^{2N}$ ) for some relevant parameter  $\delta$  (perhaps the consumption or utility increment implied by the given perturbation for an agent of the highest type). These functions are always normalised such that  $\Delta(0) = 0$ . In certain cases we will additionally allow changes to be spread through time, with the consumption and output of agents with a common reporting history  $\hat{\theta}^{t-1}$  changed at  $t-1$  (as well as at  $t$ ), according to an analogous function  $\Delta_{-1}(\delta)$  (with  $\Delta_{-1} : \mathbb{R} \rightarrow \mathbb{R}^2$ ). We wish to construct these  $\Delta$  and  $\Delta_{-1}$  functions so that they satisfy the following three properties:

1. Incentive compatibility constraints that bind under the relaxed problem for each time period up to the  $t$ th when allocations for agents with reporting history  $\hat{\theta}^{t-1}$  are  $\left( c_{t-1}^* \left( \hat{\theta}^{t-1} \right), y_{t-1}^* \left( \hat{\theta}^{t-1} \right) \right)$  at  $t-1$  and  $X_t^* \left( \hat{\theta}^{t-1} \right)$  at  $t$  will continue to bind when allocations are  $\left( c_{t-1}^* \left( \hat{\theta}^{t-1} \right), y_{t-1}^* \left( \hat{\theta}^{t-1} \right) \right) + \Delta_{-1}(\delta)$  at  $t-1$  and  $X_t^* \left( \hat{\theta}^{t-1} \right) + \Delta(\delta)$  at  $t$ . Hence the perturbed allocations

---

<sup>15</sup>We assume that these are ordered in ascending values for  $\hat{\theta}_t$ , with the lowest (reported) type's allocation in the first row of  $X^*$  and the highest type's in the  $N$ th row.

tions are candidate solutions to the relaxed problem.

2.  $\Delta(\delta)$  and  $\Delta_{-1}(\delta)$  should be both continuous and continuously differentiable in an open neighbourhood of the point  $\delta = 0$ .
3. The expected utility of all agents should remain constant from the perspective of the initial time period for all values of  $\delta$  in the neighbourhood of  $\delta = 0$ .

Since we work under the assumption that incentive compatibility constraints bind only ‘downwards’ in the relaxed problem, the first property is equivalent to requiring that any additional incentive that an agent of type  $\theta_t^n$  may have to mimic an agent of type  $\theta_t^{n-1}$  (through changes in the allocation that the latter agent receives) is offset by an *equal* increase in the utility that the agent of type  $\theta_t^n$  receives from truthful reporting.<sup>16</sup> Symmetrically, we impose that a reduction in the incentives to mimic should be matched by an equal reduction in the utility from truthful reporting – preserving continuity in the construction at  $\delta = 0$ . This ensures that if the original allocation satisfied the constraint set of the relaxed problem then the perturbed allocation must likewise. Hence if the original allocation was a *solution* to the relaxed problem, the perturbed allocation cannot deliver the same value to the policymaker at lower cost.

The second condition is required for the perturbations to be applied symmetrically. It is very similar to the requirement in consumer choice theory that optimal consumption should be at an interior point in an agent’s budget set if we are to assert that the price ratio will be *equal* to that agent’s marginal rate of substitution between two goods (and that a unique marginal rate of substitution should exist at the optimal point) – otherwise it may not be possible for the consumer to exploit any wedge that exists between the two. This requirement provides a substantial obstacle relative to the first: if we know that incentive compatibility constraints bind downwards then we know it always going to be possible to *increase* the utility of the highest type alone, or of the top  $n$  types in sufficiently skewed proportions, so that incentive compatibility constraints will remain satisfied. This could be done simply by the provision of extra consumption to these higher-type agents. But perturbations of this form will only ever give us *inequality* restrictions – to the effect that the net marginal cost of changing outcomes in such a manner must be weakly positive. Unless a symmetric *downward* shift in the utility of high types is possible, with a converse impact on the net cost of utility provision, this cannot be converted into a first-order condition that is stated with *equality*.

As the third condition states, we assume that allocations are changed in just such a way that expected utility across periods  $t - 1$  and  $t$  is held constant – in a manner already familiar in the literature from the use of perturbation analysis to derive the inverse Euler condition in the case of separable preferences (see, for instance, Golosov, Tsyvinski and Werning (2006)). Since we have assumed

---

<sup>16</sup>We use superscripts here to index the agents’ types within the set  $\Theta$ , with  $\theta_t^n$  increasing in  $n \in \{1, \dots, N\}$

a policymaker who is utilitarian, assessing outcomes from the perspective of the initial time period, this implies that in all cases the policymaker will experience no direct loss or gain from the perturbation.

A necessary condition for the original allocations  $(c_{t-1}^*, y_{t-1}^*)$  and  $X_t^*$  to have been optimal is, then, that the marginal effect on the *resource cost* of utility provision associated with the perturbations should be zero. Otherwise it would be possible to change allocations in one direction or another and raise a resource surplus, without changing the value of the policymaker's objective – contradicting the presumed optimality of the original allocation.

### 5.3 Deriving admissible perturbations: changes at the top

There is a very simple example of a perturbation that satisfies all three of the above requirements: a movement along the within-period indifference curve of the ‘top’ agent for any given reporting history. Since the famous work by Mirrlees (1971) it has been well understood that the maxim ‘no distortion at the top’ applies in a static optimal income tax setup – in the sense that  $u_c = -u_y$  for any agent whose productivity parameter takes the highest possible value in the feasible set.<sup>17</sup> This derives from the fact that no other agent envies the allocation of the highest type in equilibrium – and thus there are no benefits in moving *away* from a situation in which  $u_c = -u_y$ .<sup>18</sup> The logic generalises to the intertemporal model, as the following makes clear.

**Proposition 3** *Suppose the solution to the relaxed problem also solves the general problem. Then in all time periods  $t \geq 1$  and (if  $t > 1$ ) for all past reporting histories  $\hat{\theta}^{t-1}$ , the allocation  $(c_t^*, y_t^*)$  for the agent who reports  $\hat{\theta}_t'$  such that  $\theta_t' = \max \{\theta \in \Theta\}$  satisfies  $u_c(c_t^*, y_t^*; \theta_t') = -u_y(c_t^*, y_t^*; \theta_t')$ .*

**Proof.** Consider a perturbation to the allocation  $X_t^*(\hat{\theta}^{t-1})$  given by the  $N \times 2$  matrix of functions  $\Delta : \mathbb{R} \rightarrow \mathbb{R}^{2N}$  such that the  $n$ th row of  $\Delta(\delta)$  equals  $(0, 0)$  for all  $n \in \{1, \dots, N-1\}$  and the  $N$ th row equals  $(\delta, f(\delta))$ , with the function  $f : \mathbb{R} \rightarrow \mathbb{R}$  defined implicitly by:

$$u(c_t^* + \delta, y_t^* + f(\delta); \theta_t') = u(c_t^*, y_t^*; \theta_t') \quad (15)$$

By construction this change keeps constant the (expected) utility of all truth-telling agents in all time periods. It does affect the utility of agents who report  $\hat{\theta}_t'$  when not of type  $\theta_t'$ , but this is irrelevant to the relaxed problem, and by the initial supposition in the Proposition we know that any allocation that continues to satisfy the relaxed constraint set cannot improve upon the solution

<sup>17</sup>When this set has unbounded support the result need no longer hold, as the influential work by Saez (2001) has emphasised.

<sup>18</sup>For all other agents, reducing consumption and output together along a given indifference curve, to a point where  $u_c > -u_y$ , will reduce the utility ‘rent’ that must be provided to higher types to deter mimicking – a consideration that justifies deviating from the usual productive optimality condition in their case, as the discussion above highlighted.

to the general problem. This then implies that the value of the policymaker's objective will remain unchanged as  $\delta$  is varied away from  $\delta = 0$ . The impact of the perturbation on the resources available to the policymaker in period  $t$  (in a truth-telling equilibrium) will be  $\pi_{\Theta}(\theta'_t|\hat{\theta}^{t-1})\pi_{\Theta}(\hat{\theta}^{t-1})[f(\delta) - \delta]$ . If the original allocation is optimal then the *marginal* impact on resources as  $\delta$  moves away from zero must be zero, or else it would be possible to raise a surplus. Hence we have:

$$\pi_{\Theta}(\theta'_t|\theta^{t-1})\pi_{\Theta}(\theta^{t-1})[f'(0) - 1] = 0 \quad (16)$$

Probabilities are non-zero, so this implies:

$$f'(0) = 1 \quad (17)$$

Since utility for a highest-type truth-teller is unchanged by the perturbation we can assert the total derivative:

$$u_c(c_t^*, y_t^*; \theta'_t) + u_y(c_t^*, y_t^*; \theta'_t) f'(0) = 0 \quad (18)$$

The result follows immediately. ■

Notice that we have not had to assume anything about the type process in stating this proposition, which applies for all such processes consistent with the validity of the first-order approach. Graphically, the idea is that if the optimum involves only downwards incentive compatibility constraints binding then it must always be possible to move the allocation of the top agent at time  $t$  (for a given history) along that agent's within-period indifference curve, without jeopardising the incentives for any agent to report truthfully. This movement is additionally reversible, and (under the assumption of utilitarianism) will preserve the value of the policymaker's objective. Hence if the original allocation is optimal it must not raise surplus resources: the marginal cost of incentivising a top agent to produce an extra unit of output must exactly equal that extra unit.

The result is an interesting one in its own right, since Kocherlakota (2011) has provided a computed example in which the optimal consumption-output distortion for 'top' agents appears to be non-zero, conditional upon a particular past report.<sup>19</sup> Specifically, the author obtains a non-zero 'top' rate in the second period of a two-period (overlapping generations) model for agents whose type was not the highest in the first period. The reason for this derives from the particular productivity process that he assumes. In the first period of his model, young agents may be either type  $\theta_H$  (high type) or  $\theta_L$  (low type). In the second period, those who were low types in the previous period may now be either type  $\theta_L\theta'_L$  or type  $\theta_L\theta'_H$ , and those who were high types may be either type  $\theta_H\theta'_L$  or type  $\theta_H\theta'_H$ . This implies that the highest type that an initially low-type agent could possibly be in the second period,  $\theta_L\theta'_H$ , is *not* the highest type across all agents in the economy, which is instead  $\theta_H\theta'_H$ . This in turn means that there are conceivably agents who could mimic the second-period agent of type  $\theta_L\theta'_H$

<sup>19</sup>See Chapter 6 of Kocherlakota (2011).

with a productivity level in excess of  $\theta_L \theta'_H$ , as well as implying that two agents who receive the ‘same’ (stochastic component to their) productivity draw in the second period,  $\theta'_H$ , do not have the same within-period preference structure across consumption-output space.

By contrast, in the model used in this paper the highest within-period type that an agent could *possibly* be is independent of history, and any two agents who receive the same within-period productivity draw and have reported the same history will make identical choices. Kocherlakota’s results are influenced by the fact that changes to the second-period allocations of agents of type  $\theta_L \theta'_H$  affect the incentives for first-period truthful reporting for agents of initial type  $\theta_H$  (a point noted by the author). If we were to map from his setting to ours, the appropriate specification of  $\Theta$  would be a time-varying set.<sup>20</sup>  $\Theta = \{\theta_L, \theta_H\}$  in the first period, and  $\Theta = \{\theta_L \theta'_L, \theta_L \theta'_H, \theta_H \theta'_L, \theta_H \theta'_H\}$  in the second. So it is only agents of type  $\theta_H \theta'_H$  that we are claiming in this paper should see zero marginal rates in the second period, since  $\theta_H \theta'_H$  is, in the relevant sense to us, the maximal element in  $\Theta$  in the second period. This result (together with zero marginal rates for those of type  $\theta_H$ ) is indeed reported by Kocherlakota.

## 5.4 Uniform utility perturbations

As already noted, the most common application of perturbation analysis in the dynamic optimal tax literature to date has been in deriving the ‘inverse Euler condition’ in models with additive separability in utility between consumption and leisure/labour supply. Our next proposition will provide the natural generalisation of this analysis to the non-separable case. Before presenting it, it is useful to define the function  $\alpha(c, y; \theta)$ , with  $\alpha : \mathbb{R}_+^2 \times \Theta \rightarrow \mathbb{R}$ , as follows:

$$\alpha(c, y; \theta) = \frac{u_c(c, y; \theta) - u_c(c, y; \theta')}{u_y(c, y; \theta') - u_y(c, y; \theta)} \quad (19)$$

provided  $\theta \neq \max \{\tilde{\theta} \in \Theta\}$ , where  $\theta' = \min \{\tilde{\theta} \in \Theta : \tilde{\theta} > \theta\}$ . If  $\theta = \max \{\tilde{\theta} \in \Theta\}$  we simply define  $\alpha(c, y; \theta) = 0$ .<sup>21</sup>

This  $\alpha$  function is very useful in understanding the perturbation constructions that follow, and merits a short discussion. It gives the marginal increase in output (away from the level  $y$ ) that must accompany a unit marginal increase in consumption (away from the level  $c$ ) *if the combined marginal perturbation is to have an equal impact on utility for the agents of both types,  $\theta$  and  $\theta'$  at the given allocation.*<sup>22</sup> More specifically for our purposes, it shows how to provide

<sup>20</sup>Nothing in the analysis above has precluded the possibility of a time-varying  $\Theta$ . All we need is that the elements of this set should be treated as independent of the history of productivity reports – even if this implies distributions across  $\Theta$  for some or all agents that put zero weight on some of its elements.

<sup>21</sup>Recall again that our focus at present is on the case in which  $\Theta$  contains a finite number of elements.

<sup>22</sup>The impact of such a perturbation on the utility of type  $\theta$  will be  $u_c(c, y; \theta) + \alpha(c, y; \theta) u_y(c, y; \theta)$ , and will be  $u_c(c, y; \theta') + \alpha(c, y; \theta') u_y(c, y; \theta')$  for type  $\theta'$ . It is easy

utility at the margin along a dimension in consumption-output space that will ensure both truth-tellers ( $\theta$ -types) and would-be mimickers ( $\theta'$ -types) receive the *same* utility increment.

If consumption is additively separable in utility then  $\alpha = 0$  always holds. This is just a re-statement of the known result, used in deriving the standard inverse Euler condition, that the marginal effect of consumption changes on utility is completely independent of type under separability. In the general, non-separable problem it is not possible to find composite perturbations that have the same marginal effect on utility for *all* types in this way. But if it is sufficient to study the relaxed problem then the effects of perturbations really only matter to the extent that they change utility levels for *two* particular agents in each case: those truthfully reporting the given type, and would-be mimickers whose type is one higher. Moreover, it is always possible to ensure common utility changes for these two agents alone, even in the event of non-separability – and it proves useful to do so.

When consumption and labour supply are Edgeworth complements, so that higher levels of the latter increase the marginal utility of the former and vice-versa, we will have  $\alpha > 0$ .<sup>23</sup> That is, higher production must accompany higher consumption if the marginal increase in utility is to be the same for both truth-tellers and mimickers. This is because under complementarity the (truth-telling) lower-type agents will receive a greater marginal benefit from a unit increase in consumption at any given allocation than the (mimicking) higher-type agents – because of the higher number of hours the lower types are working to produce the given output level. To offset this disparity, one must exploit the higher marginal *dis*utility of additional output for lower types, by requiring that greater production should accompany increased consumption. Conversely, when consumption and labour supply are Edgeworth substitutes we must have  $\alpha < 0$ .

We may now state the following Proposition, which generalises the inverse Euler condition to non-separable preferences. The proof is slightly involved, but we choose to keep it in the main text because the methods used are novel and will be applied repeatedly throughout much of the subsequent analysis.

**Proposition 4** *Suppose the solution to the relaxed problem also solves the general problem. Then for all time periods  $t \geq 1$  and for all reporting histories  $\hat{\theta}^t$ , the allocations  $(c_t^*(\hat{\theta}^t), y_t^*(\hat{\theta}^t))$  and  $X_{t+1}^*(\hat{\theta}^t)$  satisfy the following condition:*

$$\begin{aligned} & R_{t+1} \beta \frac{1 - \alpha(c_t^*, y_t^*; \theta_t)}{u_c(c_t^*, y_t^*; \theta_t) + u_y(c_t^*, y_t^*; \theta_t) \alpha(c_t^*, y_t^*; \theta_t)} \\ &= \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1} | \theta^t) \frac{1 - \alpha(c_{t+1}^*, y_{t+1}^*; \theta_{t+1})}{u_c(c_{t+1}^*, y_{t+1}^*; \theta_{t+1}) + u_y(c_{t+1}^*, y_{t+1}^*; \theta_{t+1}) \alpha(c_{t+1}^*, y_{t+1}^*; \theta_{t+1})} \end{aligned} \quad (20)$$

---

to confirm that the two are equal.

<sup>23</sup>Formally, we take consumption and labour supply to be Edgeworth complements if and only if  $u_{cy} > 0$ , and Edgeworth substitutes if and only if  $u_{cy} < 0$ . Since these cross-partials hold  $\theta$  fixed, higher output is equivalent to higher labour supply. Note that equation (3) then further implies  $u_{c\theta} < 0$  for Edgeworth complements and  $u_{c\theta} > 0$  for Edgeworth substitutes.

$c_{t+1}^*$  and  $y_{t+1}^*$  are given by the relevant entries in  $X_{t+1}^* (\hat{\theta}^t)$ .<sup>24</sup>

**Proof.** We consider now a perturbation to outcomes in both time  $t$  and time  $t+1$ . Specifically, we wish to choose  $\Delta(\delta)$  and  $\Delta_{-1}(\delta)$  functions so that the agent with a truthful reporting history of  $\hat{\theta}^t$  will experience a reduction in within-period utility at time  $t$  of  $\beta\delta$  units, and an increase in within-period utility at time  $t+1$  of  $\delta$  units *for any realisation of the  $t+1$  productivity parameter*. These changes will, together, keep constant the expected utility associated with a truthful reporting strategy from the perspective of any time period up to the  $t$ th. The difficulty lies in constructing the perturbations in a way that will preserve incentive compatibility. Again, we exploit the Proposition's supposition that no allocation that satisfies the constraint set of the relaxed problem can improve upon the solution to the general problem. This implies we need only concern ourselves with continuing to satisfy the  $N-1$  constraints at  $t+1$  that prevent mimicking by types 'one higher' than any given  $\theta_{t+1} \in \Theta$ , and the similar  $t$ -dated constraint preventing mimicking of type  $\theta_t$  (assumed to be the last entry in  $\hat{\theta}^t$ ) by the immediately superior type.

Indexing the elements of  $\Theta$  in ascending order  $\{1, \dots, N\}$ , our strategy is to construct perturbations in both time periods that change the consumption and output levels of the agent reporting  $\hat{\theta}^n$  in just such a way that the impact on within-period utility will be identical whether that agent is of true type  $\theta^n$  or  $\theta^{n+1}$ . To this end, let  $\Delta_{-1}(\delta)$  be given by:

$$\Delta_{-1}(\delta) = (\phi^c(\theta_t, -\beta\delta; c_t^*, y_t^*), \phi^y(\theta_t, -\beta\delta; c_t^*, y_t^*)) \quad (21)$$

where  $\phi^c(\theta, k; c^*, y^*)$  and  $\phi^y(\theta, k; c^*, y^*)$  are defined implicitly when  $\theta \neq \max\{\theta'' \in \Theta\}$  by the pair of equalities:

$$u(c^* + \phi^c(\theta, k; c^*, y^*), y^* + \phi^y(\theta, k; c^*, y^*); \theta) = u(c^*, y^*; \theta) + k \quad (22)$$

$$u(c^* + \phi^c(\theta, k; c^*, y^*), y^* + \phi^y(\theta, k; c^*, y^*); \theta') = u(c^*, y^*; \theta') + k \quad (23)$$

for  $\theta' = \min\{\theta'' \in \Theta : \theta'' > \theta\}$ , and when  $\theta = \max\{\theta'' \in \Theta\}$  by

$$u(c^* + \phi^c(\theta, k; c^*, y^*), y^*; \theta) = u(c^*, y^*; \theta) + k \quad (24)$$

$$\phi^y(\theta, k; c^*, y^*) = 0 \quad (25)$$

That is to say,  $\phi^c(\theta, k; c^*, y^*)$  and  $\phi^y(\theta, k; c^*, y^*)$  are the consumption and output increments required to increase the utility of both mimickers and truth-tellers by  $k$  units. These functions will be uniquely defined, which follows from the fact that the single crossing property holds. Similarly, the  $n$ th row of  $\Delta(\delta)$  is given by:

$$[\phi^c(\theta_{t+1}^n, \delta; c_{t+1}^*, y_{t+1}^*), \phi^y(\theta_{t+1}^n, \delta; c_{t+1}^*, y_{t+1}^*)] \quad (26)$$

where we index by type in the natural way. By construction this perturbation

---

<sup>24</sup>We suppress dependence upon  $\hat{\theta}^t$  to keep the notation manageable.



must preserve incentive compatibility in the relaxed problem at  $t + 1$ , since the within-period utility that any agent can gain from mimicking is being changed by exactly the same amount ( $\delta$ ) as the within-period utility from truth-telling (for the mimicking strategies that need concern us). It must also preserve incentive compatibility at  $t$  under the relaxed problem, since its aggregate impact on the present value of expected utility from the perspective of period  $t$  and earlier is equal to zero (a reduction by  $\beta\delta$  units at  $t$  and an increase by  $\delta$  units at  $t + 1$ , discounted at rate  $\beta$ ), both for agents of true type  $\theta_t$  and for the potential mimickers whose type is one higher. (Note that this assertion does not require any iid assumption, since utility is increased *uniformly* at the margin across all types at  $t + 1$ .) The overall impact of the perturbation on the present value (assessed at time  $t$ ) of the resources used by the policymaker is given by the following expression:

$$\begin{aligned} & \pi_{\Theta}(\theta^t) [\phi^c(\theta_t, -\beta\delta; c_t^*, y_t^*) - \phi^y(\theta_t, -\beta\delta; c_t^*, y_t^*)] \\ & + R_{t+1}^{-1} \pi_{\Theta}(\theta^t) \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1} | \theta^t) [\phi^c(\theta_{t+1}^n, \delta; c_{t+1}^*, y_{t+1}^*) \\ & - \phi^y(\theta_{t+1}^n, \delta; c_{t+1}^*, y_{t+1}^*)] \end{aligned}$$

We require for optimality that the derivative of this expression with respect to  $\delta$  should equal zero when  $\delta = 0$ ; otherwise the policymaker could use fewer resources in obtaining the same value for aggregate utility, and still satisfy the relaxed problem's constraint set. Taking the derivative gives the optimality condition:

$$\begin{aligned} & \beta [\phi_2^c(\theta_t, 0; c_t^*, y_t^*) - \phi_2^y(\theta_t, 0; c_t^*, y_t^*)] \\ = & R_{t+1}^{-1} \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1} | \theta^t) [\phi_2^c(\theta_{t+1}^n, 0; c_{t+1}^*, y_{t+1}^*) \\ & - \phi_2^y(\theta_{t+1}^n, \delta; c_{t+1}^*, y_{t+1}^*)] \end{aligned} \tag{27}$$

where  $\phi_2^c$  denotes the derivative of  $\phi^c$  with respect to its second argument. By total differentiation of conditions (22) to (25) with respect to  $k$  it is easy to show:

$$\begin{aligned} & \phi_2^c(\theta, 0; c^*, y^*) - \phi_2^y(\theta, 0; c^*, y^*) \\ = & \frac{1 - \alpha(c^*, y^*; \theta)}{u_c(c^*, y^*; \theta) + u_y(c^*, y^*; \theta) \alpha(c^*, y^*; \theta)} \end{aligned} \tag{28}$$

The result follows. ■

It is useful to provide a heuristic explanation of this result. The important innovation is to provide a general expression for the marginal cost of incentive-compatible utility provision from the perspective of the policymaker, and to show the manner in which it is optimal to spread this cost through time. Changing consumption and output jointly at  $t$  for the agent with report history  $\theta_t$  according to the vector  $(1, \alpha(c_t^*, y_t^*; \theta_t))$  increases the within-period utility of

that agent at the margin by  $u_c(c_t^*, y_t^*; \theta_t) + u_y(c_t^*, y_t^*; \theta_t) \alpha(c_t^*, y_t^*; \theta_t)$  units. By construction, it would have the same impact on a mimicking agent with a common report history to  $t - 1$ , but a type one higher at  $t$ . The  $t$ -dated cost of providing utility in this manner at the margin for each agent with the given report history is  $1 - \alpha(c_t^*, y_t^*; \theta_t)$  (any extra output being a negative cost). Hence the term on the left-hand side of (20) is the marginal cost for every  $\beta$  units of  $t$ -dated utility provided, which is converted into  $t + 1$  resources at the prevailing real interest rate. The term on the right-hand side is, by similar reasoning, the marginal cost (assessed at  $t + 1$ ) of providing the agent with report history  $\hat{\theta}^t$  with a guaranteed utility increment of one unit across types at time  $t + 1$  (and hence a discounted  $\beta$  units guaranteed from the perspective of  $t$ ). Again, these marginal costs are obtained under the assumption that increments to a given  $t + 1$  type's utility must provide equivalent increases to the utility of mimicking agents.

Why do these marginal perturbations preserve incentive compatibility?<sup>25</sup> Consider period  $t + 1$  first: we know that for any given agent the important consideration is whether the benefits to mimicking a type one lower have changed relative to the benefits from truthful reporting. This cannot be the case, since all truth-telling agents *and those that could be tempted to mimic them* see a common marginal utility increment of one unit in that time period. This is the importance of assuming that output changes in accordance with the  $\alpha$  function alongside any changes to consumption.

Meanwhile at time  $t$  the agent whose current type is indeed  $\theta_t$  would see exactly offsetting changes to the present value of truth-telling were current utility to be increased (reduced) by an amount  $\beta$  at the margin and future utility reduced (increased) across all future types by a unit at the margin. With no changes to the allocations to agents with alternative  $t$ -dated reports, this agent would have no reason not to continue reporting truthfully. But again, by construction we have ensured that the *same* ( $\beta$ -unit) marginal change to  $t$ -dated utility is engineered for the relevant *mimicker*. And since a unit of utility is gained for all types at  $t + 1$  at the margin, this mimicker will likewise witness no change in the benefits to mimicking type  $\theta_t$ . Hence incentive compatibility is preserved for any perturbation that increases (reduces) utility by  $\beta$  units at  $t$  and reduces (increases) it by one unit at  $t + 1$  for agents with the given reporting history. Since this perturbation is additionally having no impact on utility for any agent from the perspective of period  $t$  and earlier, it must also be having no impact on the policymaker's objective function – so a necessary condition for optimality is that it cannot generate a surplus in net present value. This is what condition (20) is expressing.

Note that, like the ‘no distortion at the top’ condition, this result applies for general type processes – so long as the first-order approach remains valid.

---

<sup>25</sup>Strictly they are only sure to preserve incentive compatibility in the relaxed problem. We exploit the fact that if the solution to the relaxed problem solves the general problem then even if our perturbations violate some of the general problem's constraints, they cannot possibly improve upon the general problem's optimum so long as they remain admissible under the relaxed problem's constraint set.

#### 5.4.1 Implications for optimal savings wedges

Quite aside from its theoretical implications, expression (20) is of interest in its own right. On a simple analytical level, it helps fill a widely-recognised gap in the existing theory. Golosov, Tsyvinski and Werning (2006) have written that “Little is known about the solution of the optimal problem when preferences are not separable [between consumption and leisure],” before making use of numerical simulations to show that some results (notably that savings ‘wedges’ should be positive) need not carry across from the separable to the non-separable case. Similarly, Kocherlakota (2011) has noted that “It would definitely be desirable to be able to construct optimal tax systems in dynamic settings in which preferences are nonseparable between consumption and labor inputs.” This result, it is hoped, will allow this to be achieved much more easily.

More importantly from an economic perspective, we are able to give qualified analytical support to the numerical result of Golosov, Tsyvinski and Werning that the optimal savings wedge *could* be negative for some agents, at least in the weak sense that under some preference structures we cannot analytically rule *out* the inequality:

$$u_c(c_t^*, y_t^*; \theta_t) > \beta R_{t+1} \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1} | \theta^t) u_c(c_{t+1}^*, y_{t+1}^*; \theta_{t+1}) \quad (29)$$

holding in certain time periods for certain realisations of  $\theta^\infty$ . This would suggest tax instruments are being used to hold consumption at  $t$  *below* the level that would obtain in the event that the consumer could save freely at the gross real interest rate  $R_{t+1}$ , given the distribution of consumption across states in  $t+1$ ; this can be interpreted as a subsidisation of savings. To understand why this is so, it is worth recalling exactly why the optimal savings wedge is positive under separability.

Taking the mathematical treatment first, by Jensen’s inequality we know that  $\sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1} | \theta^t) \left[ \frac{1}{u_c(c_{t+1}^*, y_{t+1}^*; \theta_{t+1})} \right] \geq \left[ \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1} | \theta^t) u_c(c_{t+1}^*, y_{t+1}^*; \theta_{t+1}) \right]^{-1}$ , with a strict inequality holding so long as the marginal utility of consumption varies in  $\theta_{t+1}$  (which will be true in all models of interest). From this a simple rearrangement of (20) in the case of separable preferences ( $\alpha(c, y; \theta) = 0$ ) confirms that savings are indeed deterred.<sup>26</sup>

$$u_c(c_t^*, y_t^*; \theta_t) \leq \beta R_{t+1} \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1} | \theta^t) u_c(c_{t+1}^*, y_{t+1}^*; \theta_{t+1}) \quad (30)$$

(again, with a strict inequality holding so long as the marginal utility of consumption varies in  $\theta_{t+1}$ ).

The *economic* reason why the usual Euler condition (with an equality in the above relationship) does not hold in this environment derives from the linked

---

<sup>26</sup>See, for instance, Golosov, Kocherlakota and Tsyvinski (2003) for a fuller treatment of the separable utility case.

problems of missing markets and over-insurance. Because each agent's productivity draw in each period is unobservable – and hence reports of it unverifiable – the idiosyncratic risk associated with future draws cannot be insured against. Absent any market intervention, the only way for individuals to prevent themselves from experiencing very low consumption in the event that they are unlucky in the future is to engage in private saving – their concerns dominated by future states of the world in which they are unlucky. This means all individuals in the economy are 'saving for a rainy day' simultaneously, even though it is (almost surely) impossible for them all to experience low productivity levels simultaneously. *Ex-post* there will be a sizeable measure of individuals who were *not* unlucky, and thus who have a large quantity of accumulated savings that they would not have chosen to hold, had they been able access complete insurance markets. This excess stock of savings reduces the marginal benefits to these individuals from working, since the consumption returns from doing so are not that valuable to them. Thus over time more productive agents are content to put in less and less effort – an outcome that is not constrained efficient.<sup>27</sup> The policymaker prefers to rein in savings at the margin, making it easier to provide future production incentives for higher types.

A more direct way to understand the same result is simply to consider why the consumption Euler equation is not a necessary optimality condition for the policymaker's problem. The Euler condition states that spreading resources through time, with equal consumption increments across states at  $t + 1$ , cannot raise utility. But this is not a choice open to our policymaker – who instead must ensure that spreading *utility* through time, with equal utility increments across states at  $t + 1$  (provided in a manner consistent with equal gains for truth-tellers and 'downwards' mimickers) cannot raise surplus *resources*. Providing equal consumption increments across states at  $t + 1$  would generally provide greater marginal utility to those whose initial consumption was lower, raising the benefits to higher types from mimicking. In the separable case the marginal cost of utility provision *in a manner that preserves incentive compatibility* is the inverse of the marginal utility of consumption: only when consumption is provided *differentially* across states at  $t + 1$  in proportions according to this marginal cost can incentive compatibility be preserved. In the more general case this marginal cost is the expression contained in equation (20), with utility changes effected through a combination of consumption and output perturbations.

Perhaps slightly disappointingly, it turns out that a simple re-statement of inequality (30) in the non-separable case is only possible in very particular circumstances. Fortunately there is a natural generalisation of the 'deterred savings' inequality that will apply more widely; but first we present the arguments that pertain to this standard consumption Euler condition.

**Proposition 5** *Suppose the solution to the relaxed problem also solves the general problem, and additionally that in all time periods  $s \geq 1$  and for all re-*

---

<sup>27</sup> There are clear parallels here with the general intuition provided by Greenwald and Stiglitz (1986) for missing markets and/or informational asymmetries implying a scope for Pareto-improving policy interventions (relative to market outcomes).

porting histories  $\widehat{\theta}^s$  the allocations  $(c_s^*(\widehat{\theta}^s), y_s^*(\widehat{\theta}^s))$  imply  $u_c(c_s^*, y_s^*; \theta_s) \geq u_y(c_s^*, y_s^*; \theta_s)$ . Then for all time periods  $t \geq 1$  and for all reporting histories  $\widehat{\theta}^t$ , the allocations  $(c_t^*(\widehat{\theta}^t), y_t^*(\widehat{\theta}^t))$  and  $X_{t+1}^*(\widehat{\theta}^t)$  will satisfy inequality (30) if one of the following conditions holds:

1. Preferences are additively separable between consumption and labour supply.
2. Consumption and labour supply are Edgeworth substitutes, and  $\theta_t = \max\{\theta \in \Theta\}$ .

**Proof.** The result follows from the arguments above in the event that condition 1 holds, and does not require the extra assumption that intratemporal wedges are weakly positive.

The definition of Edgeworth substitutes gives  $u_{cy} < 0$  under condition 2. By equation (3) this implies  $u_{c\theta} > 0$ . We also know  $u_{y\theta} > 0$ , so in general we will have  $\alpha(c, y; \theta) \leq 0$ , with a strict inequality except when  $\theta = \max\{\theta' \in \Theta\}$ . This, together with the assumption that intratemporal wedges are weakly positive, gives:

$$\frac{1 - \alpha(\theta)}{u_c(\theta) + u_y(\theta) \alpha(\theta)} \geq \frac{1}{u_c(\theta)} \quad (31)$$

(where we now suppress dependence upon  $c$  and  $y$  in the relevant functions to ease notation). Hence:

$$\begin{aligned} & \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1} | \theta^t) \frac{1 - \alpha(\theta_{t+1})}{u_c(\theta_{t+1}) + u_y(\theta_{t+1}) \alpha(\theta_{t+1})} \\ & \geq \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1} | \theta^t) \frac{1}{u_c(\theta_{t+1})} \\ & \geq \left[ \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1} | \theta^t) u_c(\theta_{t+1}) \right]^{-1} \end{aligned} \quad (32)$$

where the last result uses Jensen's inequality, and will hold strictly provided the marginal utility of consumption varies in  $\theta_{t+1}$ . If  $\theta_t = \max\{\theta \in \Theta\}$  then Proposition 3 implies  $u_c(c_t^*, y_t^*; \theta_t) = u_y(c_t^*, y_t^*; \theta_t)$ , so we have:

$$\begin{aligned} R_{t+1} \beta \frac{1}{u_c(\theta_t)} &= R_{t+1} \beta \frac{1 - \alpha(\theta_t)}{u_c(\theta_t) + u_y(\theta_t) \alpha(\theta_t)} \\ &= \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1} | \theta^t) \frac{1 - \alpha(\theta_{t+1})}{u_c(\theta_{t+1}) + u_y(\theta_{t+1}) \alpha(\theta_{t+1})} \\ &\geq \left[ \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1} | \theta^t) u_c(\theta_{t+1}) \right]^{-1} \end{aligned} \quad (33)$$

The result then follows from trivial manipulation. ■

We show subsequently that the assumption  $u_c(c_s^*, y_s^*; \theta_s) \geq u_y(c_s^*, y_s^*; \theta_s)$  is indeed satisfied at any optimum: it is an immediate corollary of Proposition 11 below.

Thus we have a result that when consumption and labour supply are substitutes there will always be a positive savings wedge imposed on the highest-type agent, in the sense implied by inequality (30). Beyond this, though, it is hard to say much of specific relevance to the *consumption* Euler condition. But this condition isn't the *only* way to characterise an optimal savings decision in an economy free from government intervention. For instance, optimality under autarky also requires that a consumer cannot produce an extra unit of output at time  $t$ , save it, produce  $R_{t+1}$  units fewer at  $t + 1$ , and increase the net present value of their utility by doing so. This consideration implies an alternative intertemporal optimality condition expressed in terms of an individual's output:

$$u_y(\theta_t) = \beta R_{t+1} \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1} | \theta^t) u_y(\theta_{t+1}) \quad (34)$$

More significantly for our purposes, *any combination* of a reduction in consumption and increase in output at  $t$ , coupled with any distribution (in each state of the world) of the saved proceeds at  $t + 1$  between extra consumption and reduced output is possible, and similarly must not increase utility at an optimum under autarky (for resource movements towards either period). In particular, in a world with no taxation the following optimality condition must hold:

$$\begin{aligned} & \frac{u_c(\theta_t) + u_y(\theta_t) \alpha(\theta_t)}{1 - \alpha(\theta_t)} \\ = & \beta R_{t+1} \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1} | \theta^t) \frac{u_c(\theta_{t+1}) + u_y(\theta_{t+1}) \alpha(\theta_{t+1})}{1 - \alpha(\theta_{t+1})} \end{aligned} \quad (35)$$

The numerator in the object  $\frac{u_c(\theta_t) + u_y(\theta_t) \alpha(\theta_t)}{1 - \alpha(\theta_t)}$  is the marginal effect on the agent's utility at the given allocation of a unit increase in consumption, coupled with an increase in output of  $\alpha(\theta_t)$  units. The denominator is the net cost to the agent of this change, under the maintained 'no tax' assumption that all of the  $\alpha(\theta_t)$  units of extra output are retained by the agent; the entire fraction then gives the marginal effect on utility per unit cost. The condition is just stating that no set of joint combinations of consumption and output changes can be used to spread resources through time and raise a surplus for the agent. So an agent's intertemporal ('savings') decisions are implicitly being distorted whenever equation (35) does not hold, with saving implicitly being discouraged whenever the left-hand side is less than the right.

Of course, any such dynamic distortion may well interact with concurrent distortions at the labour-consumption margin *within* a period, but there is nothing inherently correct about focusing exclusively on deviations from the traditional

consumption Euler equation in assessing the extent of savings distortions. Any equation that states that the marginal rate of substitution between a given pair of composite output-consumption bundles in two consecutive periods must equal their relative price ratio (in this case  $R_{t+1}$ ), as equation (35) does, is of equal validity to the consumption Euler equation in characterising optimal savings behaviour under autarky.

The useful feature of equation (35) is that we can say something far more general about deviations from *this* expression at the optimum than we can about deviations from an Euler equation stated in terms of consumption alone. Specifically, we have the following.

**Proposition 6** *Suppose the solution to the relaxed problem also solves the general problem. Then for all time periods  $t \geq 1$  and for all reporting histories  $\hat{\theta}^t$ , if consumption and labour supply are either Edgeworth substitutes or additively separable in preferences then savings will be deterred at the optimum, in the sense that the allocations  $(c_t^*(\hat{\theta}^t), y_t^*(\hat{\theta}^t))$  and  $X_{t+1}^*(\hat{\theta}^t)$  will satisfy the following condition:*

$$\begin{aligned} & \frac{u_c(\theta_t) + u_y(\theta_t)\alpha(\theta_t)}{1 - \alpha(\theta_t)} \\ & \leq \beta R_{t+1} \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1}|\theta^t) \frac{u_c(\theta_{t+1}) + u_y(\theta_{t+1})\alpha(\theta_{t+1})}{1 - \alpha(\theta_{t+1})} \end{aligned} \quad (36)$$

with the inequality holding strictly so long as the object  $\frac{u_c(\theta_{t+1}) + u_y(\theta_{t+1})\alpha(\theta_{t+1})}{1 - \alpha(\theta_{t+1})}$  varies for different draws of  $\theta_{t+1} \in \Theta$ .

**Proof.** If consumption and labour supply are substitutes then  $\alpha(\theta_t) < 0$ , so for the preferences we are focusing on we must always have:

$$\frac{u_c(\theta_t) + u_y(\theta_t)\alpha(\theta_t)}{1 - \alpha(\theta_t)} > 0 \quad (37)$$

(recalling that  $u_y < 0$ ). Thus by Jensen's inequality we have the following:

$$\begin{aligned} & \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1}|\theta^t) \left[ \frac{u_c(\theta_{t+1}) + u_y(\theta_{t+1})\alpha(\theta_{t+1})}{1 - \alpha(\theta_{t+1})} \right]^{-1} \\ & \geq \left[ \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1}|\theta^t) \frac{u_c(\theta_{t+1}) + u_y(\theta_{t+1})\alpha(\theta_{t+1})}{1 - \alpha(\theta_{t+1})} \right]^{-1} \end{aligned} \quad (38)$$

with a strict inequality provided  $\frac{u_c(\theta_{t+1}) + u_y(\theta_{t+1})\alpha(\theta_{t+1})}{1 - \alpha(\theta_{t+1})}$  varies for different draws of  $\theta_{t+1}$ . The left-hand side of (38) is also the right-hand side of equation (20); the inequality in the Proposition then follows straightforwardly from using (20) in (38). ■

Note that this result has been stated irrespective of the manner and extent to which income is being taxed *within* periods  $t$  and  $t + 1$ : unlike the prior Proposition we do not need any assumption that tax wedges are weakly positive for savings to be deterred in the given sense.

More significantly, note that we are *not* able to state the result for the case of Edgeworth complements: in that case we cannot rule out the possibility that  $\frac{u_c(\theta_{t+1}) + u_y(\theta_{t+1})\alpha(\theta_{t+1})}{1 - \alpha(\theta_{t+1})} < 0$  may hold at the optimum for some values of  $\theta_{t+1}$ , preventing us from applying Jensen's inequality.<sup>28</sup> As it happens, Proposition 18 below implies  $\frac{u_c(\theta_{t+1}) + u_y(\theta_{t+1})\alpha(\theta_{t+1})}{1 - \alpha(\theta_{t+1})} > 0$  will also hold at the optimum under complements provided types additionally follow an iid process for all agents; but this relies on arguments that we have yet to establish.

In economic terms, the result suggests the problem of over-saving in the absence of perfect insurance markets carries over directly to the case of substitutes, and it will also apply under some circumstances with complements – so long as the object  $\frac{u_c(\theta_{t+1}) + u_y(\theta_{t+1})\alpha(\theta_{t+1})}{1 - \alpha(\theta_{t+1})}$  is known to be positive. But we cannot be confident that savings will be being deterred at the margin *if the marginal cost of incentive-compatible utility provision could turn negative*. Though that possibility may at first appear unlikely, we show in a computed example below that it can indeed obtain under Markov shock processes and complementarity. The problem in this event is that, starting from the equilibrium allocation, an undistorted decision by agents to ‘save’ at the margin (as we have defined it) involves increasing their  $t + 1$  utility across all states through a uniform change in the quantity of resources at their disposal in that period, with these resources distributed between consumption and output in proportions corresponding to the relevant  $\alpha(\theta_{t+1})$  terms. But it is possible that utility may be increased on average across  $t + 1$  states in this manner *even when the uniform quantity of resources to be allocated is negative*. This could happen whenever the optimum allows some agents to increase their utility despite increasing their production at the margin by more than their consumption – because of high equilibrium distortions at the labour supply-consumption margin. In this case extra ‘savings’ – in the sense of incremental *utility* deferral – in fact correspond to a lower stock of *resources* being deferred. It would not be surprising if in this case the standard intuition relating to over-insurance did not apply.

This argument does highlight clearly the importance of distinguishing between marginal and average distortions. All we have said is that at the optimum under complementarity individuals could conceivably defer utility through constant marginal resource changes across future states, and pay a negative cost for doing so. But plainly this would never be a feature of the equilibrium allocation under autarky, which must always involve  $u_c + u_y = 0$  for all agents in all periods – making it impossible for utility to increase at the margin along a

---

<sup>28</sup>In fact subsequent results (specifically Proposition 18) do establish that  $\frac{u_c(\theta_{t+1}) + u_y(\theta_{t+1})\alpha(\theta_{t+1})}{1 - \alpha(\theta_{t+1})} > 0$  holds at the optimum for complements in the case of iid shocks, but this does not generalise to Markov transition processes. But this bound cannot be stated until we have presented our general perturbation methodology more fully.



vector that sees output rise by more than consumption. Knowing whether the *total* quantity of resources saved at the utilitarian optimum is less than the *total* quantity that would be saved under an autarkic equilibrium is as important as knowing whether the optimum is characterised by additional marginal savings being discouraged – but it is only the latter that we have been able to shed light on.

Moving away from its implications for marginal tax wedges, it will also be interesting to consider what Proposition 4 implies for the ‘immiseration’-type results that emerge in the special case that  $R_t = \beta^{-1}$  for all  $t$ . In that case equation (20) is a martingale, to which martingale convergence results may be applicable if bounds can be placed upon it. Under separable preferences the martingale is in the inverse of the marginal utility of consumption, which is bounded below at zero under usual Inada conditions. It is well known (see, for instance, Farhi and Werning (2007)) that this implies almost all agents will see their marginal utility of consumption converge to the lower bound along an optimal path – and thus that consumption tends to zero for almost all agents. This ‘immiseration’ was first demonstrated as a potential property of optimal allocations under asymmetric information in a moral hazard context by Thomas and Worrall (1990), and it will turn out to generalise fairly robustly to the non-separable case – with important qualifications. But unfortunately the proofs rely on other arguments that are still to be established, so we defer treatment of this important area until later in the paper.

## 5.5 Intermediate perturbations: the case of iid types

### 5.5.1 Heuristic overview

We have shown above how it is possible to choose two particular pairs of  $\Delta(\delta)$  and  $\Delta_{-1}(\delta)$  schedules, in each case satisfying the three requirements of local incentive compatibility preservation, continuity, and zero impact on *ex ante* expected utility. The first was obtained by arguing that movements in either direction along the within-period indifference curve of the highest-type agent are always incentive-compatible (under the relaxed problem) and feasible. These perturbations will have zero impact on the within-period utility of all agents in the period that the  $\Delta(\delta)$  schedule is applied (and in all other periods). The second was obtained by arguing we could reduce (increase) the utility of an agent with a given history to time  $t$  by  $\beta\delta$  units according to  $\Delta_{-1}(\delta)$ , and raise (lower) utility by  $\delta$  units for all realisations of the productivity parameter at  $t + 1$ , in both cases in a manner that is incentive compatible under the relaxed problem and feasible. These perturbations will raise within-period utility for all contemporary realisations of  $\theta$  by  $\delta$  units in the period that the relevant  $\Delta(\delta)$  schedule is applied. In both cases the results were quite general, in that they were derived without making any specific assumptions about the distribution of agent types through time.

This sub-section shows that it is also possible to find an entire set of perturbations ‘intermediate’ between these two extremes, in the sense that these

can raise the within-period utility of the top  $n$  types by  $\delta$  units in the period that the relevant  $\Delta(\delta)$  schedule is applied, whilst holding constant the utility of all other types (for all  $n \in \{1, \dots, N-1\}$ ). For the arguments to be valid we must make the additional simplifying assumption that type processes are iid for all agents. This is clearly unrealistic, but fortunately the logic generalises the Markov case with relatively minor additional complications.

The intuition that we exploit is the following. Suppose one were to perturb the within-period allocation at time  $t+1$  of an agent with prior reporting history  $\hat{\theta}^t$  and whose current report is  $\hat{\theta}_{t+1}^n$ ,<sup>29</sup> in a manner that keeps the utility of an agent truthfully reporting this type constant. This can be achieved by a movement along this agent's within-period indifference curve in consumption-output space. If the solution to the relaxed problem also solves the general problem then we can focus on changes that preserve incentive compatibility under the relaxed problem, knowing that they cannot improve on the best outcome when the general constraint set is imposed. To remain within the relaxed problem's constraint set we must simultaneously change the utility at  $t+1$  of all *higher-type* agents (who share the same prior reporting history,  $\hat{\theta}^t$ ) by an amount equal to the change in the utility that the  $n+1$ th agent can obtain by mimicking the  $n$ th. These latter utility changes must, in turn, be delivered along a dimension in consumption-output space consistent with the same impact being felt by truth-tellers and (relevant) mimickers (in the case of separability, for instance, via consumption increments alone).

To preserve prior incentive compatibility there must also be a perturbation in period  $t$  to the utility of an agent with reporting history  $\hat{\theta}^t$ , so as to keep the present value of reporting  $\hat{\theta}^t$  constant in that period (and earlier) – with this perturbation again being constructed to ensure equal utility increments for truth-tellers and mimickers. The iid assumption will ensure that both truth-tellers and mimickers at  $t$  experience the same *ex-ante* change to their expected within-period utility at  $t+1$  from reporting  $\hat{\theta}^t$ , even though the proposed perturbation has differential effects across types at  $t+1$ .<sup>30</sup>

Figure 3 illustrates diagrammatically the type of perturbation we have in mind for the latter period, under the assumption that there are two types in  $\Theta$ .

Suppose we know that the optimal allocation gives the lower-type agent a bundle at point A and the higher-type agent a bundle at point B in the relevant period.<sup>31</sup> Then suppose that the allocation to the lower type is perturbed along the lower type's indifference curve through A,  $I(\theta_1)$ . This will change the utility that the higher-type agent can obtain by mimicking, corresponding graphically to a movement by the higher type onto a distinct indifference curve from  $I'(\theta_2)$

<sup>29</sup>Superscripts again denote ranking in  $\Theta$ , with higher values for  $n$  corresponding to higher values of  $\theta$ .

<sup>30</sup>Under more general type distributions this will not be true, since the probability distribution across future states will differ between mimickers and truth-tellers at  $t$ .

<sup>31</sup>As drawn this implies the higher-type agent strictly envies the lower type within the given period, so we are implicitly assuming higher future utility is also provided by way of compensation to prevent mimicking.

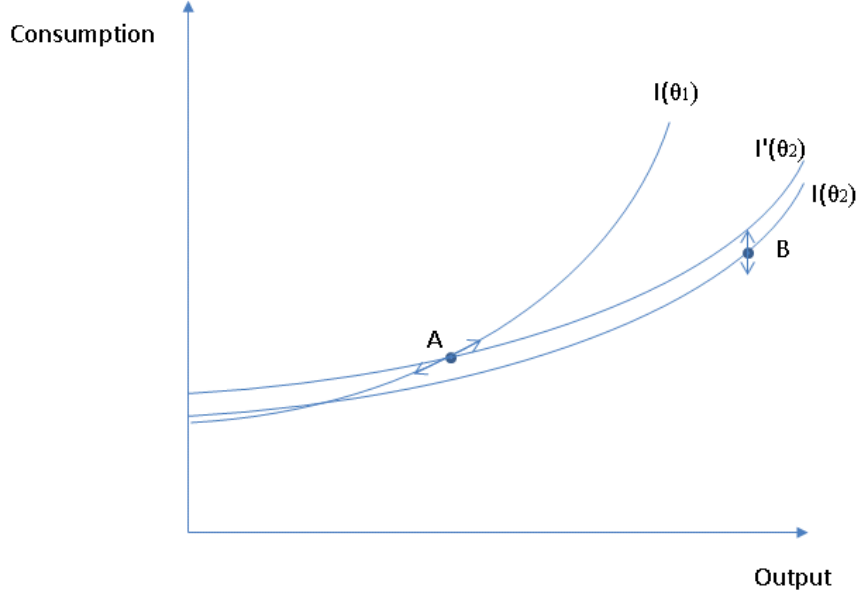


Figure 3: Preserving downwards incentive compatibility

at the lower type's (perturbed) allocation, as movement takes place along  $I(\theta_1)$ . To offset this, the higher type must simultaneously be given a utility change when reporting truthfully, corresponding to a movement in the locus of point B by exactly the amount necessary for the higher-type agent to see the *same* change in utility at point A as at point B.

When more than two agents are present one must additionally provide still higher types with compensation for the given changes – but note that the within-period incentives of lower types will be unaffected. Any agent whose allocation is perturbed along his or her within-period indifference curve is, by construction, left no worse off from truthful reporting, and under the relaxed problem's constraint set this agent is not concerned by changes to allocations higher up in the type distribution. Still lower types are likewise unaffected. Provided we can additionally ensure incentives in prior periods are unaffected, this class of perturbation can be used to obtain additional optimality conditions.

### 5.5.2 Analytical treatment

We now present the arguments more formally. For each potential choice of  $n$  and any arbitrary report history  $\hat{\theta}^t$  for some  $t \geq 1$ , we wish to find a set of time- $t + 1$  perturbations, stacked in a function  $\Delta(\delta)$ , that will hold equilibrium consumption and output levels constant for all agents of type  $\theta_{t+1}^{n-1}$  and

lower, whilst changing them for type  $\theta_{t+1}^n$  and higher.<sup>32</sup> This implies that the marginal impact on any given  $\Delta(\delta)$  function of moving  $\delta$  away from zero must depend on the choice of  $n$ , which in turn implies that if we are successful we will have obtained a new set of  $N - 1$  linearly independent dimensions along which the optimal allocation can be perturbed at the margin whilst preserving incentive compatibility locally – in turn providing  $N - 1$  new, distinct optimality conditions.

Before presenting the formal argument it is useful to define the function  $\tau : \mathbb{R}_+^2 \times \Theta \rightarrow \mathbb{R}$  by  $\tau(c, y, \theta) \equiv 1 + \frac{u_y(c, y, \theta)}{u_c(c, y, \theta)}$ . This can be thought of as the implicit within-period marginal income tax rate faced by an agent of type  $\theta$  receiving an allocation  $(c, y)$  – which is seen by noting:

$$u_c(c, y, \theta) (1 - \tau(c, y, \theta)) = -u_y(c, y, \theta) \quad (39)$$

Or, in words, the marginal utility of consumption multiplied by the real disposable income that an agent receives per unit of extra output that they produce is equal to the marginal disutility of production. By defining  $\tau$  in this way we are not implying that a non-linear marginal income tax rate of this form should necessarily form part of any ultimate decentralisation scheme, but it is useful to deploy a variable with a such a clear practical interpretation.

Then we have the following Proposition, the proof of which is in the appendix:

**Proposition 7** *Suppose the solution to the relaxed problem also solves the general problem, and that type draws are iid across agents and time. Then for all time periods  $t \geq 1$ , all reporting histories  $\hat{\theta}^t$ , and all  $\theta_{t+1}^n \in \Theta : \theta_{t+1}^n \neq \max\{\theta' \in \Theta\}$  (so  $n < N$ ), the optimal allocations  $(c_t^*(\hat{\theta}^t), y_t^*(\hat{\theta}^t))$  and  $X_{t+1}^*(\hat{\theta}^t)$  satisfy the following condition:*

$$\begin{aligned} & -\pi_\Theta(\theta_{t+1}^n) \frac{\tau(\theta_{t+1}^n)}{u_c(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1}) (1 - \tau(\theta_{t+1}^n)) + u_y(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})} \quad (40) \\ & + \sum_{m=n+1}^N \pi_\Theta(\theta_{t+1}^m) \frac{1 - \alpha(\theta_{t+1}^m)}{u_c(\theta_{t+1}^m) + u_y(\theta_{t+1}^m) \alpha(\theta_{t+1}^m)} \\ & = \beta R_{t+1} \pi_\Theta(\theta_{t+1} \geq \theta_{t+1}^{n+1}) \frac{1 - \alpha(\theta_t)}{u_c(\theta_t) + u_y(\theta_t) \alpha(\theta_t)} \end{aligned}$$

where  $u(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})$  is used to denote the within-period utility an agent of type  $\theta_{t+1}^{n+1}$  can obtain by mimicking an agent of type  $\theta_{t+1}^n$ , and its derivatives are defined correspondingly.<sup>33</sup>

<sup>32</sup> Again, this applies only to the set of agents who have previously reported  $\hat{\theta}^t$ .

<sup>33</sup> We again suppress the dependence of  $\tau$ ,  $\alpha$ ,  $u_c$  and  $u_y$  on equilibrium allocations and report histories, in order to keep the notation manageable.

The result merits some discussion. The term on the right-hand side is the marginal cost to the policymaker of raising utility at time  $t$  (for agents with the relevant report history) by an amount exactly equal to the discounted impact on expected utility of raising  $t+1$  utility by a unit at the margin for all agents whose type is  $\theta_{t+1}^{n+1}$  and higher. The term on the left-hand side is the marginal cost of raising  $t+1$  utility in this fashion. The latter, in turn, has two components to it. The first is in fact a negative cost, so long as the effective marginal income tax rate is non-negative (and Proposition 11 below confirms that this is generally so). Any movement along the  $n$ th agent's indifference curve must raise resources so long as the slope of that indifference curve is less than one. The numerator  $\tau(\theta_{t+1}^n)$  in the first fraction gives the net marginal increase in revenue to the policymaker (at  $t+1$ ) for every unit increase in the consumption of agent  $\theta_{t+1}^n$ , as we move along that agent's indifference curve.<sup>34</sup> The denominator (which is always positive) gives the marginal impact that this change has on the utility of mimickers whose type is one higher. Hence the entire term gives the additional quantity of resources raised from agents of type  $\theta_{t+1}^n$  for every unit by which mimickers' utilities are increased. Notice that, normalised for scale, it must equal agent  $\theta_{t+1}^n$ 's marginal Hicksian (compensated) demand change for output minus that for consumption as the effective price of consumption in terms of output is increased at the equilibrium allocation.<sup>35</sup>

The second term on the left-hand side of (40) is more familiar from our earlier results: it gives the marginal cost of incentive-compatible utility provision to all agents above type  $\theta_{t+1}^n$  – utility provision that is necessary for the complete perturbation to preserve incentive compatibility. The implications of the Proposition for this term are in fact best seen by applying (20), so as instead to obtain the following condition for all  $n \in \{1, \dots, N\}$  and  $t \geq 1$ .<sup>36</sup>

$$\begin{aligned}
& - \frac{\pi_{\Theta}(\theta_{t+1}^n)}{\pi_{\Theta}(\theta_{t+1} \geq \theta_{t+1}^{n+1})} \frac{\tau(\theta_{t+1}^n)}{u_c(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1}) (1 - \tau(\theta_{t+1}^n)) + u_y(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})} \quad (41) \\
& + \sum_{m=n+1}^N \frac{\pi_{\Theta}(\theta_{t+1}^m)}{\pi_{\Theta}(\theta_{t+1} \geq \theta_{t+1}^{n+1})} \frac{1 - \alpha(\theta_{t+1}^m)}{u_c(\theta_{t+1}^m) + u_y(\theta_{t+1}^m) \alpha(\theta_{t+1}^m)} \\
& = \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1}) \frac{1 - \alpha(\theta_{t+1})}{u_c(\theta_{t+1}) + u_y(\theta_{t+1}) \alpha(\theta_{t+1})}
\end{aligned}$$

Expressed in this way, we have a relationship between the expected marginal cost of utility provision across *all* types at  $t+1$  (on the right-hand side), and the expected marginal cost conditional upon type being higher than  $\theta_{t+1}^n$  (the second term on the left-hand side). Provided taxes are positive for the agent

---

<sup>34</sup>To see this, recall that the slope of the relevant indifference curve is  $-\frac{u_y(\theta_{t+1}^n)}{u_c(\theta_{t+1}^n)} =$

$(1 - \tau(\theta_{t+1}^n))$

<sup>35</sup>This follows directly from the fact that we are considering movements along the indifference curve.

<sup>36</sup>The case of  $n = N$  reduces to the ‘no distortion at the top’ result.

of type  $\theta_{t+1}^n$ , the equation implies the expectation conditional upon being type  $\theta_{t+1}^{n+1}$  or higher will be greater than the unconditional expectation. This is just a re-statement in the non-separable case of the fact that more resources must be used to increase by a given amount the welfare of those whose utility is already relatively high.

When utility is separable between consumption and labour supply, the previous condition collapses to a much simpler object:

$$\begin{aligned} & \frac{\pi_{\Theta}(\theta_{t+1}^n)}{\pi_{\Theta}(\theta_{t+1} \geq \theta_{t+1}^{n+1})} \frac{\tau(\theta_{t+1}^n)}{u_y(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1}) - u_y(\theta_{t+1}^n)} \\ &= E \left[ \frac{1}{u_c(\theta_{t+1})} | \theta_{t+1} \geq \theta_{t+1}^{n+1} \right] - E \left[ \frac{1}{u_c(\theta_{t+1})} \right] \end{aligned} \quad (42)$$

So at the optimum, the implicit tax rate will be higher the higher is the difference between the expected marginal cost of utility provision to agents *above* type  $\theta_{t+1}^n$ , and its average across all agents. This makes sense: if one distortion from the first-best is high, in that high-type agents have relatively low marginal utilities of consumption, then other distortions are more likely to be beneficial. The effective tax rate will also be higher the greater is the productivity differential between truth-tellers of type  $\theta_{t+1}^n$  and their mimickers, since high levels of this productivity gap imply distortions to the allocation of the lower type are more effective at the margin in ‘pulling down’ the utility from mimicking – and hence more desirable. Finally, the relative measure of agents of a higher type than  $\theta_{t+1}^n$  will matter: the higher is this measure, the more desirable are higher effective taxes on  $\theta_{t+1}^n$ , since the resources gained by reducing the utility of higher types are then more sizeable by comparison with the resources lost from distorting the production of the lower type.

All of these arguments are essentially familiar from the analysis of *static* optimal income tax models (see, in particular, Roberts (2000) and Saez (2001)) – though our focus on the marginal cost of utility provision is an innovation on that literature, and allows for a much simpler presentation of the the necessary optimality condition.<sup>37</sup> Indeed, though the arguments have not yet fully established it, we will subsequently show that condition (41) is necessary for intratemporal optimality regardless of whether earlier timer periods have existed, and this will be done in a manner that does not rely at all on subsequent time periods existing – so that the condition must also apply at any optimum in a static income tax model (for which the first-order approach is valid). Though these models tend to assume a continuous type distribution, we show in a companion paper (Brendon (2011)) that the limit of (41) as types become arbitrarily close to one another is a necessary intratemporal optimality condition when the first-order approach is valid in that setting too (again under the assumption of iid types). For this reason the arguments in the current paper could well prove equally useful in clarifying the analysis and implementation of static optimal

---

<sup>37</sup>The simplification appears to arise from the fact our perturbations isolate Hicksian substitution effects by construction. I thank Kevin Roberts for highlighting this.

income tax problems as dynamic ones.

### 5.5.3 Characterising the set of perturbations

An interesting feature of condition (40) is that it nests our earlier two as extreme cases. Suppose we set  $n = 0$ , and extend the set  $\Theta$  to include some arbitrary element, denoted  $\theta^0$ , which is strictly less than all other elements of  $\Theta$  and whose probability in any period is always zero under the  $\pi_\Theta$  measure. Then  $\pi_\Theta(\theta_{t+1}^0) = 0$  and  $\pi_\Theta(\theta_{t+1} \geq \theta_{t+1}^{n+1}) = 1$  by definition, and the equation reduces to the generalised version of the inverse Euler equation, (20). If, on the other hand, we set  $n = N$  then the last two terms drop out, and it reduces to a requirement that  $\tau(\theta_{t+1}^N) = 0$  – the ‘no distortion at the top’ result from Proposition 3.

It proves useful for the analysis of the non-iid case to elaborate on this point further. Suppose  $\phi^c(\theta, k; c, y)$  and  $\phi^y(\theta, k; c, y)$  are defined for all  $\theta \in \Theta$  as in the proof of Proposition 4, and  $\varphi^c(\theta, k; c, y)$  and  $\varphi^y(\theta, k; c, y)$  for all  $\theta \neq \max\{\tilde{\theta} \in \Theta\}$  as in the proof of Proposition 7.<sup>38</sup> If we define  $\varphi^c(\theta, k; c, y)$  and  $\varphi^y(\theta, k; c, y)$  for  $\theta = \max\{\tilde{\theta} \in \Theta\}$  by:

$$\varphi^c(\theta, k; c, y) = k \quad (43)$$

$$u(c + \varphi^c(\theta, k; c, y), y + \varphi^y(\theta, k; c, y); \theta) = u(c, y; \theta) \quad (44)$$

and the derivatives with respect to  $k$  by  $\varphi_k^c$  and  $\varphi_k^y$  correspondingly, then the complete set of (non-marginal)  $\Delta(\delta)$  perturbations that we consider at  $t + 1$  when obtaining Proposition 7 is given by the  $N + 1$  matrices of the following form:

$$\begin{bmatrix} 0 & 0 \\ \dots & \dots \\ 0 & 0 \\ \dots & \dots \\ 0 & 0 \\ \varphi^c(\theta_{t+1}^N, \delta) & \varphi^y(\theta_{t+1}^N, \delta) \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ \dots & \dots \\ 0 & 0 \\ \dots & \dots \\ \varphi^c(\theta_{t+1}^{N-1}, \delta) & \varphi^y(\theta_{t+1}^{N-1}, \delta) \\ \phi^c(\theta_{t+1}^N, \delta) & \phi^y(\theta_{t+1}^N, \delta) \end{bmatrix}$$

<sup>38</sup> Recall that  $\phi^c(\theta, k; c, y)$  and  $\phi^y(\theta, k; c, y)$  gave, respectively, the consumption and output changes necessary to increase the within-period utility of an agent of type  $\theta$  by  $k$  units, starting from an allocation  $(c, y)$ . For  $\theta \neq \max\{\tilde{\theta} \in \Theta\}$ ,  $\varphi^c(\theta, k; c, y)$  and  $\varphi^y(\theta, k; c, y)$  gave the corresponding consumption and output changes needed to increase the utility of a mimicking agent (one type higher than  $\theta$ ) by a unit, whilst holding constant the utility of  $\theta$  types.

The derivatives of these functions with respect to their  $i$ th argument are denoted  $\phi_i^c$ ,  $\phi_i^y$  etc.

$$\begin{bmatrix} 0 & 0 \\ \dots & \dots \\ \varphi^c(\theta_{t+1}^n, \delta) & \varphi^y(\theta_{t+1}^n, \delta) \\ \phi^c(\theta_{t+1}^{n+1}, \delta) & \phi^y(\theta_{t+1}^{n+1}, \delta) \\ \dots & \dots \\ \phi^c(\theta_{t+1}^N, \delta) & \phi^y(\theta_{t+1}^N, \delta) \end{bmatrix}, \begin{bmatrix} \varphi^c(\theta_{t+1}^1, \delta) & \varphi^y(\theta_{t+1}^1, \delta) \\ \phi^c(\theta_{t+1}^2, \delta) & \phi^y(\theta_{t+1}^2, \delta) \\ \dots & \dots \\ \phi^c(\theta_{t+1}^{n+1}, \delta) & \phi^y(\theta_{t+1}^{n+1}, \delta) \\ \dots & \dots \\ \phi^c(\theta_{t+1}^N, \delta) & \phi^y(\theta_{t+1}^N, \delta) \end{bmatrix}$$

$$\begin{bmatrix} \phi^c(\theta_{t+1}^1, \delta) & \phi^y(\theta_{t+1}^1, \delta) \\ \phi^c(\theta_{t+1}^2, \delta) & \phi^y(\theta_{t+1}^2, \delta) \\ \dots & \dots \\ \phi^c(\theta_{t+1}^{n+1}, \delta) & \phi^y(\theta_{t+1}^{n+1}, \delta) \\ \dots & \dots \\ \phi^c(\theta_{t+1}^N, \delta) & \phi^y(\theta_{t+1}^N, \delta) \end{bmatrix}$$

to each of which corresponds a  $\Delta_{-1}(\delta)$  perturbation given by

$$[\phi^c(\theta_t, -\beta\delta\pi_\Theta(\theta_{t+1} \geq \theta_{t+1}^{n+1})), \phi^y(\theta_t, -\beta\delta\pi_\Theta(\theta_{t+1} \geq \theta_{t+1}^{n+1}))]$$

where  $n \in \{1, \dots, N+1\}$  identifies the lowest agent type at  $t+1$  whose utility is being increased (with  $N+1$  corresponding to movements along the ‘top’ indifference curve only). In what follows we index the entire matrix of perturbations by this  $n$ , so that the first of the matrices above gives  $\Delta^{N+1}(\delta)$ , with corresponding  $\Delta_{-1}^{N+1}(\delta)$ :

$$\Delta_{-1}^{N+1}(\delta) = [0, 0] \quad (45)$$

the second gives  $\Delta^N(\delta)$ , with corresponding  $\Delta_{-1}^N(\delta)$ :

$$\begin{aligned} & \Delta_{-1}^N(\delta) \\ &= [\phi^c(\theta_t, -\beta\delta\pi_\Theta(\theta_{t+1} \geq \theta_{t+1}^N)), \phi^y(\theta_t, -\beta\delta\pi_\Theta(\theta_{t+1} \geq \theta_{t+1}^N))] \end{aligned} \quad (46)$$

and so on. The last matrix, corresponding to uniform utility provision across agents, we denote  $\Delta^1(\delta)$ , with corresponding  $\Delta_{-1}^1(\delta)$ :

$$\Delta_{-1}^1(\delta) = [\phi^c(\theta_t, -\beta\delta), \phi^y(\theta_t, -\beta\delta)] \quad (47)$$

The *marginal* effects on allocations of moving  $\delta$  away from zero will, for each of these perturbations, be given by equivalent matrices in which  $\phi_2^c(\cdot, 0)$  and  $\phi_2^y(\cdot, 0)$  replace  $\phi^c(\cdot, \delta)$  and  $\phi^y(\cdot, \delta)$  respectively, and  $\varphi_2^c(\cdot, 0)$  and  $\varphi_2^y(\cdot, 0)$  replace  $\varphi^c(\cdot, \delta)$  and  $\varphi^y(\cdot, \delta)$ . It is then clear by inspection that the  $N+1$  marginal changes to allocations at  $t+1$ , which we may denote  $\Delta^{n'}(0)$  for  $n \in \{1, \dots, N+1\}$ , are linearly independent from one another. Thus each of the associated first-order conditions will be providing distinct information about the character of the optimal allocation.

The sense in which the perturbations analysed in Proposition 7 are ‘intermediate’ between those of Propositions 3 and 4 should now be apparent. A compact way to see matters is in terms of an  $(N+1) \times N$  matrix listing the



set of possible marginal impacts on the within-period utility levels of agents of different types that is afforded by the marginal perturbations  $\Delta^{n'}(0)$  for  $n \in \{1, \dots, N+1\}$ . We label this matrix  $\hat{J}$ :

$$\hat{J} = \begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ 0 & 1 & \dots & 1 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix} \quad (48)$$

Element  $\hat{J}_{n,m}$  is then the marginal impact on the within-period utility of an agent of type  $\theta_{t+1}^m$  caused by a marginal move in  $\delta$  away from zero in accordance with the  $\Delta^n(\delta)$  schedule. Note the fact that the non-zero entries are all equal to one arises from the way we have defined  $\delta$  in each case.

This matrix representation allows us to develop the analysis in a way that will prove very helpful when we move away from iid types, and we next present two technical Lemmas that are useful to this end. In stating them we denote by  $\pi_{\Theta}^{vec}$  the  $N$ -dimensional vector stacking the probabilities  $\pi_{\Theta}(\theta_t^n)$  in order from  $n = 1$  to  $n = N$  (these are independent of time under the maintained iid assumption). Then we have the following, the proof of which is in the appendix:

**Lemma 8** *Suppose that type draws are iid across agents and time. Then for any vector  $\nu \in \mathbb{R}^N$  (whose  $n$ th element is denoted  $\nu_n$ ), all time periods  $t \geq 1$  and any given reporting history  $\hat{\theta}^t$ , it is possible to perturb the optimal allocations  $(c_t^*(\hat{\theta}^t), y_t^*(\hat{\theta}^t))$  and  $X_{t+1}^*(\hat{\theta}^t)$  in a manner that will preserve the incentive compatibility constraints of the relaxed problem in all periods whilst raising the within-period utility of an agent of type  $\theta_{t+1}^n$  by an amount  $\nu_n \delta$  at  $t+1$  and raising the within-period utility of the agent at  $t$  by an amount  $-\beta \nu' \pi_{\Theta}^{vec} \delta$ , for any scalar  $\delta$  satisfying  $|\delta| < \varepsilon$  for some  $\varepsilon > 0$ . Additionally, considering period 1 in isolation, for all vectors  $\nu \in \mathbb{R}^N$  that satisfy  $\nu' \pi_{\Theta}^{vec} = 0$  it is possible to perturb the allocations  $X_1^*$  in a manner that will preserve the incentive-compatibility constraints of the relaxed problem whilst raising the within-period utility of an agent of type  $\theta_1^n$  by an amount  $\nu_n \delta$  in period 1, again for any scalar  $\delta$  satisfying  $|\delta| < \varepsilon$  for some  $\varepsilon > 0$ .*

This result is useful for two reasons. First, because it implies a set of ‘intra-temporal’ optimality conditions that must hold in period 1 (the first), which it was not possible to obtain through a proof that appealed along the way to perturbations to allocations in a *prior* time period.<sup>39</sup> Second, because for each ‘utility increment’ vector  $\nu$  there will always be an equivalent  $N \times 1$  vector  $\gamma$ , satisfying:

$$\gamma = (J^{-1})' \nu \quad (49)$$

---

<sup>39</sup>This is the purpose of stating separately the final sentence in the Proposition.

Since  $J^{-1}$  is plainly invertible, the Lemma implies that we can pick composite sums of the marginal utility changes that are induced by each of the  $\Delta^{n'}(0)$  marginal perturbations (for  $n \in \{1, \dots, N\}$ ), in proportions corresponding to any  $\gamma \in \mathbb{R}^N$  (that is,  $\gamma_1$  units of the utility change from marginal perturbation  $\Delta^{1'}$ , plus  $\gamma_2$  units of  $\Delta^{2'}$ , and so on), and know that the overall vector of utility changes described in this way, given by  $J'\gamma$ , can be implemented through some incentive-compatible perturbation (both at the margin and for discrete utility increments).

Moreover, if one solves for the marginal effect on the policymaker's surplus at  $t+1$  that is associated with the perturbation that changes within-period utility away from the optimum according to the vector  $\nu$  (per unit increase in  $\delta$ ), we can also show that this is equal to the effect on the policymaker's surplus of an additive combination of the  $\Delta^{n'}(0)$  marginal perturbations for  $n = \{1, \dots, N\}$ , in proportions that likewise correspond to the entries in the associated  $\gamma$ . Together with movements along the  $N$ th agent's indifference curve (which have no effect on the utility of any agent and thus cannot assist in the provision of utility according to any vector  $\nu$ ), the complete set of  $\Delta^{n'}(0)$  marginal perturbations for  $n \in \{1, \dots, N+1\}$  will therefore *span* the entire set of  $(N+1)$  dimensions along which the consumption and output of all agents can be jointly perturbed at the margin away from the optimal allocation  $X_{t+1}^*$  without undermining the within-period incentive compatibility requirements of the restricted problem at  $t+1$ . This is very useful for technical purposes, since it implies any complex marginal perturbation that to  $X_{t+1}^*$  that is known to satisfy within-period incentive compatibility at  $t+1$  can always be written as a composite of the  $\Delta^{n'}(0)$  matrices – complementing the previous Lemma, which implied that any composite of the  $\Delta^{n'}(0)$  matrices was within-period incentive-compatible.

We summarise the result in the following Lemma, the proof of which is again relegated to the appendix.

**Lemma 9** *Consider perturbations of the form outlined in Proposition 8, generating utility changes at  $t+1$  according to the vector  $\delta\nu$ . For any such  $\nu \in \mathbb{R}^N$  the marginal resource cost of this perturbation to the policymaker at  $t+1$  can always be expressed as the following additive combination of the  $\Delta^{n'}(0)$  matrices for  $n \in \{1, \dots, N\}$ :*

$$\pi_{\Theta}(\theta^t) (\pi_{\Theta}^{vec})' \left[ \sum_{n=1}^N \gamma_n \Delta^{n'}(0) \right] k$$

where  $k$  is here defined as the  $2 \times 1$  vector  $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$  and  $\gamma_n$  is the  $n$ th element in the vector  $\gamma$ , which in turn satisfies:

$$\gamma = (J^{-1})' \nu \quad (50)$$

Moreover, any perturbation to a constrained-optimal allocation that changes allocations at  $t$  and  $t+1$  across agents with a common reporting history to  $t$  whilst

remaining within the constraint set of the relaxed problem must have marginal effects on allocations at  $t + 1$  that are expressible as a linear combination of the  $\Delta^{n'}(0)$  marginal effects alone.

Note that this Lemma does not rely on the iid assumption: if a given perturbation to utilities admissible under Lemma 8 is additionally incentive-compatible under a more general Markov type process then it will also yield a surplus according to the expression above.<sup>40</sup>

Together these results allow us to provide a more general statement of the requirements for ‘intratemporal’ optimality in the iid case – in the style of condition (41) above. The particular usefulness of that condition comes from the fact that it gives a relationship that must hold at the optimum between variables that are *entirely particular to one time period*. In this way we are able to divide up the set of  $N + 1$  optimality conditions that we have derived into  $N$  intratemporal conditions, and just one intertemporal condition. Because an intermediate step in our original derivation of condition (41) was the use of utility perturbations in the *preceeding* time period, at present we strictly can only state it for periods after the first. With the aid of Lemma 8 this can easily be overcome, as the next Proposition shows. First, let  $\gamma$  be any  $N \times 1$  vector with the following property:

$$\gamma' J \pi_{\Theta}^{vec} = 0 \quad (51)$$

where  $J$  is the  $N \times N$  matrix obtained by deleting the last (zero) row from  $\hat{J}$ . Then we have the following result.

**Proposition 10** *Suppose the solution to the relaxed problem also solves the general problem, and that type draws are iid across agents and time. Then for all time periods  $t \geq 1$  and all reporting histories  $\hat{\theta}^t$ , the matrices  $\{\Delta^{n'}(0)\}_{n=1}^N$  associated with the optimal allocation matrix  $X_{t+1}^*(\hat{\theta}^t)$  satisfy the following condition:*

$$(\pi_{\Theta}^{vec})' \left[ \sum_{n=1}^N \gamma_n \Delta^{n'}(0) \right] k = 0 \quad (52)$$

where  $k$  is again the  $2 \times 1$  vector  $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$  and  $\gamma_n$  is the  $n$ th element of any vector  $\gamma$  that satisfies equation (51).

**Proof.** Appealing to Lemma 8, we know that the vector  $\nu$  that solves  $\nu = J'\gamma$  can be applied at the margin to augment the utilities of agents with a common prior type history in any period  $t \geq 1$ , through a change to output and consumption bundles that preserves incentive compatibility. By the definition of  $\gamma$  we will have  $\nu' \pi_{\Theta}^{vec} = 0$ , and thus incentive compatibility in any prior periods is assured without the need for any further perturbations. The result then follows directly from Lemma 9. ■

---

<sup>40</sup>The proof of the Lemma admits this possibility.

Notice that there will, in general, exist  $N - 1$  linearly independent  $\gamma$  vectors satisfying equation (51). Equation (41) (which we may now assert for all time periods, including the first) gives the  $N - 1$  possibilities for which  $\gamma_1 = -1$ ,  $\gamma_n = [\pi_\Theta(\theta_t \geq \theta_t^n)]^{-1}$  for some  $n \in \{2, \dots, N\}$  and any  $t \geq 1$ , and there are zero entries elsewhere. When we move to the non-iid case the set of admissible  $\gamma$  vectors is reduced in an important way, and this matrix representation proves invaluable in characterising composite movements that *are* still possible in that case.

Recall that there will additionally exist an  $N$ th intratemporal condition – the ‘no distortion at the top’ result – to which any set of  $N - 1$  perturbations that satisfy (10) should be added. For all time periods after the first, the generalised inverse Euler condition, (35), provides a further cross-restriction, linking outcomes for agents with a given prior history to their common allocation in the previous period. There are additionally  $N - 1$  binding incentive compatibility constraints across any  $N$  agents who share a common prior history. Finally, there is a single intertemporal budget constraint that the policymaker must satisfy (which may be thought of loosely as ‘substituting’ for the dynamic Euler condition in the very first time period) – ensuring that we always have precisely  $2N$  equations to tie down the  $2N$  variables that are to be determined across the set of agents with a common prior history at any given point in time. In this sense we have provided a complete analytical description of the solution. This is likely to be of great use practically, since it obviates the need to apply dynamic programming techniques in arriving at a numerical solution to any given example. In a finite-horizon model, all that is needed is to solve a known set of simultaneous equations, although in general the number of equations will grow exponentially in the number of time periods.<sup>41</sup>

In an infinite-horizon model one will still need a way to approximate agents’ value functions conditional on any shock history (since these feature in the binding incentive compatibility constraint). But note that in the iid case history dependence can be summarised by a single variable – the marginal cost of uniform incentive-compatible utility provision. This follows from the fact that we can solve for outcomes from period  $t$  onwards for agents with a given history  $\hat{\theta}^{t-1}$  based on a set of optimality conditions and constraints that all depend only on outcomes from  $t$  onwards, plus the generalised inverse Euler condition between  $t - 1$  and  $t$ . The marginal cost of utility provision must be sufficient to summarise the past completely.

## 5.6 Optimal effective income tax rates

The results of the previous subsection allow us to demonstrate a further quite general result with important economic implications.

**Proposition 11** *Suppose the solution to the relaxed problem also solves the general problem. Then for all time periods  $t \geq 1$ , all reporting histories  $\hat{\theta}^{t-1}$*

---

<sup>41</sup>For  $T$  periods there will generally be  $\sum_{t=1}^T 2N^t$  variables to determine.

and all  $\theta_t^n \in \Theta$  the implicit marginal tax rate  $\tau(\theta_t^n)$  satisfies  $\tau(\theta_t^n) \geq 0$ .

**Proof.** Consider the perturbation given by applying just the  $n$ th row of the schedule  $\Delta^{n+1}(\delta)$  at time  $t$  – that is, a movement along the within-period indifference curve of the  $n$ th agent. For negative values of  $\delta$  (only) this will keep us within the constraint set of the relaxed problem, since the net impact on the utility obtainable from reporting  $\hat{\theta}^t$  at  $t$  is zero for truth-tellers and strictly negative for ‘downwards mimickers’, and expected utility in prior periods is left completely unaffected by the fact that all agents at  $t$  are indifferent to this perturbation. Hence the marginal cost as  $\delta$  is moved marginally *below* zero must be weakly positive, given that the optimal solution in the relaxed constraint set solves the general problem. From our earlier results, this implies:

$$-[\varphi_2^c(\theta_{t+1}^n, 0) - \varphi_2^y(\theta_{t+1}^n, 0)] \geq 0 \quad (53)$$

The proof of Proposition 7 shows:

$$\varphi_2^c(\theta_t^n, 0) - \varphi_2^y(\theta_t^n, 0) = -\frac{\tau(\theta_t^n)}{u_c(\hat{\theta}_t^n; \theta_t^{n+1})(1 - \tau(\theta_t^n)) + u_y(\hat{\theta}_t^n; \theta_t^{n+1})} \quad (54)$$

where  $u(\hat{\theta}_t^n; \theta_t^{n+1})$  (and associated partial derivatives) denotes the utility function of an agent whose type is  $\theta_{t+1}^{n+1}$  mimicking one of type  $\theta_{t+1}^n$ . Hence:

$$\frac{\tau(\theta_{t+1}^n)}{u_c(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})(1 - \tau(\theta_{t+1}^n)) + u_y(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})} \geq 0 \quad (55)$$

We have:

$$\begin{aligned} (1 - \tau(\theta_{t+1}^n)) &= -\frac{u_y(\theta_{t+1}^n)}{u_c(\theta_{t+1}^n)} \\ &> -\frac{u_y(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})}{u_c(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})} \end{aligned} \quad (56)$$

where the last inequality is an application of the single-crossing condition. Hence the denominator in condition (55) will be strictly positive, and the result follows. ■

So unlike the savings distortion the direction of the intratemporal distortion on production is completely unambiguous: the optimal effective marginal income tax rate is never negative. Note that we have not had to make any iid assumption in stating this result. This follows from the fact the chosen perturbation does not change equilibrium utility levels across agents, so differences in previous periods between the probability distributions of truth-tellers and mimickers are irrelevant to continued incentive compatibility.

In a sense the result itself should not be surprising. We have already seen that the first-best involves effective marginal tax rates of zero on current income, and there are benefits from moving away from this situation under imperfect information only to the extent that doing so reduces the information rent that higher types are able to extract as compensation for not mimicking. This was the message of Figure 2 above. Since a ‘downwards’ movement along the within-period indifference curve of lower types reduces the utility of higher-type mimickers, it is always better to move to a point where this indifference curve has a slope  $(\frac{dc}{dy})$  that is less than one.

## 6 General perturbations with Markov types

Whilst the iid model is instructive, it is plainly unrealistic as a description of the way individuals’ productivities evolve in practice. To attain some greater realism we need to generalise to allow for persistence in types. The simplest way to do this is to assume the productivity measure  $\pi_\Theta$  incorporates a Markov structure (so  $\pi_\Theta(\theta_{t+1}|\theta^t) = \pi_\Theta(\theta_{t+1}|\theta_t)$ ). Recall from the earlier discussion that our confidence in the first-order approach cannot be so sure in this case – we had to assume increasing differences in the value function at the relaxed problem’s optimum for sufficiency, which was not a condition directly related to the ‘fundamentals’ of the model. We proceed all the same, and leave a more satisfactory resolution of the sufficiency question for subsequent work.

When types follow a general Markov process we are faced with an extra dimension of complication. For agents with a given reporting history  $\hat{\theta}^{t-1}$  we may be able to define a perturbation to allocations at  $t$  that has zero impact on the expected utility at  $t-1$  of a relevant *truth-telling* agent, but the probability distribution under which this expectation is calculated is now particular to that agent. An agent who is, at the optimum, on the cusp of *falsely* reporting  $\hat{\theta}^{t-1}$  will take expectations of the future returns from a mimicking strategy under a different probability distribution to the truth-teller – and thus may well experience a change in the *ex-ante* expected utility from mimicking subsequent to the perturbation. This would undermine local incentive compatibility at time  $t-1$ , for movements in one direction or the other.

In general our aim is, once again, to find a set of distinct functions  $\Delta : \mathbb{R} \rightarrow \mathbb{R}^{2N}$  and  $\Delta_{-1} : \mathbb{R} \rightarrow \mathbb{R}^2$  that can be used to perturb the consumption and output allocations across all agents with a given reporting history  $\hat{\theta}^{t-1}$ , at  $t$  and  $t-1$  respectively, subject to these functions satisfying the three conditions set out at the start of Section 5.2: the preservation of incentive compatibility, continuous differentiability in  $\delta$  in the region of  $\delta = 0$ , and no net impact on expected utility for any agent from the perspective of period  $t-1$  and earlier. It is the first of these conditions – incentive compatibility – that we will no longer necessarily satisfy through applications of the  $\Delta^n$  and  $\Delta_{-1}^n$  schedules defined above. But in certain regards the earlier analysis *does* go through unchanged. We focus on these similarities with the iid problem before turning to the differences.

## 6.1 Equivalences between the Markov and iid cases

Perhaps the most obvious situation in which Markov and iid cases will be equivalent to one another is when we consider perturbations to the allocations at  $t+1$  and (possibly)  $t$  of an agent whose allocation was not ‘envied’ in  $t$ . This could either be because  $t+1 = 1$  (i.e., there was no prior period from the perspective of our policymaker) or because the agent’s type was the highest possible at  $t$  (and thus, by our maintained focus on the ‘restricted problem’, was not envied). In stating this formally it is useful to define  $\pi_{\Theta}^{vec}(\theta_t^n)$  as the vector of  $t+1$  probabilities over  $\Theta$  conditional upon a productivity draw of  $\theta_t^n$  at time  $t$  (stacked in identical fashion to  $\pi_{\Theta}^{vec}$ ). Then we can state the following:

**Proposition 12** *Suppose the solution to the relaxed problem also solves the general problem. Then for all time periods  $t \geq 1$  and any reporting history  $\hat{\theta}^t$  whose terminal entry  $\hat{\theta}_t$  is the maximal element of  $\Theta$ ,  $\theta^N$ , the matrices  $\{\Delta^{n'}(0)\}_{n=1}^N$  associated with the optimal allocation matrix  $X_{t+1}^*(\hat{\theta}^t)$  satisfy the following condition:*

$$\left(\pi_{\Theta}^{vec}(\theta_t^N)\right)' \left[ \sum_{n=1}^N \gamma_n \Delta^{n'}(0) \right] k = 0 \quad (57)$$

where  $k$  is again the  $2 \times 1$  vector  $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$  and  $\gamma_n$  is the  $n$ th element of any vector  $\gamma$  that satisfies the equation  $\gamma' J \pi_{\Theta}^{vec}(\theta_t^N) = 0$ .

Similarly in period 1 the matrices  $\{\Delta^{n'}(0)\}_{n=1}^N$  associated with the initial optimal allocation matrix  $X_1^*$  satisfy the following condition:

$$\left(\pi_{\Theta}^{vec}\right)' \left[ \sum_{n=1}^N \gamma_n \Delta^{n'}(0) \right] k = 0 \quad (58)$$

where  $\gamma_n$  is the  $n$ th element of any vector  $\gamma$  that satisfies the equation  $\gamma' J \pi_{\Theta}^{vec} = 0$  and  $\pi_{\Theta}^{vec}$  is the initial (unconditional) probability vector across productivity draws.

The proof of these claims merely repeats the logic contained in Proposition 10, so is omitted. All we need note is that if the agent whose  $t+1$  allocations are being perturbed was not envied by any other agent at time  $t$  then we do not need to concern ourselves with ensuring the perturbation is utility-neutral at  $t$  for a potential mimicker, and this concern is the only additional problem generated by a switch to Markov transition probabilities. The agent will not have been envied if he or she was of the highest possible type at  $t$ , or if  $t = 0$  – and so  $t+1$  is in fact the first period of the problem.

This result implies all of the ‘intermediate’ intratemporal optimality conditions from the iid case (that is, those associated with differential changes to utility levels for different productivity draws at  $t+1$ ) carry over to the Markov problem for a particular subset of reporting histories. We have additionally

already shown that the two ‘extreme’ perturbations – that is, changes along the top indifference curve and uniform utility provision, which led to the ‘no distortion at the top’ and generalised inverse Euler results respectively – both carry over for all histories under Markov type processes. So all that remains is to understand how the ‘intermediate’ perturbations are affected when agents’ prior allocations *were* envied.

Note also that a static model (with just one time period) will have identical optimality requirements to the intratemporal conditions outlined for the first period here – though the constraint set will be slightly different, since future utilities cannot be used to provide incentives. Again, we are hopeful that the results of this paper can be of use in clarifying the static problem too.

## 6.2 Differences between the Markov and iid cases

There are two important ways in which optimality requirements do change when we switch to the Markov problem. First, the dimensionality of the space within which outcomes can be perturbed to generate *intratemporal* optimality conditions is reduced by one for all agents who were envied in the previous time period. Second, and offsetting this loss of an intratemporal condition, an additional *intertemporal* condition arises, ensuring that the cost to the policymaker of preventing mimicking is spread optimally through time. We explain these points in turn.

### 6.2.1 Intratemporal optimality: a dimension lost

If we are considering a perturbation that applies exclusively in period  $t + 1$  to the allocations of agents with a common reporting history  $\hat{\theta}^t$ , such that  $\hat{\theta}_t = \theta_t^n \neq \theta_t^N$  (where the latter is the maximal element of  $\Theta$ ), we need to make sure that this perturbation does not affect the incentive at  $t$  for truthful reporting – either for an agent whose true type is  $\theta_t^n$  or for one whose true type is  $\theta_t^{n+1}$  (and thus is indifferent at the conjectured optimum between reporting  $\hat{\theta}_t^{n+1}$  or  $\hat{\theta}_t^n$ ). This implies that the expected utility consequences of the perturbation must be zero under both the ‘truth-teller’s’ probability measure  $\pi_\Theta(\cdot|\theta_t^n)$  and the ‘mimicker’s’ measure  $\pi_\Theta(\cdot|\theta_t^{n+1})$ . In the iid case we were able at the margin to implement any linear composite of the ‘basic’ perturbation matrices  $\{\Delta^{n'}(0)\}_{n=1}^N$  provided the vector of relative weights given to each, the  $N \times 1$  vector  $\gamma$ , satisfied  $\gamma' J \pi_\Theta^{vec} = 0$  for the unique probability vector  $\pi_\Theta^{vec}$ . Recall that the  $n$ th row of the matrix  $J$  details the marginal utility consequence of the perturbation  $\Delta^n$  for each type at  $t + 1$ , so this restriction on  $\gamma$  ensures the net effect of the composite perturbation on expected utility is zero under the common probability measure. In general one can always find  $N - 1$  linearly independent  $\gamma$  vectors that satisfy this condition.

By the same logic, when shocks are Markov we can preserve incentive compatibility for both truth-tellers and mimickers provided we perturb outcomes at the margin according to a composite of the basic perturbations for which the



weight vector  $\gamma$  jointly satisfies two conditions:

$$\gamma' J \pi_{\Theta}^{vec}(\theta_t^n) = \gamma' J \pi_{\Theta}^{vec}(\theta_t^{n+1}) = 0 \quad (59)$$

In general one can always find  $N - 2$  linearly independent  $\gamma$  vectors for which this condition is satisfied. Hence the movement to Markov probabilities has denied us the capacity to carry out intratemporal perturbations in precisely one dimension. As in the iid case, corresponding to any such  $\gamma$  vector will again be an equivalent  $\nu$  vector that directly lists the marginal utility effects on agents at  $t + 1$ , given by  $\nu = J' \gamma$ . Lemma 8 can then be easily adjusted to cover intratemporal perturbations in the Markov case:

**Lemma 13** *For all time periods  $t \geq 1$ , all reporting histories  $\hat{\theta}^t$  such that  $\hat{\theta}_t = \theta_t^n \neq \theta_t^N$ , and any vector  $\nu$  that satisfies  $\nu' \pi_{\Theta}^{vec}(\theta_t^n) = \nu' \pi_{\Theta}^{vec}(\theta_t^{n+1}) = 0$  it is possible to perturb the optimal allocations  $X_{t+1}^*(\hat{\theta}^t)$  in a manner that will preserve the incentive compatibility constraints of the relaxed problem in all periods whilst raising the within-period utility of an agent of type  $\theta_{t+1}^n$  by an amount  $\nu_n \delta$  at  $t + 1$  for any  $\delta$  satisfying  $|\delta| < \varepsilon$  for some  $\varepsilon > 0$  and leaving utility in all other periods constant.*

We omit to include a proof, since the logic is identical to that of Lemma 8, except that it is applied here only to the subset of within-period perturbations admissible in the Markov case. The important point is just that the specified non-marginal perturbations can be carried out whilst preserving incentive compatibility for the relaxed problem in earlier periods (even though the basic perturbations  $\{\Delta^n(\delta)\}_{n=1}^N$  and  $\{\Delta_{-1}^n(\delta)\}_{n=1}^N$  may no longer be admissible). Thus the *marginal* utility effects associated with them are implementable in a manner that will keep us within the constraint set of the relaxed problem, and so must come at zero marginal resource cost when the solution to the relaxed problem is known to coincide with the solution to the general problem.

Note also that the proof of Lemma 9 will go through essentially unchanged when we restrict attention to the subset of the possible utility increment vectors  $\nu$  that will permit incentive compatibility to be preserved in the Markov case. This implies that the marginal cost to the policymaker of implementing that vector at  $t + 1$  for all agents with a given reporting history  $\hat{\theta}^t = \theta^t$  (such that  $\hat{\theta}_t = \theta_t^n$ ) will again be:

$$\pi_{\Theta}(\theta^t) (\pi_{\Theta}^{vec}(\theta_t^n))' \left[ \sum_{n=1}^N \gamma_n \Delta^{n'}(0) \right] k$$

So even though weighted *pairwise* sums of the basic perturbations  $\{\Delta^n(\delta)\}_{n=1}^N$  may no longer be compatible with incentive compatibility,<sup>42</sup> any composite per-

<sup>42</sup>For instance, the iid condition (41) is a pairwise sum of the marginal effects of the  $\Delta^1$  and  $\Delta^n$  perturbations for some  $n \in \{2, \dots, N\}$ , with weights  $-\pi_{\Theta}(\theta_{t+1} > \theta_{t+1}^n | \theta_t)$  and 1 respectively.

turbations that still *are* incentive compatible in the Markov case can also still have their marginal costs expressed as a linear combination of the marginal costs of these basic perturbations. Underneath all of our linear algebra remains a set of marginal movements along within-period indifference curves, and marginal compensation payments to mimickers for these.

The required intratemporal optimality conditions can now be stated formally:

**Proposition 14** *Suppose the solution to the relaxed problem also solves the general problem. Then for all time periods  $t \geq 1$  and any reporting history  $\hat{\theta}^t$  such that  $\hat{\theta}_t = \theta_t^n \neq \theta_t^N$ , the matrices  $\{\Delta^{n'}(0)\}_{n=1}^N$  associated with the optimal allocation matrix  $X_{t+1}^*(\hat{\theta}^t)$  satisfy the following condition:*

$$(\pi_{\Theta}^{vec}(\theta_t^n))' \left[ \sum_{m=1}^N \gamma_m \Delta^{m'}(0) \right] k = 0 \quad (60)$$

where  $k$  is again the  $2 \times 1$  vector  $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$  and  $\gamma_m$  is the  $m$ th element of any vector  $\gamma$  that satisfies the two restrictions  $\gamma' J \pi_{\Theta}^{vec}(\theta_t^n) = 0$  and  $\gamma' J \pi_{\Theta}^{vec}(\theta_t^{n+1}) = 0$ .

The proof again follows directly from earlier arguments so is omitted here. Together with the ‘no distortion at the top’ condition, it implies we now have  $N - 1$  linearly independent optimality conditions that must hold within each time period across types that share a common prior report history. The generalised inverse Euler condition gives a further condition (in all periods except the first),<sup>43</sup> and there are  $N - 1$  binding incentive compatibility constraints. Together this implies that we are one equation short of tying down the  $2N$  variables that are to be determined in each period (with the exception of the first, and for any reporting history that did *not* feature the maximal element of  $\Theta$  in the preceding period). The final step in our characterisation is to provide this missing equation.

## 6.2.2 Intertemporal optimality: exploiting dynamic dependencies

Recall again the basic problem faced by our utilitarian policymaker. As we saw in Section 3, the first-best solution would involve all agents facing a within-period marginal income tax rate of zero, so that the marginal utility value of a unit of extra product is equal to its marginal utility cost. At the same time, the marginal utility of consumption would be equalised across agents. When types are unobservable these objectives are mutually incompatible. The ability of higher-type agents to mimic implies they would only report their types truthfully if given substantially more utility than lower types. But by raising the tax wedge on lower types – reducing their consumption and output levels

<sup>43</sup> Again, an intertemporal resource constraint can loosely be thought of as substituting for a dynamic optimality condition in the first time period.

simultaneously along a within-period indifference curve – one can ensure that the marginal benefits to higher types from mimicking are reduced, appealing to the intuition that we developed when presenting Figure 2. This in turn reduces the utility rents that higher types can extract from the policymaker – these rents being spread at the optimum across the contemporary and subsequent periods, in a manner that satisfies the inverse Euler condition. Seen in this light, the problem is one of resolving the trade-off between the provision of wasteful amounts of current and future utility to higher types, and the use of wasteful tax wedges that impede the production of lower types.

When productivity shocks are Markov there is a third alternative available to the policymaker. Instead of reducing higher types' utility rents through tax wedges on lower types, it is possible to do it by 'twisting' the provision of utility across states in subsequent periods, so that the expected benefits to mimickers from a given report are reduced, even whilst the expected benefits to truth-tellers are held constant. That is, if an agent were to report some  $\hat{\theta}^t$  such that  $\hat{\theta}_t = \theta_t^n \neq \theta_t^N$ , it is always possible to shift allocations across states in period  $t + 1$  (relative to the least-cost means of providing a given level of expected utility to truth-tellers) so that agents whose true type is  $\theta_t^{n+1}$  see a reduction in their expected utility from mimicking under the measure  $\pi_\Theta(\cdot|\theta_t^{n+1})$ , whilst expected utility under the measure  $\pi_\Theta(\cdot|\theta_t^n)$  remains unchanged. The theory of the second best suggests there will in general be net benefits to distorting  $t + 1$  allocations in this manner.

Before stating the main argument we must provide an equivalent to Lemma 13 to confirm incentive compatibility for dynamic perturbations. We have the following, the proof of which is in the appendix:

**Lemma 15** *For all time periods  $t \geq 1$ , all reporting histories  $\hat{\theta}^t$  such that  $\hat{\theta}_t = \theta_t^n \neq \theta_t^N$ , and any vector  $\nu$  that satisfies  $\nu' \pi_\Theta^{vec}(\theta_t^n) = 0$  and  $\nu' \pi_\Theta^{vec}(\theta_t^{n+1}) = 1$  it is possible to perturb the optimal allocations  $(c_t^*(\hat{\theta}^t), y_t^*(\hat{\theta}^t))$  and  $X_{t+1}^*(\hat{\theta}^t)$  in a manner that will preserve the incentive compatibility constraints of the relaxed problem in all periods whilst raising the within-period utility of an agent of type  $\theta_{t+1}^n$  by an amount  $\nu_n \delta$  at  $t + 1$  for any  $\delta$  satisfying  $|\delta| < \varepsilon$  for some  $\varepsilon > 0$  and leaving equilibrium utility in all other periods constant.*

This result immediately takes us to the final optimality condition that we desire.

**Proposition 16** *Suppose the solution to the relaxed problem also solves the general problem. Then for all time periods  $t \geq 1$  and any reporting history  $\hat{\theta}^t$  such that  $\hat{\theta}_t = \theta_t^n \neq \theta_t^N$ , the marginal perturbation matrices  $\{\Delta^{n'}(0)\}_{n=1}^N$  associated with the optimal  $t + 1$  allocation matrix  $X_{t+1}^*(\hat{\theta}^t)$  together with the*

optimal  $t$  allocation pair  $(c_t^*(\hat{\theta}^t), y_t^*(\hat{\theta}^t))$  must satisfy the following condition:

$$\begin{aligned} & \beta R_{t+1} \frac{\tau(\theta_t^n)}{u_c(\hat{\theta}_t^n; \theta_t^{n+1}) (1 - \tau(\theta_t^n)) + u_y(\hat{\theta}_t^n; \theta_t^{n+1})} \\ = & (\pi_{\Theta}^{vec}(\theta_t^n))' \left[ \sum_{m=1}^N \gamma_m \Delta^{m'}(0) \right] k \end{aligned} \quad (61)$$

where  $k$  is the  $2 \times 1$  vector  $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$  and  $\gamma_m$  is the  $m$ th element of any vector  $\gamma$  that satisfies the two restrictions  $\gamma' J \pi_{\Theta}^{vec}(\theta_t^n) = 0$  and  $\gamma' J \pi_{\Theta}^{vec}(\theta_t^{n+1}) = -1$ .

**Proof.** We consider a composite perturbation pair, denoted  $\Delta(\delta)$  and  $\Delta_{-1}(\delta)$ , such that  $\Delta(\delta)$  raises the within-period utility of an agent of type  $\theta_{t+1}^m$  by an amount  $\nu_m \delta$  at  $t+1$ , where  $\nu_m$  is the  $m$ th entry of the vector  $\nu = J' \gamma$ . By earlier arguments (c.f. proof of Lemma 9), the marginal cost of this  $\Delta(\delta)$  perturbation as  $\delta$  is moved away from 0, assessed from the perspective of time  $t$ , will be:

$$R_{t+1}^{-1} \pi_{\Theta}(\theta^t) (\pi_{\Theta}^{vec}(\theta_t^n))' \left[ \sum_{m=1}^N \gamma_m \Delta^{m'}(0) \right] k$$

By Lemma 15 we know that we can remain within the constraint set of the relaxed problem through these perturbations, and the fact that the solution to the relaxed problem also solves the general problem will then imply marginal changes cannot raise a surplus. The proof of Lemma 15 shows that incentive compatibility at  $t$  is preserved by moving allocations along the indifference curve of the relevant truth-telling agent with the given report history  $\hat{\theta}^t$ , and doing so by an amount sufficient to increase the within-period utility of a mimicker by  $\beta \delta$  units. Retaining earlier definitions of the functions  $\varphi^c$  and  $\varphi^y$ , the cost of this perturbation, assessed at time  $t$ , will be:

$$\pi_{\Theta}(\theta^t) [\varphi^c(\theta_t^n, \beta \delta) - \varphi^y(\theta_t^n, \beta \delta)]$$

and so the marginal cost as  $\delta$  is moved away from zero is:

$$\beta \pi_{\Theta}(\theta^t) [\varphi^c(\theta_t^n, 0) - \varphi^y(\theta_t^n, 0)]$$

which we have already established (c.f. proof of Proposition 7) is equal to:

$$-\beta \pi_{\Theta}(\theta^t) \frac{\tau(\theta_t^n)}{u_c(\hat{\theta}_t^n; \theta_t^{n+1}) (1 - \tau(\theta_t^n)) + u_y(\hat{\theta}_t^n; \theta_t^{n+1})}$$

The result then follows from the fact that the total present value of the marginal cost of the perturbation must be zero at an optimum. ■

It is well known that the shift from iid to Markov transition probabilities

complicates substantially the computation of optimal dynamic policy in models such as this – the point is explored at length, for instance, by Fernandes and Phelan (2000) in the context of a dynamic agency model, and by Kapička (2010) in the context of dynamic Mirrleesian problems. Equation (61) provides one interpretation for why this is so: when shocks are Markov the policymaker has the capacity to spread through time the costs of any given utility advantage that mimickers have over truth-tellers, and it is always optimal to exploit this. That fact introduces an extra dynamic optimality requirement, on top of the generalised inverse Euler condition.<sup>44</sup> This implies one needs to know much more information about past productivity draws when solving for an optimal within-period allocation in the Markov case than in the iid case, since one must ascertain not just the average level of the marginal cost of utility provision to implement across agent types within a period, but also the extent to which allocations should be ‘twisted’ to reduce prior benefits to mimicking.

It is also worth emphasising that the benefits to twisting allocations in this way are time-inconsistent. As Proposition 12 shows, if  $t = 1$  there would be no incentives to set a value for  $(\pi_{\Theta}^{vec}(\theta_{t-1}^n))' \left[ \sum_{m=1}^N \gamma_m \Delta^{m'}(0) \right] k$  different from zero (given  $\gamma' J \pi_{\Theta}^{vec}(\theta_{t-1}^n) = 0$ ), so in all subsequent periods an ‘uncommitted’ policymaker would have an incentive to revert to the least-cost means of providing a given utility distribution to agents with a known prior history.

In general, the optimality consideration highlighted here is likely to result in greater equality at  $t + 1$ , conditional upon a given history to  $t$ , the higher is the marginal tax rate for an agent at  $t$ . This is because, as just discussed, higher marginal rates are really a means for the policymaker to reduce the utility gap that has to exist between agents of adjacent types in order to prevent mimicking by the more productive. But one can also reduce this gap by reducing the benefits higher types could expect to obtain in future periods subsequent to mimicking, assessed under their type-specific probability distribution. Assuming this latter distribution places greater weight on higher-type outcomes in the future than does the distribution specific to truth-tellers (as would be the case if  $\pi_{\Theta}(\cdot|\theta_t^n)$  was first-order stochastically dominated by  $\pi_{\Theta}(\cdot|\theta_t^{n+1})$ , for instance), one can disadvantage mimickers at  $t$  whilst leaving truth-tellers unaffected in expected utility terms by shifting  $t + 1$  utility away from higher types and towards lower types. Thus the ‘twisting’ that we have highlighted seems very likely to move outcomes towards greater equality in future *utilities* the higher are initial tax rates (and thus the greater is the distortion the policymaker is willing to accept).<sup>45</sup>

<sup>44</sup>Kapička (2010) makes a similar observation when using a first-order value function method to study a specific example of a dynamic Mirrleesian model. The idea is also implicit in the general treatment of dynamic incentive provision under the first-order approach by Pavan, Segal and Toikka (2011).

<sup>45</sup>Note that greater equality in utilities will generally require greater distortions to the production efficiency of lower types, with downward movements along their within-period indifference curves relative to the iid case, so that higher types are not given an incentive to mimic. Farhi and Werning (2010) use a simulated model with Markov transitions to show that the average income taxes across types do indeed increase through time – entirely consistent

## 7 Martingale convergence results

The final major area on which it is worth focusing attention is the evolution of optimal outcomes over time, and in particular at the limit as the time horizon becomes large. Suppose that the real interest rate were in all time periods equal to the inverse of the discount factor  $\beta$ . Then the generalised inverse Euler equation can be written as:

$$\begin{aligned} & \frac{1 - \alpha(\theta_t)}{u_c(\theta_t) + u_y(\theta_t)\alpha(\theta_t)} \\ &= \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1}|\theta_t) \frac{1 - \alpha(\theta_{t+1})}{u_c(\theta_{t+1}) + u_y(\theta_{t+1})\alpha(\theta_{t+1})} \end{aligned} \quad (62)$$

That is to say, we have a martingale in the marginal cost of (locally incentive compatible) utility provision. When preferences are separable between consumption and labour supply,  $\alpha(\theta_t) = 0$  holds, and the expression collapses to a martingale in the inverse of the marginal utility of consumption – an object that is strictly positive and (under the Inada conditions that we have assumed) bounded below at 0. As many authors have observed, this boundedness allows the application of Doob’s martingale convergence theorem, which implies almost sure convergence in the inverse marginal utility of consumption to a finite (possibly random) limit. If one can also show that the optimum will never involve consumption staying fixed at a non-zero value (which is a likely consequence of the policymaker’s ever-present need to provide incentives<sup>46</sup>), convergence to zero consumption becomes the only possibility.

To generalise these results to the case at hand we need to put a bound on the object in (62) – the marginal cost of utility provision – for preference structures more general than the separable case. A first step is the following.

**Lemma 17** *Under an optimal plan that solves the restricted problem,  $u_c(\theta_t) + u_y(\theta_t)\alpha(\theta_t) > 0$  always holds.*

**Proof.** By definition

$$\alpha(\theta_t^n) = \frac{u_c(\theta_t^n) - u_c(\hat{\theta}_t^n; \theta_t^{n+1})}{u_y(\hat{\theta}_t^n; \theta_t^{n+1}) - u_y(\theta_t^n)} \quad (63)$$

for  $n < N$ , and  $\alpha(\theta_t^N) = 0$ . In the latter case the result follows immediately

---

with our theoretical result.

<sup>46</sup>In a useful discussion, Kocherlakota (2011) notes the possibility of convergence in consumption to one of the endpoints of some bounded interval of the real line in the event that the marginal disutility of labour supply is bounded away from zero and total labour supply has an upper limit. The intuition here is that when agents are sufficiently ‘wealthy’ or sufficiently poor they will, respectively, work zero or the maximum possible number of hours whatever their productivity draw – so stable consumption is possible following convergence to these limits.

from  $u_c(\theta_t) > 0$ . In the former case we have from equation (5):

$$\frac{u_y(\hat{\theta}_t^n; \theta_t^{n+1})}{u_y(\theta_t^n)} < \frac{u_c(\hat{\theta}_t^n; \theta_t^{n+1})}{u_c(\theta_t^n)} \quad (64)$$

Rewriting our object of interest, we have:

$$u_c(\theta_t) + u_y(\theta_t) \alpha(\theta_t) = \frac{u_c(\hat{\theta}_t^n; \theta_t^{n+1}) - u_c(\theta_t^n) \frac{u_y(\hat{\theta}_t^n; \theta_t^{n+1})}{u_y(\theta_t^n)}}{1 - \frac{u_y(\hat{\theta}_t^n; \theta_t^{n+1})}{u_y(\theta_t^n)}} \quad (65)$$

The numerator of the right-hand side is clearly positive by the preceding inequality, and the denominator likewise by the fact the marginal disutility of production is lower for higher types (c.f. inequality (2)). ■

Given the definition of  $\alpha(\theta_t)$  this allows us almost immediately to state a bound when consumption and labour supply are Edgeworth substitutes. But when they are Edgeworth complements our scope for doing so proves surprisingly limited. Taken together we have the following result.

**Lemma 18**  $\frac{1-\alpha(\theta_t)}{u_c(\theta_t)+u_y(\theta_t)\alpha(\theta_t)} > 0$  always holds under an optimal plan that solves the restricted problem, unless (a) consumption and labour supply are Edgeworth complements, and (b) productivities follow a non-iid process.

**Proof.** With separability between consumption and labour supply  $\alpha(\theta_t) = 0$ , and the assumption  $u_c(\theta_t) > 0$  is enough to confirm the result. When consumption and labour supply are Edgeworth substitutes we have  $\alpha(\theta_t^n) < 0$  (the marginal utility of consumption is higher for mimickers than truth-tellers, since the former need not work so hard to produce a given level of output), and the result follows from Lemma 17. When consumption and labour supply are Edgeworth complements it is possible to prove the bound only for the iid case. The reasoning is far more involved, and we relegate it to an appendix. ■

Having put a zero lower bound on the marginal cost of utility provision for these specific cases, when  $R_t = \beta^{-1}$  for all  $t$  a direct application of Doob's martingale convergence theorem implies the object  $\frac{1-\alpha(\theta_t)}{u_c(\theta_t)+u_y(\theta_t)\alpha(\theta_t)}$  must converge almost surely along all realisations of  $\theta^\infty$  to some value  $X \in [0, \infty)$ , where  $X$  is potentially a random variable. We want to be able to say more about the value of  $X$ . In fact, it turns out – as in the separable case – that  $X$  must equal zero. The next Proposition establishes this.

**Proposition 19** Suppose  $R_t = \beta^{-1}$  for all  $t \geq 1$ . Then  $\frac{1-\alpha(\theta_t)}{u_c(\theta_t)+u_y(\theta_t)\alpha(\theta_t)} \xrightarrow{a.s.} 0$  holds under any optimal plan that solves the restricted problem, unless (a) consumption and labour supply are Edgeworth complements, and (b) productivities follow a non-iid process.

**Proof.** See appendix. ■

This result is an obvious generalisation of the ‘immiseration’ results obtained by studying convergence of the standard inverse Euler condition. Moreover, almost sure *consumption* immiseration (in the sense that inverse of the marginal utility of consumption – and hence consumption itself – must tend to zero for almost all agents) is a direct implication of this result, when one recalls that  $\frac{1-\alpha(\theta_t)}{u_c(\theta_t)+u_y(\theta_t)\alpha(\theta_t)} = \frac{1}{u_c(\theta_t)}$  when  $\theta_t = \theta_t^N$  (the highest type): the outcome for an agent who draws the top productivity parameter in the  $t$ th period *must* be immiseration (almost surely) at the limit as  $t$  becomes large, and incentive compatibility then demands that all lower types with the same history must have a still worse lot. So the more complicated nature of the expression for the marginal cost of utility provision in the non-separable case does not undermine the extreme predictions regarding long-run consumption when martingale convergence *can* be applied. The political difficulties associated with long-run commitment to a scheme with such severe future outcomes are plainly immense, even abstracting from the more fundamental question of whether the welfare of the initial period’s cohort of agents *ought* to be the exclusive concern for public policy.<sup>47</sup> For this reason alone the immiseration result is a troubling one: it is hard to imagine a scheme more likely to result in government default than one that demands its future citizens should be enslaved to pay the debts of the past.<sup>48</sup>

Perhaps the more surprising result of this section, though, is that when productivity follows a Markov process and consumption and labour supply are Edgeworth complements – so that those who are working longer hours with a given level of consumption have a higher marginal utility of consumption – we *cannot* put a zero lower bound on the marginal cost of utility provision. Indeed, it is quite possible that this marginal cost may turn negative. This possibility we are able to confirm through a finite-horizon computed example, the details of which we now present.

## 7.1 Computed example

We assume that production is linear in labour supply, with the marginal product of labor equal to  $\theta$ , and that the utility function takes the form outlined in King, Plosser and Rebelo (1988):

$$u(c, y; \theta) = \frac{c^{1-\varsigma}}{1-\varsigma} \exp \left\{ (\varsigma - 1) v \left( \frac{y}{\theta} \right) \right\} \quad (66)$$

<sup>47</sup>This latter question is explored in detail by Farhi and Werning (2007).

<sup>48</sup>Clearly if consumption is reaching zero at the limit then the within-period surplus raised for almost all histories must be substantial. We know, for instance, that ‘top’ agents will certainly be producing very large quantities of output, since  $u_c + u_y = 0$  for these types.

This surplus must be being used either to service interest on outstanding debts or to fund the lavish consumption of some measure-zero subset of agents whose luck has never been out. The latter is probably even less politically plausible than the former.



with the labour disutility schedule  $v$  defined by:

$$v(l) = \frac{l^{1+v}}{1+v} \quad (67)$$

This function implies that consumption and labour supply are Edgeworth complements provided  $\varsigma > 1$ , and are Edgeworth substitutes for  $\varsigma < 1$ .

A substantial practical advantage of the solution method presented in this paper is that it provides a complete set of equations necessary to solve any given example – so provided there is a finite number of types and of time periods, for any given parameterisation we can obtain a solution simply by solving these equations numerically. Specifically, if  $T$  is the total number of time periods and  $N$  the cardinality of  $\Theta$  then we will have  $\sum_{t=1}^T 2N^t$  variables to tie down in total (in each period, an output and consumption level for an agent of each current type, for each history). The method presented above delivers precisely this number of equations, which can be jointly solved to machine accuracy using standard non-linear solution algorithms. Unlike methods that exploit value function iteration, the approach is equally fast whether shocks follow an iid or a Markov process, with the latter simply involving a slightly different set of equations.

For our example we assume two types (identical across all time periods):  $\theta_L$  and  $\theta_H$ , with  $\theta_L < \theta_H$ . Transition probabilities are denoted as follows:

$$\begin{aligned} \pi_{\Theta}(\theta_t = \theta_H) &= P^H & \text{if } t = 1 \\ \pi_{\Theta}(\theta_t = \theta_H | \theta_{t-1} = \theta_H) &= P_H^H & \text{if } t > 1 \\ \pi_{\Theta}(\theta_t = \theta_H | \theta_{t-1} = \theta_L) &= P_L^H & \text{if } t > 1 \end{aligned}$$

We set  $T = 6$ , implying 252 variables to determine. Since at this stage the purpose of the example is more to find a counterexample to  $\frac{1-\alpha(\theta_t)}{u_c(\theta_t)+u_y(\theta_t)\alpha(\theta_t)} > 0$  than to claim realism *per se*, and since this counterexample is more likely to arise in our finite horizon the greater is the value of  $\varsigma$ ,<sup>49</sup> we choose the relatively high value:  $\varsigma = 10$ . For the other parameters we choose values  $v = 2$  and  $\beta = 0.99$ . We normalise  $\theta_L = 1$  and set  $\theta_H = 2$ . The initial probability  $P^H$  we set to 0.5, with strong type persistence thereafter:  $P_H^H = 0.9$  and  $P_L^H = 0.1$ .

Figure 4 is a histogram summarising the distribution of the marginal cost of utility provision across agents in the 6th (and final) period of the simulation, with bins 0.1 units wide (the units here being the single consumption good). The high degree of persistence accounts for this distribution's clear bimodal character.<sup>50</sup> What is of more interest is that the marginal cost of utility pro-

<sup>49</sup>High values of  $\varsigma$  imply strong complementarity, and thus a much lower marginal utility of consumption for mimickers at a given allocation than for truth-tellers. To offset this requires utility provision along a vector that will increase production requirements significantly alongside any extra consumption provision (this exploits the higher marginal disutility of production on the part of truth-tellers), with greater production increases the greater are complementarities. Since the marginal cost of utility provision is lower the more output is increased for a given consumption increase, higher complementarity is likely to be associated in general with lower marginal costs.

<sup>50</sup>Roughly three fifths of agents draw the same type in all six periods.

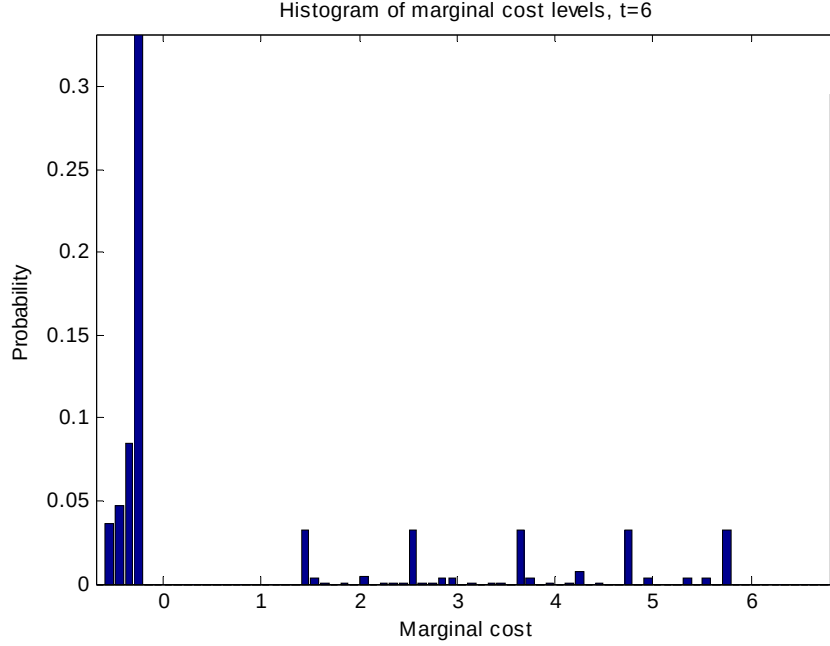


Figure 4: Distribution of marginal cost of utility provision in 6th period

vision (provision, that is, in a manner that preserves *within*-period incentive compatibility) is negative for exactly half of the agents in this period. These agents are the half of the population with contemporaneous productivity  $\theta_L$ .<sup>51</sup>

A negative marginal cost of utility provision also obtains for almost all low-type agents in the 5th period of the simulation, so the result is not dependent upon the period in question being the last. On the surface it is a very counter-intuitive outcome (surely the policymaker can provide utility to a subset of agents and generate a surplus?), so it is worth providing a detailed explanation for it. Recall that when consumption and labour supply are Edgeworth complements, a provision of utility by consumption increments alone at a given output level would benefit low types by more than (mimicking) high types, since the latter supply less labour to produce the given quantity of output – and thus do not benefit from complementarities to so great an extent. Hence to preserve incentive compatibility (for utility movements in either direction) any consumption increment must be accompanied at the margin by an increase in production, which causes greater marginal disutility to lower types than higher (the former are already working longer hours, so their marginal disutility of effort is greater), offsetting the utility imbalance.

<sup>51</sup>Recall that high-type agents must be associated with positive marginal costs, since  $\alpha(\theta_H) = 0$  always holds.

The results of the simulation suggest that the choices of low types are, at the optimum, being distorted sufficiently far away from a point at which the slope of their within-period indifference curve equals one that even movement along a vector giving *equal* consumption and output increments would still raise their utility by more than it would raise the utility of high-type mimickers – and so output must be increased by *more* than consumption at the margin to obtain balance. Notice that this suggests the output of low-type agents is being restricted substantially at the optimum: the lower is output the lower is the *difference* in the marginal effect on utility of an increase in it between truth-tellers and mimickers, and so the more it must be raised for an incentive-compatibility-preserving perturbation.

Why is it not possible to exploit the negative cost of utility provision to generate a surplus? The main reason for this is just that there does not exist a means to provide utility to a given agent in a way that generates resources *whilst at the same time offsetting any effects on incentive compatibility constraints*. A gift of extra utility to a low-type agent in the 6th period would induce high-type agents with the relevant prior history to switch to a mimicking strategy. The cost of preventing this, through an equal utility increment to a high-type agent, may directly offset the generation of a surplus. Even if not, incentive compatibility in the 5th period would also be violated if we are considering allocations to those whose prior type report was  $\hat{\theta}_L$ . Equally in the 5th period, a gift of utility to a low-type agent whose marginal cost is negative could be incentive-compatible if accompanied by a reduction in utility across all agents in the subsequent period; but the aggregated present value of the (negative) costs of these perturbations will be zero, by the generalised inverse Euler condition. Ultimately, no matter what composite marginal perturbation one tries to construct, either local incentive compatibility must be violated, or no surplus raised.

The important question that follows from these results is whether the *potential* for a more benign long-run outcome than immiseration is indeed likely to be realised in the event of complementarity: just because we cannot prove it by martingale convergence does not mean immiseration can be ruled out. One can only conjecture in the absence of a full solution to the infinite-horizon model, but there are reasonable economic grounds for believing immiseration will be avoided. Specifically, note that the tendency towards immiseration (when it does hold) must derive in part from the finite stock of resources at the policymaker's disposal. As time progresses, either a prior tendency to front-load utility provision through debt finance, or promises of very high utility levels to a measure-zero (perpetually lucky) subset of agents, or some combination of the two, results in the maximum possible surplus being extracted from almost all agents. But if in the case of complementarities the marginal cost of reducing the utility of agents turns negative then a tendency to immiserate may well be counter-productive – costing resources rather than generating them. Clearly the policymaker has no direct desire to see immiseration occur, so it seems unlikely that this cost will be worth paying.

## 8 A last word on savings wedges and immiseration

The results of the previous section – in particular Lemma 18 – allows for a slight extension to the set of circumstances in which we can claim it is optimal to deter savings (in some meaningful sense). We can state the following.

**Proposition 20** *Suppose the solution to the relaxed problem also solves the general problem. Then for all time periods  $t \geq 1$  and for all reporting histories  $\hat{\theta}^t$ , if consumption and labour supply are Edgeworth complements then savings will be deterred at the optimum, in the sense that the allocations  $(c_t^*(\hat{\theta}^t), y_t^*(\hat{\theta}^t))$  and  $X_{t+1}^*(\hat{\theta}^t)$  will satisfy inequality (36), with that inequality holding strictly so long as the object  $\frac{u_c(\theta_{t+1}) + u_y(\theta_{t+1})\alpha(\theta_{t+1})}{1 - \alpha(\theta_{t+1})}$  varies for different draws of  $\theta_{t+1} \in \Theta$ .*

The proof is identical to that of Proposition 6, which can be applied whenever the bound  $\frac{1 - \alpha(\theta_t)}{u_c(\theta_t) + u_y(\theta_t)\alpha(\theta_t)} > 0$  holds – which we now know to be the case under complementarity and iid productivity draws by Lemma 18. What is interesting here is that the cases in which we can say with certainty that it is optimal to deter savings (relative to some optimality criterion that would have to hold under autarky) are precisely the cases in which we can confirm immiseration as a limiting outcome: essentially, all situations *except* that of Markov productivity draws and complementarity. This is unlikely to be a coincidence. If savings are being distorted at the optimum, the policymaker is implicitly choosing to ‘front-load utility’ in expectation. This is just a direct reading of inequality (36). But if utility is being front-loaded it would not be at all surprising if the policymaker’s wealth were deteriorating continually over time – so that outstanding obligations eventually become cripplingly large as time passes. In this case agents in the economy would have to put in large amounts of work for little or (at the limit) no return, just to preserve the tax scheme’s solvency. This implies immiseration. Only when the optimality of ‘front-loading’ utility no longer necessarily goes through can we escape this almost sure immiseration.

## 9 Conclusion

The main contribution of this paper is a methodological one. Dynamic models with asymmetric information are a growing source of interest to macroeconomists, and the dynamic version of the Mirrlees income tax problem has generated particular interest. But practically all of the analysis of these models to date has relied on the recursive computation of value functions, defined by a Bellman-type operator appropriately augmented to ensure past promises are kept. These methods are extremely powerful and widely applicable, but their results can be difficult to interpret, simply because it is not always clear exactly which trade-offs have contributed to generating a given policy function or time-path for a variable of interest. Our analysis gives an alternative means to gain

insight into this class of problems, through carefully-chosen perturbations to optimal allocations. In particular, we appeal to the revelation principle to treat the optimum as one in which individuals make direct reports of their types, and investigate how to perturb allocations along a dimension chosen to ensure there will be no changes to these reports – at least for small perturbations. This approach allows us to obtain a complete set of optimality conditions that, together with the binding incentive compatibility restrictions and the resource constraint, are sufficient to characterise the problem’s solution. The method is analogous to solving dynamic consumption choice problems by noting that the marginal rate of substitution must equal the price ratio between any two goods whenever the consumer is at an interior optimum: in our case as there, the optimality conditions obtained do not directly make use of information from the problem’s constraints – this being introduced at a subsequent stage in solving the model.

The equivalent of the requirement of an interior optimum in our setting is that we must know in advance exactly which incentive compatibility constraints bind at the optimum. In the static Mirrlees problem the single crossing condition is known to ensure these constraints bind ‘downwards’ locally, and we present sufficient conditions relating to the optimal allocation that can be checked to verify whether this extends to the dynamic case for any given example. We proceed under the assumption that it does, but *ex ante* knowledge of this essential characteristic of the solution is undoubtedly the chief disadvantage of the method we present.

The optimality conditions that we derive are easiest to understand through a graphical representation of the problem in output-consumption space. They are a set of cross-restrictions on (a) the cost to the policymaker of moving ‘along’ each agent’s within-period indifference curve, reducing that agent’s consumption and output jointly, and (b) the cost of providing a unit of utility to each agent in such a way that a mimicking higher-type agent would receive the same utility increment. Appropriately-chosen composites of these movements, either within or across periods, can ensure local incentive compatibility always continues to hold, and so cannot be applied in the neighbourhood of the optimum in a way that would generate a surplus for the policymaker.

This analytical method is likely to be very useful from a computational perspective, since it eliminates any need to solve maximisation problems directly when calculating the optimum to a given problem. Instead, one need only impose (jointly) the complete set of equations known to characterise that optimum. When the problem has a finite and sufficiently small number of time periods, and relatively small set of productivity types, the solution can be established to machine accuracy by solving a quite manageable set of simultaneous equations. In an infinite horizon problem functional approximation will still be necessary, since future values feature in incentive compatibility constraints, but these values should be expressible as functions of a relatively small set of variables, and will not have to be defined by any maximisation or supremum operator (or similar). In the iid case, for instance, intertemporal optimality can be ensured by linking outcomes at  $t + 1$  for agents with a common history simply to the mar-

ginal cost of utility provision to those agents at  $t$ . This marginal cost variable alone should then be enough to establish the value function.

But the focus of the paper has been on exploiting the analytical results that a perturbation approach can expose, and here there are several. On a higher theoretical level, we have shown that when productivity draws are iid the problem separates into intratemporal and intertemporal dimensions, with the set of intratemporal optimality restrictions that must hold being identical to those that are necessary in a static model, and a single dynamic optimality condition all that is required to ensure an optimal use of resources through time. In the more realistic case that productivity draws follow a Markov process with persistence, one extra dynamic optimality condition emerges – reflecting an extra ability that the policymaker now has to exploit differences in productivity measures between mimickers and truth-tellers, in order to spread distortions through time. Accompanying this is a reduction by one in the number of intratemporal optimality conditions that can be stated. Rather like the use of separability in utility functions to simplify the statement of optimal consumption choices, this partition of the problem can, it is hoped, make the character of its solution much easier to understand.

From a more practical perspective, we have shown that many of the well-known results from static income tax theory generalise to the dynamic case. In particular, regardless of whether the shock process is Markov or iid we can show that effective within-period marginal income tax rates are always weakly positive at the optimum – in the sense that the solution always involves individuals being willing to produce at the margin for a return that is (weakly) less than their marginal product. Moreover, agents whose type is the highest always have a zero effective marginal tax rate, and these are the only agents who do so.

Turning to savings taxes, it is already well-known that in the event of separability between consumption and labour supply it is optimal to apply a positive tax wedge to savings, in the sense that the marginal utility of consumption in period  $t$  is below its expected value at  $t+1$  (allowing for discounting and the interest rate): this follows from the well-known ‘inverse Euler equation’ that holds in that case, combined with Jensen’s inequality. We have been able to generalise this result in two regards. First, and rather limited in its scope, we have shown that the marginal utility of consumption for an agent whose productivity type is the highest possible must also be below its expected value in the next period when consumption and labour supply are Edgeworth substitutes. But one need not focus simply on the *consumption* Euler equation as characterising dynamic optimality: the marginal rate of substitution between output levels in one period and the next, or between arbitrary vector combinations of consumption and output in one period and the next, must likewise equal the intertemporal price ratio at any autarkic allocation. Specifically, the inverse of the marginal cost of incentive-compatible utility provision is the marginal utility associated with a particular joint change in consumption and output, and the existence of an optimality condition relating to this object allows us to confirm that savings are always deterred at the optimum (in an economically meaningful sense) unless consumption and labour supply are Edgeworth complements *and* productivity

draws are non-iid.

This latter result has strong connections with the final area that we have investigated in detail: allocations in the long run. Once again, except in the case that consumption and labour supply are Edgeworth complements and productivity draws are Markov, we have been able to put a zero lower bound on the marginal cost of incentive-compatible utility provision – which in turn will follow a martingale process in the event that the real interest rate equals the inverse of the discount factor  $\beta$ . Martingale convergence theorems then imply almost sure immiseration for all agents in the economy under standard preference assumptions. With complementarity and Markov shocks we have shown by counterexample that the marginal cost of utility provision can in fact turn negative, and so convergence to miserable outcomes need not take place. Indirectly this result seems to shed some light on the cause of immiseration under alternative assumptions: the fact that immiseration need not occur in precisely the same case that savings need not be deterred at the optimum suggests a connection between the implicit decision on the part of the policymaker to front-load the provision of utility when savings *are* being deterred – a strategy that is likely to involve some initial borrowing – and immiseration as the costs of servicing the resulting public debt burden accumulate.

Finally, we note that the methods used in this paper can be applied more widely, albeit with some adaptation. For instance, a companion paper (Brendon (2011)) outlines a similar perturbation method applicable to dynamic Mirrlees problems in which the type space  $\Theta$  is a continuum. Reassuringly all of the results from this paper extend to that case in the natural way, and we are able to provide an expression for optimal marginal tax rates in the *static* model that is considerably simpler than those available in the literature to date. A second area of applicability is to dynamic agency models, where a similar set of optimality conditions can be derived under the assumption that the ‘first order approach’ is valid.

## References

- [1] Abreu, D., D. Pearce and E. Stacchetti (1990), ‘Toward a Theory of Discounted Repeated Games with Imperfect Monitoring’, *Econometrica*, 58, 1041–1063.
- [2] Atkeson, A. and R.E. Lucas Jr. (1992), ‘On Efficient Distribution with Private Information’, *Review of Economic Studies*, 59, 427–453.
- [3] Atkinson, A.B. and J.E. Stiglitz (1980), *Lectures on Public Economics*, London, McGraw-Hill.
- [4] Brendon, C.F. (2011), ‘Applying perturbation methods to dynamic tax problems with a continuum of types’, Manuscript (available on request).

- [5] Broer, T., M. Kapička and J. Klein (2011), ‘Consumption Risk Sharing under Private Information when Earnings are Persistent’, Manuscript (available at <http://paulklein.se/newsite/research/risksharing.pdf>).
- [6] da Costa, C., and I. Werning (2002), ‘Commodity Taxation and Social Insurance’, Manuscript (available at <http://economics.mit.edu/files/1265>).
- [7] Dasgupta, P. (1982), ‘Utilitarianism, Information and Rights’, in A. Sen and B. Williams (eds.), *Utilitarianism and Beyond*, Cambridge, Cambridge University Press.
- [8] Diamond, P.A. and J.A. Mirrlees (1978), ‘A Model of Social Insurance with Variable Retirement’, *Journal of Public Economics*, 10, 295–336.
- [9] Farhi, E. and I. Werning (2007), ‘Inequality and Social Discounting’, *Journal of Political Economy*, 115, 365–402.
- [10] Farhi, E. and I. Werning (2010), ‘Insurance and Taxation over the Life Cycle’, Manuscript (available at <http://econ-www.mit.edu/files/6588>).
- [11] Fernandes, A. and C. Phelan (2000), ‘A Recursive Formulation for Repeated Agency with History Dependence’, *Journal of Economic Theory*, 91, 223–247.
- [12] Golosov, M., N. Kocherlakota and A. Tsyvinski (2003), ‘Optimal Indirect and Capital Taxation’, *Review of Economic Studies*, 70, 569–587.
- [13] Golosov, M., M. Troshkin and A. Tsyvinski (2011), ‘Optimal Taxation: Merging Micro and Macro Approaches’, *Journal of Money, Credit and Banking*, 43, 147–174.
- [14] Golosov, M., A. Tsyvinski and I. Werning (2006), ‘New Dynamic Public Finance: A User’s Guide’, in *NBER Macroeconomic Annual 2006*, Cambridge MA, MIT Press.
- [15] Greenwald, B., and J.E. Stiglitz (1986), ‘Externalities in Economies with Imperfect Information and Incomplete Markets’, *Quarterly Journal of Economics*, 101, 229–264.
- [16] M. Kapička (2010), ‘Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach’, Manuscript (available at <http://www.econ.ucsb.edu/~mkapicka/persistent.pdf>), and forthcoming in *Review of Economic Studies*.
- [17] Kocherlakota, N.R. (1996), ‘Implications of Efficient Risk Sharing without Commitment’, *Review of Economic Studies*, 63, 595–609.
- [18] Kocherlakota, N.R. (2005), ‘Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation’, *Econometrica*, 73, 1587–1621.



- [19] Kocherlakota, N.R. (2011), *The New Dynamic Public Finance*, Princeton, Princeton University Press.
- [20] Mankiw, N.G., M. Weinzierl and D. Yagan (2009), ‘Optimal Taxation in Theory and Practice’ *Journal of Economic Perspectives*, 23, 147–174,
- [21] Marcet, A. and R. Marimon (1998), ‘Recursive Contracts’, Manuscript, Universitat Pompeu Fabra.
- [22] Mele, A. (2011), ‘Repeated Moral Hazard and Recursive Lagrangeans’, Manuscript (available at <http://sites.google.com/site/meleantonio/#research>).
- [23] Messner, M., N. Pavoni and C.M. Sleet (2011), ‘Recursive Methods for Incentive Problems’, Manuscript (available at SSRN: <http://ssrn.com/abstract=1792816>).
- [24] Mirrlees, J.A. (1971), ‘An Exploration in the Theory of Optimum Income Taxation’, *Review of Economic Studies*, 38, 175–208.
- [25] Pavan, A., I. Segal and J. Toikka (2011), ‘Dynamic Mechanism Design: Incentive Compatibility, Profit Maximization and Information Disclosure’, Manuscript (available at <http://www.stanford.edu/~isegal/dmd.pdf>).
- [26] Phelan, C. (2006), ‘Opportunity and Social Mobility’, *Review of Economic Studies*, 73, 487–505.
- [27] Roberts, K. (2000), ‘A Reconsideration of the Optimal Income Tax’, in Hammond, P.J., and G.D. Myles (eds.), *Incentives and Organization: Papers in Honour of Sir James Mirrlees* (Oxford: Oxford University Press).
- [28] Saez, E. (2001), ‘Using Elasticities to Derive Optimal Income Tax Rates’, *Review of Economic Studies*, 68, 205–229.
- [29] Sleet, C.M. and Ş. Yeltekin (2010a), ‘The Recursive Lagrangian Method: Discrete Time’, Manuscript (available at <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.167.7942>).
- [30] Sleet, C.M. and Ş. Yeltekin (2010b), ‘Misery and Luxury: Long Run Outcomes with Private Information’, Manuscript (available at <https://student-3k.tepper.cmu.edu/gsiadoc/WP/2011-E19.pdf>).
- [31] Thomas, J.P. and T. Worrall (1990), ‘Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-Agent Model’, *Journal of Economic Theory*, 51, 367–390.

## A Appendix

### A.1 Proof of property 4 of first-best allocation (decreasing utility in type)

It is useful first to show that normality of leisure implies  $u_{cc} + u_{cy} < 0$ . The consumer's within-period problem (when income is untaxed at the margin) is:

$$\max_{c,y} u(c, y; \theta)$$

subject to

$$c = y + \omega$$

for some endowment  $\omega$ . At any interior optimum we will have  $u_c = -u_y$ , and differentiating both this and the budget constraint totally with respect to  $\omega$  gives:

$$\frac{dy}{d\omega} = -\frac{u_{cc} + u_{cy}}{u_{cc} + 2u_{cy} + u_{yy}}$$

The denominator here is negative by the negative definiteness of the partial Hessian, and so  $\frac{dy}{d\omega} < 0$  only if  $u_{cc} + u_{cy} < 0$ , as required.

We can now analyse the impact of an increase in  $\theta$  on first-best outcomes by taking a total derivative of utility with respect to  $\theta$ , under the twin restrictions that  $u_c$  and  $u_y$  remain unchanged. With a little algebra it can be shown that these restrictions imply:

$$\begin{aligned} \frac{dc}{d\theta} &= \frac{u_{yy}u_{c\theta} - u_{cy}u_{y\theta}}{u_{cy}^2 - u_{cc}u_{yy}} \\ \frac{dy}{d\theta} &= \frac{u_{cc}u_{y\theta} - u_{cy}u_{c\theta}}{u_{cy}^2 - u_{cc}u_{yy}} \end{aligned}$$

The overall effect on utility at the margin is:

$$\begin{aligned} \frac{du}{d\theta} &= u_\theta + u_c \frac{dc}{d\theta} + u_y \frac{dy}{d\theta} \\ &= u_\theta + u_y \frac{u_{y\theta}(u_{cc} + u_{cy}) - u_{c\theta}(u_{yy} + u_{cy})}{u_{cy}^2 - u_{cc}u_{yy}} \end{aligned}$$

(where we have used that  $u_c = -u_y$  at the optimum). Negative definiteness of the partial Hessian  $\begin{bmatrix} u_{cc} & u_{cy} \\ u_{cy} & u_{yy} \end{bmatrix}$  implies  $u_{cy}^2 - u_{cc}u_{yy} < 0$ ,<sup>52</sup> and we have

---

<sup>52</sup>Consider movements in the vector  $\begin{bmatrix} -\frac{u_{yc}}{u_{cc}} \\ 1 \end{bmatrix}$  here.

$u_y < 0$ , so if  $u_{cc} + u_{cy} < 0$  then condition (2) gives:

$$\begin{aligned}\frac{du}{d\theta} &< u_\theta + u_y \frac{u_{yy} u_\theta (u_{cc} + u_{cy}) - u_{c\theta} (u_{yy} + u_{cy})}{u_{cy}^2 - u_{cc} u_{yy}} \\ &= u_\theta \left( 1 + \frac{u_{yy} (u_{cc} + u_{cy}) - u_{cy} (u_{yy} + u_{cy})}{u_{cy}^2 - u_{cc} u_{yy}} \right) \\ &= 0\end{aligned}$$

where we have additionally used condition (3). The result then follows immediately.

## A.2 Proof of Proposition 1

For the sake of clarity we index the  $N$  elements of  $\Theta$  in ascending order, so  $\theta_t^n > \theta_t^m$  whenever  $n > m$  for all  $n, m \in \{1, \dots, N\}$ . We have imposed that

$$W(\theta_t^n; \theta_t^n, \hat{\theta}^{t-1}) = W(\theta_t^{n-1}; \theta_t^n, \hat{\theta}^{t-1})$$

for all  $n \in \{2, \dots, N\}$ , and wish to show that this implies

$$W(\theta_t^n; \theta_t^n, \hat{\theta}^{t-1}) \geq W(\theta_t^m; \theta_t^n, \hat{\theta}^{t-1})$$

for all  $m \in \{1, \dots, N\}$ , given the increasing differences condition

We first consider the case in which  $n > 1$ , and show

$$W(\theta_t^n; \theta_t^n, \hat{\theta}^{t-1}) \geq W(\theta_t^m; \theta_t^n, \hat{\theta}^{t-1})$$

for all  $m \in \{1, \dots, n-1\}$ . For  $m = n-1$  this holds by assumption. For  $m = n-2$  we have by increasing differences:

$$\begin{aligned}& W(\theta_t^{n-1}; \theta_t^n, \hat{\theta}^{t-1}) - W(\theta_t^{n-1}; \theta_t^{n-1}, \hat{\theta}^{t-1}) \\ &> W(\theta_t^{n-2}; \theta_t^n, \hat{\theta}^{t-1}) - W(\theta_t^{n-2}; \theta_t^{n-1}, \hat{\theta}^{t-1})\end{aligned}$$

But

$$W(\theta_t^{n-1}; \theta_t^n, \hat{\theta}^{t-1}) = W(\theta_t^n; \theta_t^n, \hat{\theta}^{t-1})$$

and

$$W(\theta_t^{n-2}; \theta_t^{n-1}, \hat{\theta}^{t-1}) = W(\theta_t^{n-1}; \theta_t^{n-1}, \hat{\theta}^{t-1})$$

so prior inequality implies

$$W(\theta_t^n; \theta_t^n, \hat{\theta}^{t-1}) > W(\theta_t^{n-2}; \theta_t^n, \hat{\theta}^{t-1})$$

as required. Taking  $m = n - 3$ , we then have by increasing differences:

$$\begin{aligned} & W\left(\theta_t^{n-2}; \theta_t^n, \hat{\theta}^{t-1}\right) - W\left(\theta_t^{n-2}; \theta_t^{n-2}, \hat{\theta}^{t-1}\right) \\ & > W\left(\theta_t^{n-3}; \theta_t^n, \hat{\theta}^{t-1}\right) - W\left(\theta_t^{n-3}; \theta_t^{n-2}, \hat{\theta}^{t-1}\right) \end{aligned}$$

Again, by

$$W\left(\theta_t^{n-3}; \theta_t^{n-2}, \hat{\theta}^{t-1}\right) = W\left(\theta_t^{n-2}; \theta_t^{n-2}, \hat{\theta}^{t-1}\right)$$

this inequality collapses to

$$W\left(\theta_t^{n-2}; \theta_t^n, \hat{\theta}^{t-1}\right) > W\left(\theta_t^{n-3}; \theta_t^n, \hat{\theta}^{t-1}\right)$$

and we can apply the earlier result

$$W\left(\theta_t^n; \theta_t^n, \hat{\theta}^{t-1}\right) > W\left(\theta_t^{n-2}; \theta_t^n, \hat{\theta}^{t-1}\right)$$

to assert

$$W\left(\theta_t^n; \theta_t^n, \hat{\theta}^{t-1}\right) > W\left(\theta_t^{n-3}; \theta_t^n, \hat{\theta}^{t-1}\right)$$

as required. The same argument can clearly be applied for all  $m \in \{1, \dots, n-1\}$ .

When  $n < N$  we must in the same way consider the cases of  $m \in \{n+1, \dots, N\}$ . For  $m = n+1$ , we have immediately by the binding restriction on  $n+1$ -types, together with increasing differences:

$$\begin{aligned} 0 &= W\left(\theta_t^{n+1}; \theta_t^{n+1}, \hat{\theta}^{t-1}\right) - W\left(\theta_t^n; \theta_t^{n+1}, \hat{\theta}^{t-1}\right) \\ &> W\left(\theta_t^{n+1}; \theta_t^n, \hat{\theta}^{t-1}\right) - W\left(\theta_t^n; \theta_t^n, \hat{\theta}^{t-1}\right) \end{aligned}$$

as required. By similar logic, for  $m = n+2$  we have:

$$\begin{aligned} 0 &= W\left(\theta_t^{n+2}; \theta_t^{n+2}, \hat{\theta}^{t-1}\right) - W\left(\theta_t^{n+1}; \theta_t^{n+2}, \hat{\theta}^{t-1}\right) \\ &> W\left(\theta_t^{n+2}; \theta_t^n, \hat{\theta}^{t-1}\right) - W\left(\theta_t^{n+1}; \theta_t^n, \hat{\theta}^{t-1}\right) \end{aligned}$$

and the condition

$$W\left(\theta_t^n; \theta_t^n, \hat{\theta}^{t-1}\right) > W\left(\theta_t^{n+1}; \theta_t^n, \hat{\theta}^{t-1}\right)$$

then delivers the required result. Again, we can apply an identical argument inductively for all remaining  $m < N$ . This completes the proof.

### A.3 Proof of Proposition 7

We again consider a pair of perturbation schedules  $\Delta_{-1}(\delta)$  and  $\Delta(\delta)$  applied at  $t$  and  $t+1$  respectively to the allocations of agents with the relevant reporting

history  $\widehat{\theta}^t$ . We set the first  $n - 1$  rows of  $\Delta(\delta)$  to 0 for all  $\delta$ . The  $n$ th row is then constructed to equal

$$(\varphi^c(\theta_{t+1}^n, \delta; c_{t+1}^*, y_{t+1}^*), \varphi^y(\theta_{t+1}^n, \delta; c_{t+1}^*, y_{t+1}^*))$$

where these values are defined implicitly (and – by the single-crossing condition – uniquely) by the following two equations:

$$u(c + \varphi^c(\theta, k; c, y), y + \varphi^y(\theta, k; c, y); \theta) = u(c, y; \theta) \quad (68)$$

$$u(c + \varphi^c(\theta, k; c, y), y + \varphi^y(\theta, k; c, y); \theta') = u(c, y; \theta') + k \quad (69)$$

with  $\theta' = \min\{\theta'' \in \Theta : \theta'' > \theta\}$ . (So  $\varphi^c(\theta, k; c, y)$  and  $\varphi^y(\theta, k; c, y)$  are perturbations from the allocation  $(c, y)$  that keep utility constant for an agent of type  $\theta$ , whilst increasing it by  $k$  units for an agent whose type is one higher.) The  $m$ th row in  $\Delta(\delta)$  is given for  $m \in \{n + 1, \dots, N\}$  by

$$(\phi^c(\theta_{t+1}^m, \delta; c_{t+1}^*, y_{t+1}^*), \phi^y(\theta_{t+1}^m, \delta; c_{t+1}^*, y_{t+1}^*))$$

where these functions are defined in the proof of Proposition 4. We additionally assume a time- $t$  perturbation schedule  $\Delta_{-1}(\delta)$  given by:

$$\begin{aligned} & \Delta_{-1}(\delta) \\ = & (\phi^c(\theta_t, -\beta\delta\pi_\Theta(\theta_{t+1} > \theta_{t+1}^n | \theta_t); c_t^*, y_t^*), \phi^y(\theta_t, -\beta\delta\pi_\Theta(\theta_{t+1} > \theta_{t+1}^n | \theta_t); c_t^*, y_t^*)) \end{aligned} \quad (70)$$

These perturbations will preserve incentive compatibility (according to the relaxed problem's constraint set) at  $t + 1$ . For the  $n$ th agent this holds because the perturbation is constructed so as not to affect his or her utility from truth-telling, whilst utility from mimicking the  $n - 1$ th agent is held constant by the fact agents below the  $n$ th see no change to their allocations. For agents whose types are higher than the  $n$ th, the  $\Delta(\delta)$  schedule is constructed to ensure there are equal utility gains to mimicking and truth-telling, so that 'downwards' incentive compatibility constraints cannot be violated through these perturbations, whilst we are free to ignore other constraints under the supposition that the general problem's optimum cannot be improved upon for any allocation that satisfies the relaxed problem's constraint set.

The perturbations will also preserve incentive compatibility under the relaxed problem in the earlier period  $t$ . Given the iid assumption, the discounted value of the  $\Delta(\delta)$  perturbation at time  $t$  is  $\beta\delta\pi_\Theta(\theta_{t+1} > \theta_{t+1}^n | \theta_t)$  to *both* the agent of type  $\theta_t$  and the agent whose type is one higher and chooses to mimic (that is,  $\delta$  units of extra utility received if and only if one's type exceeds  $\theta_{t+1}^n$ ). By perturbing the  $t$ -dated utility received by both truth-teller and mimicker by  $-\beta\delta\pi_\Theta(\theta_{t+1} > \theta_{t+1}^n | \theta_t)$  units, we ensure the impact on the net present value of utility is zero for both. Hence truth-telling will remain an optimal strategy for both.

The present value (from the perspective of time  $t$ ) of the cost to the policy-

maker of applying the  $\Delta_{-1}(\delta)$  and  $\Delta(\delta)$  perturbations is:

$$\begin{aligned} & R_t^{-1} \left\{ \pi_{\Theta}(\theta_{t+1}^n | \theta_t) [\varphi^c(\theta_{t+1}^n, \delta) - \varphi^y(\theta_{t+1}^n, \delta)] \right. \\ & + \sum_{m=n+1}^N \pi_{\Theta}(\theta_{t+1}^m | \theta_t) [\phi^c(\theta_{t+1}^m, \delta) - \phi^y(\theta_{t+1}^m, \delta)] \left. \right\} \\ & + \phi^c(\theta_t, -\beta\delta\pi_{\Theta}(\theta_{t+1} > \theta_{t+1}^n | \theta_t)) - \phi^y(\theta_t, -\beta\delta\pi_{\Theta}(\theta_{t+1} > \theta_{t+1}^n | \theta_t)) \end{aligned}$$

(where we have suppressed dependence upon equilibrium allocations in the  $\phi$  and  $\varphi$  functions to ease the notation). The perturbations have zero net impact on the present value of utility assessed from the perspective of the initial time period, so a necessary condition for optimality is that the derivative of this surplus with respect to  $\delta$  should equal zero when  $\delta = 0$ :

$$\begin{aligned} & R_t^{-1} \left\{ \pi_{\Theta}(\theta_{t+1}^n | \theta_t) [\varphi_2^c(\theta_{t+1}^n, 0) - \varphi_2^y(\theta_{t+1}^n, 0)] \right. \\ & + \sum_{m=n+1}^N \pi_{\Theta}(\theta_{t+1}^m | \theta_t) [\phi_2^c(\theta_{t+1}^m, 0) - \phi_2^y(\theta_{t+1}^m, 0)] \left. \right\} \\ & - \beta\pi_{\Theta}(\theta_{t+1} > \theta_{t+1}^n | \theta_t) [\phi_2^c(\theta_t, 0) - \phi_2^y(\theta_t, 0)] \\ & = 0 \end{aligned} \quad (71)$$

Again, differentiating the equations defining  $\phi$  we can show:

$$\phi_2^c(\theta, 0) - \phi_2^y(\theta, 0) = \frac{1 - \alpha(\theta)}{u_c(\theta) + u_y(\theta)\alpha(\theta)} \quad (72)$$

and total differentiation of (68) and (69) gives:

$$\begin{aligned} \phi_2^c(\theta, 0) - \phi_2^y(\theta, 0) &= \frac{1 + \frac{u_c(\theta)}{u_y(\theta)}}{u_c(\widehat{\theta}; \theta') - u_y(\widehat{\theta}; \theta') \frac{u_c(\theta)}{u_y(\theta)}} \\ &= -\frac{\tau(\theta)}{u_c(\widehat{\theta}; \theta')(1 - \tau(\theta)) + u_y(\widehat{\theta}; \theta')} \end{aligned} \quad (73)$$

where  $\theta' = \min\{\theta'' \in \Theta : \theta'' > \theta\}$ . Using these results in the optimality condition we have:

$$\begin{aligned} & -\pi_{\Theta}(\theta_{t+1}^n | \theta_t) \frac{\tau(\theta_{t+1}^n)}{u_c(\widehat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})(1 - \tau(\theta_{t+1}^n)) + u_y(\widehat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})} \\ & + \sum_{m=n+1}^N \pi_{\Theta}(\theta_{t+1}^m | \theta_t) \frac{1 - \alpha(\theta_{t+1}^m)}{u_c(\theta_{t+1}^m) + u_y(\theta_{t+1}^m)\alpha(\theta_{t+1}^m)} \\ & = \beta R_{t+1} \pi_{\Theta}(\theta_{t+1} > \theta_{t+1}^n | \theta_t) \frac{1 - \alpha(\theta_t)}{u_c(\theta_t) + u_y(\theta_t)\alpha(\theta_t)} \end{aligned} \quad (74)$$

QED.

#### A.4 Proof of Lemma 8

Our focus is restricted to remaining within the ‘relaxed’ constraint set, so we need only show that it is possible to change the consumption and output levels of each agent in such a way that utilities change in the manner described in the Proposition, and ‘downwards’ incentive compatibility restrictions remain satisfied for all  $\delta$  in an open neighbourhood of 0. This requires that the following two conditions are satisfied at  $t + 1$  for all  $n \in \{1, \dots, N\}$ :

$$u(c_{n,t+1}^* + \delta_n^c(\delta), y_{n,t+1}^* + \delta_n^y(\delta); \theta_{t+1}^n) = u(c_{n,t+1}^*, y_{n,t+1}^*; \theta_{t+1}^n) + \nu_n \delta \quad (75)$$

$$u(c_{n,t+1}^* + \delta_n^c(\delta), y_{n,t+1}^* + \delta_n^y(\delta); \theta_{t+1}^{n+1}) = u(c_{n,t+1}^*, y_{n,t+1}^*; \theta_{t+1}^{n+1}) + \nu_{n+1} \delta \quad (76)$$

where  $\delta_n^c(\delta)$  and  $\delta_n^y(\delta)$  are the perturbations to the  $n$ th agent’s consumption and output levels respectively. For the  $N$ th agent we just need:

$$u(c_{N,t+1}^* + \delta_N^c(\delta), y_{N,t+1}^*; \theta_{t+1}^N) = u(c_{N,t+1}^*, y_{N,t+1}^*; \theta_{t+1}^N) + \nu_N \delta \quad (77)$$

and we normalise  $\delta_N^y(\delta) = 0$ .<sup>53</sup>

Equations (75) and (77) here are just stating that the truth-telling agent should be moved onto a within-period indifference curve consistent with the perturbed utility level obtaining, whilst condition (76) states that the specific perturbed allocation should be at a point on this indifference curve such that the change in the utility of a mimicking higher-type agent is equal to the change in that higher-type agent’s truth-telling utility. By the single-crossing condition higher-type agents see their utility change monotonically through movements along the indifference curve of a lower-type agent, so for small enough  $\delta$  these equations must solve for unique values of  $\delta_n^c(\delta)$  and  $\delta_n^y(\delta)$  for all  $n$ . (The limit on the magnitude of  $\delta$  comes from the fact that there is a minimum level of utility a mimicking agent must obtain along a given lower-type agent’s indifference curve.) These values will preserve incentive compatibility at  $t + 1$ . The impact on discounted expected utility from the perspective of time  $t$  for an agent who has reported  $\hat{\theta}^t$  is to increase it by an amount  $\beta \nu' \pi_{\Theta}^{vec} \delta$ . If  $\nu' \pi_{\Theta}^{vec} = 0$  then we are done (confirming the last statement in the Proposition). Otherwise, to preserve ‘downward’ incentive compatibility at  $t$  (and earlier) we must reduce within-period utility in that period by an equal amount, in a manner that has an equal impact on the agent whose true type is  $\theta_t$  and a mimicker whose type is one higher. This can be done through the perturbation

$$[\phi^c(\theta_t, -\beta \nu' \pi_{\Theta}^{vec} \delta), \phi^y(\theta_t, -\beta \nu' \pi_{\Theta}^{vec} \delta)]$$

as already established.

---

<sup>53</sup>This is analogous to the normalisation  $\phi^y(\theta, k; c^*, y^*) = 0$  in equation (25).

## A.5 Proof of Lemma 9

Note first from the definition of  $\gamma$  that  $\gamma_n = \nu_n - \nu_{n-1}$  for all  $n \in \{1, \dots, N\}$ , where we define  $\nu_0 = 0$ . The marginal resource cost at  $t+1$  of the perturbation set out in Proposition 8 is:

$$\pi_{\Theta}(\theta^t) \sum_{n=1}^N \pi_{\Theta}(\theta_{t+1}^n | \theta_t) \left( \frac{d\delta_n^c(0)}{d\delta} - \frac{d\delta_n^y(0)}{d\delta} \right) \quad (78)$$

where the  $\delta_n^c(\delta)$  and  $\delta_n^y(\delta)$  functions satisfy restrictions (75) to (77), and  $\delta_N^y(\delta)$  is again set to zero. Totally differentiating those restrictions, one can show:

$$\frac{d\delta_n^c(0)}{d\delta} = \frac{\frac{u_y(\theta_{t+1}^n)}{u_c(\theta_{t+1}^n)} \frac{\nu_{n+1}}{u_c(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})} - \frac{u_y(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})}{u_c(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})} \frac{\nu_n}{u_c(\theta_{t+1}^n)}}{\frac{u_y(\theta_{t+1}^n)}{u_c(\theta_{t+1}^n)} - \frac{u_y(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})}{u_c(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})}} \quad (79)$$

$$\frac{d\delta_n^y(0)}{d\delta} = \frac{\frac{\nu_n}{u_c(\theta_{t+1}^n)} - \frac{\nu_{n+1}}{u_c(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})}}{\frac{u_y(\theta_{t+1}^n)}{u_c(\theta_{t+1}^n)} - \frac{u_y(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})}{u_c(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})}} \quad (80)$$

for  $n \in \{1, \dots, N-1\}$ , and

$$\frac{d\delta_N^c(0)}{d\delta} = \frac{\nu_N}{u_c(\theta_{t+1}^N)} \quad (81)$$

With some manipulation it is then possible to show for all  $n \in \{1, \dots, N-1\}$ :

$$\begin{aligned} \frac{d\delta_n^c(0)}{d\delta} - \frac{d\delta_n^y(0)}{d\delta} &= \nu_n \frac{1 - \alpha(\theta_{t+1}^n)}{u_c(\theta_{t+1}^n) + \alpha(\theta_{t+1}^n) u_y(\theta_{t+1}^n)} \\ &\quad - (\nu_{n+1} - \nu_n) \frac{\tau(\theta_{t+1}^n)}{u_c(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1}) (1 - \tau(\theta_{t+1}^n)) + u_y(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})} \\ &= \nu_n (\phi_2^c(\theta_{t+1}^n, 0) - \phi_2^y(\theta_{t+1}^n, 0)) \\ &\quad - (\nu_{n+1} - \nu_n) (\varphi_2^c(\theta_{t+1}^n, 0) - \varphi_2^y(\theta_{t+1}^n, 0)) \end{aligned} \quad (82)$$

and

$$\begin{aligned} \frac{d\delta_N^c(0)}{d\delta} - \frac{d\delta_N^y(0)}{d\delta} &= \frac{d\delta_N^c(0)}{d\delta} \\ &= \frac{\nu_N}{u_c(\theta_{t+1}^N)} \\ &= \nu_N (\phi_2^c(\theta_{t+1}^N, 0) - \phi_2^y(\theta_{t+1}^N, 0)) \end{aligned} \quad (83)$$



where we apply the earlier definitions of the  $\phi^c$ ,  $\phi^y$ ,  $\varphi^c$  and  $\varphi^y$  functions. Hence:

$$\begin{aligned}
& \sum_{n=1}^N \pi_{\Theta}(\theta_{t+1}^n | \theta_t) \left( \frac{d\delta_n^c(0)}{d\delta} - \frac{d\delta_n^y(0)}{d\delta} \right) \\
&= \sum_{n=1}^N \pi_{\Theta}(\theta_{t+1}^n | \theta_t) \left[ \nu_n (\phi_2^c(\theta_{t+1}^n, 0) - \phi_2^y(\theta_{t+1}^n, 0)) \right. \\
&\quad \left. - (\nu_{n+1} - \nu_n) (\varphi_2^c(\theta_{t+1}^n, 0) - \varphi_2^y(\theta_{t+1}^n, 0)) \right] \\
&= \sum_{n=1}^{N-1} \pi_{\Theta}(\theta_{t+1}^n | \theta_t) \left[ \sum_{m=1}^n \gamma_m (\phi_2^c(\theta_{t+1}^n, 0) - \phi_2^y(\theta_{t+1}^n, 0)) \right. \\
&\quad \left. - \gamma_{n+1} (\varphi_2^c(\theta_{t+1}^n, 0) - \varphi_2^y(\theta_{t+1}^n, 0)) \right] \\
&\quad + \pi_{\Theta}(\theta_{t+1}^N | \theta_t) \sum_{m=1}^N \gamma_m (\phi_2^c(\theta_{t+1}^N, 0) - \phi_2^y(\theta_{t+1}^N, 0))
\end{aligned} \tag{84}$$

This last expression can equivalently be written in matrix form:

$$(\pi_{\Theta}^{vec})' \left[ \sum_{n=1}^N \gamma_n \Delta^{n'}(0) \right] k \tag{85}$$

The first part of the result follows. For the second we need to show additionally that it is not possible to move in any dimension *not* described by a  $\Delta^{n'}(0)$  matrix and preserve incentive compatibility for the relaxed problem. But the only degree of freedom we have to vary the above changes is in relaxing the normalisation that  $\delta_N^y(\delta) = 0$ , whilst still satisfying

$$u(c_{N,t+1}^* + \delta_N^c(\delta), y_{N,t+1}^* + \delta_N^y(\delta); \theta_{t+1}^N) = u(c_{N,t+1}^*, y_{N,t+1}^*; \theta_{t+1}^N) + \nu_N \delta \tag{86}$$

This gives:

$$u_c(\theta_{t+1}^N) \frac{d\delta_N^c(0)}{d\delta} + u_y(\theta_{t+1}^N) \frac{d\delta_N^y(0)}{d\delta} = \nu_N \tag{87}$$

But at the top  $u_c(\theta_{t+1}^N) = -u_y(\theta_{t+1}^N)$ , so:

$$\begin{aligned}
\frac{d\delta_N^c(0)}{d\delta} - \frac{d\delta_N^y(0)}{d\delta} &= \frac{\nu_N}{u_c(\theta_{t+1}^N)} \\
&= \nu_N (\phi_2^c(\theta_{t+1}^N, 0) - \phi_2^y(\theta_{t+1}^N, 0)) \\
&= \nu_N (\phi_2^c(\theta_{t+1}^N, 0) - \phi_2^y(\theta_{t+1}^N, 0)) \\
&\quad + \kappa (\varphi_2^c(\theta_{t+1}^N, 0) - \varphi_2^y(\theta_{t+1}^N, 0))
\end{aligned} \tag{88}$$

for arbitrary  $\kappa$ . So if  $\frac{\delta_N^y(0)}{d\delta}$  differs from zero then  $\frac{\delta_N^c(0)}{d\delta}$  can differ from  $\frac{\nu_N}{u_c(\theta_{t+1}^N)}$ , but only in a manner that raises no net resources – which is equivalent to a movement along the ‘top’ indifference curve, given that the optimum involves no distortion at the top. So the only additional dimension in which outcomes can be perturbed at the margin is that described by  $\Delta^{N+1'}(0)$ .

## A.6 Proof of Lemma 15

Again, the Lemma requires us to focus only on the need to ensure ‘downwards’ incentive compatibility continues to hold locally at  $t$  and  $t + 1$ . The latter is simpler: it requires that the following conditions are satisfied for agents with the relevant reporting history for all  $m \in \{1, \dots, N\}$ :

$$\begin{aligned} u(c_{m,t+1}^* + \delta_{m,t+1}^c(\delta), y_{m,t+1}^* + \delta_{m,t+1}^y(\delta); \theta_{t+1}^m) &= u(c_{m,t+1}^*, y_{m,t+1}^*; \theta_{t+1}^m) + \nu_m \delta \\ u(c_{m,t+1}^* + \delta_{m,t+1}^c(\delta), y_{m,t+1}^* + \delta_{m,t+1}^y(\delta); \theta_{t+1}^{m+1}) &= u(c_{m,t+1}^*, y_{m,t+1}^*; \theta_{t+1}^{m+1}) + \nu_{m+1} \delta \end{aligned} \quad \begin{matrix} (89) \\ (90) \end{matrix}$$

where  $\delta_{m,t+1}^c(\delta)$  and  $\delta_{m,t+1}^y(\delta)$  are the perturbations to the  $m$ th agent’s consumption and output levels respectively. For the  $N$ th agent we just need:

$$u(c_{N,t+1}^* + \delta_{N,t+1}^c(\delta), y_{N,t+1}^*; \theta_{t+1}^N) = u(c_{N,t+1}^*, y_{N,t+1}^*; \theta_{t+1}^N) + \nu_N \delta \quad (91)$$

and we normalise  $\delta_{N,t+1}^y(\delta) = 0$ .

The proof of Proposition 8 shows that these conditions can indeed be satisfied by appropriate choice of  $\delta_{m,t+1}^c(\delta)$  and  $\delta_{m,t+1}^y(\delta)$  schedules. There remains the problem of incentive compatibility (under the relaxed problem) at  $t$ . From the perspective of that time period the  $t + 1$  perturbations are increasing expected utility for potential mimickers by  $\beta\delta$  units, whilst leaving that of truth-tellers constant. To offset this effect we need to move along the indifference curve of the  $n$ th agent at  $t$  to such an extent that a mimicker’s utility is reduced by an offsetting amount (whilst, by definition, leaving the utility of a truth-teller unaffected in this period also). That requires  $\delta_{n,t}^c(\delta)$  and  $\delta_{n,t}^y(\delta)$  schedules that satisfy:

$$u(c_{n,t}^* + \delta_{n,t}^c(\delta), y_{n,t}^* + \delta_{n,t}^y(\delta); \theta_t^n) = u(c_{n,t}^*, y_{n,t}^*; \theta_t^n) \quad (92)$$

$$u(c_{n,t}^* + \delta_{n,t}^c(\delta), y_{n,t}^* + \delta_{n,t}^y(\delta); \theta_t^{n+1}) = u(c_{n,t}^*, y_{n,t}^*; \theta_t^{n+1}) - \beta\delta \quad (93)$$

Again, by the single crossing condition the utility of the agent of type  $\theta_t^{n+1}$  changes monotonically as one moves along a lower-type agent’s indifference curve, so for small enough  $\delta$  in an open neighbourhood of  $\delta = 0$  this is always possible – with a limit provided by the fact that there is a minimum to the utility that mimickers can obtain on the given lower-type indifference curve.

## A.7 Proof of Lemma 18

It remains to establish the result for the case in which consumption and labour supply are Edgeworth complements (in which case  $\alpha(\theta_t) > 0$ ) and productivities follow an iid process. In order to put a zero lower bound on the marginal cost of utility provision in this case we need to verify that  $\alpha(\theta_t) < 1$  – that is, that the marginal cost of incentive-compatible utility provision never turns negative under an optimal plan. Suppose instead that  $\alpha(\theta_t) \geq 1$  were to hold for some  $\theta_t$  and a given report history. We argue that in this situation it is always possible for the policymaker to generate surplus resources at the margin, whilst preserving incentive compatibility – contradicting optimality.

If  $\alpha(\theta_t) \geq 1$  then the generalised inverse Euler equation implies we must also have:

$$\sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1} | \theta_t) \left[ \frac{1 - \alpha(\theta_{t+1})}{u_c(\theta_{t+1}) + u_y(\theta_{t+1}) \alpha(\theta_{t+1})} \right] \geq 1 \quad (94)$$

With iid shocks it is always possible for us to find an  $(N + 1) \times 1$  vector  $\gamma$  that satisfies the following:

$$\left[ \sum_{n=1}^{N+1} \gamma_n \Delta^{n'}(0) \right] = \tilde{\Delta} \quad (95)$$

for some  $N \times 2$   $\tilde{\Delta}$  matrix whose first column is an  $N \times 1$  vector of strictly positive scalars and whose second is an  $N \times 1$  vector of zeros. To see this, note that the marginal vector  $\Delta^{n'}(0)$  has zeros in all rows up to the  $(n - 2)$ th, and an entry in the  $(n - 1)$ th row that is linearly independent of the  $(n - 1)$ th row in all of the other  $\Delta^{m'}(0)$  matrices. So we can choose the entries of  $\gamma$  by first finding an additive combination of  $\Delta^{1'}(0)$  and  $\Delta^{2'}(0)$  sufficient to increase the consumption of the agent of type  $\theta_{t+1}^1$  without changing that agent's output requirements, then add to this sufficient units of  $\Delta^{3'}(0)$  for the output level of the agent of type  $\theta_{t+1}^2$  to remain unchanged (this perturbation has no impact on type  $\theta_{t+1}^1$ ), then sufficient units of  $\Delta^{4'}(0)$  for the output level of the agent of type  $\theta_{t+1}^3$  to remain unchanged, and so on. Since consumption is increasing for the agent of type  $\theta_{t+1}^1$ , it must also increase for all higher-type agents – otherwise we could not continue to satisfy incentive compatibility. This  $\tilde{\Delta}$  perturbation increases the *ex ante* expected utility level of an agent with the relevant reporting history, so it will be possible to leave that expected utility level unchanged (preserving incentive compatibility at  $t$ ) by a linear combination of  $\Delta^{n'}(0)$  and  $\tilde{\Delta}$ , with a positive coefficient on the former and a negative on the latter. But if  $\tilde{\Delta}$  is providing utility through consumption increments alone it must come at a positive cost, whilst we have from (94) that the cost of  $\Delta^{n'}(0)$ . Hence we raise a surplus, contradicting optimality.<sup>54</sup>

<sup>54</sup> Notice that this argument cannot be applied in the case of non-iid productivity processes, since the perturbation operating on consumption levels alone does not generate a uniform level of incremental utility provision across  $t + 1$  types, and thus will have differential effects

## A.8 Proof of Proposition 19

We know Doob's convergence theorem applies to the non-negative martingale  $\frac{1-\alpha(\theta_t)}{u_c(\theta_t)+u_y(\theta_t)\alpha(\theta_t)}$ , so need only show that it is not possible for this object to converge to any non-zero value. The following Lemma is useful:

**Lemma 21**  $\frac{\tau(\theta_t^n)}{u_c(\hat{\theta}_t^n; \theta_t^{n+1})(1-\tau(\theta_t^n))+u_y(\hat{\theta}_t^n; \theta_t^{n+1})} \xrightarrow{a.s.} 0$  holds under an optimal plan that solves the restricted problem.

**Proof.** In the iid case this follows directly from equation (41):

$$\begin{aligned}
& \lim_{t \rightarrow \infty} \left[ -\pi_{\Theta}(\theta_{t+1}^n | \theta_t) \frac{\tau(\theta_{t+1}^n)}{u_c(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})(1-\tau(\theta_{t+1}^n))+u_y(\hat{\theta}_{t+1}^n; \theta_{t+1}^{n+1})} \right] \quad (96) \\
&= - \sum_{m=n+1}^N \pi_{\Theta}(\theta_{t+1}^m | \theta_t) \lim_{t \rightarrow \infty} \left[ \frac{1-\alpha(\theta_{t+1}^m)}{u_c(\theta_{t+1}^m)+u_y(\theta_{t+1}^m)\alpha(\theta_{t+1}^m)} \right] \\
&\quad + \pi_{\Theta}(\theta_{t+1} > \theta_{t+1}^n | \theta_t) \sum_{\theta_{t+1} \in \Theta} \pi_{\Theta}(\theta_{t+1} | \theta_t) \lim_{t \rightarrow \infty} \left[ \frac{1-\alpha(\theta_{t+1})}{u_c(\theta_{t+1})+u_y(\theta_{t+1})\alpha(\theta_{t+1})} \right] \\
&= 0
\end{aligned}$$

In the Markov case we know that equation (41) must hold in periods immediately following those in which  $\theta = \theta^N$ , and so if one indexes by  $T$  the (infinite) set of periods in which this is the case, and denotes by  $t(T)$  the (conventional) time period corresponding to the  $T$ th occasion on which  $\theta = \theta^N$  has obtained along

---

on the expected utility levels of mimickers and truth-tellers at time  $t$ . One could generalise to consider the complete set of perturbations guaranteed to increase utility for all agents at  $t+1$  at a positive cost – that is, movements that increase consumption by a greater amount than output at the margin for all agents, whilst simultaneously increasing utility. But even this set is not sufficiently large to prove the result: the movements it permits do not span an entire half-space in the  $(c_{t+1}, y_{t+1})$  plane for each agent.

the given sample path, we must have:

$$\begin{aligned}
& \lim_{T \rightarrow \infty} \left[ -\pi_{\Theta} \left( \theta_{t(T)+1}^n | \theta_{t(T)} \right) \right. \\
& \quad \left. \cdot \frac{\tau \left( \theta_{t(T)+1}^n \right)}{u_c \left( \hat{\theta}_{t(T)+1}^n; \theta_{t(T)+1}^{n+1} \right) \left( 1 - \tau \left( \theta_{t(T)+1}^n \right) \right) + u_y \left( \hat{\theta}_{t(T)+1}^n; \theta_{t(T)+1}^{n+1} \right)} \right] \\
&= - \sum_{m=n+1}^N \pi_{\Theta} \left( \theta_{t(T)+1}^m | \theta_{t(T)} \right) \lim_{T \rightarrow \infty} \left[ \frac{1 - \alpha \left( \theta_{t(T)+1}^m \right)}{u_c \left( \theta_{t(T)+1}^m \right) + u_y \left( \theta_{t(T)+1}^m \right) \alpha \left( \theta_{t(T)+1}^m \right)} \right] \\
& \quad + \pi_{\Theta} \left( \theta_{t(T)+1} > \theta_{t(T)+1}^n | \theta_{t(T)} \right) \\
& \quad \cdot \sum_{\theta_{t(T)+1} \in \Theta} \pi_{\Theta} \left( \theta_{t(T)+1} | \theta_{t(T)} \right) \lim_{T \rightarrow \infty} \left[ \frac{1 - \alpha \left( \theta_{t(T)+1} \right)}{u_c \left( \theta_{t(T)+1} \right) + u_y \left( \theta_{t(T)+1} \right) \alpha \left( \theta_{t(T)+1} \right)} \right] \\
&= 0
\end{aligned} \tag{97}$$

But if

$$\frac{\tau \left( \theta_{t(T)+1}^n \right)}{u_c \left( \hat{\theta}_{t(T)+1}^n; \theta_{t(T)+1}^{n+1} \right) \left( 1 - \tau \left( \theta_{t(T)+1}^n \right) \right) + u_y \left( \hat{\theta}_{t(T)+1}^n; \theta_{t(T)+1}^{n+1} \right)} = 0$$

holds at the limit as  $T$  becomes large then we must also, at the same limit, have an identical set of zero restrictions in period  $t(T) + 2$ , by equations (61) and (60). By induction this can then be extended to period  $t(T) + n$  for all  $n > 1$ , and the result follows. ■

This Lemma implies two alternatives: either

$$\tau \left( \theta_t^n \right) \xrightarrow{a.s.} 0$$

or

$$u_c \left( \hat{\theta}_t^n; \theta_t^{n+1} \right) \left( 1 - \tau \left( \theta_t^n \right) \right) + u_y \left( \hat{\theta}_t^n; \theta_t^{n+1} \right) \xrightarrow{a.s.} \infty$$

Suppose the latter were true. By equation (65) we have:

$$\begin{aligned}
u_c \left( \theta_t^n \right) + u_y \left( \theta_t^n \right) \alpha \left( \theta_t^n \right) &= \frac{u_c \left( \hat{\theta}_t^n; \theta_t^{n+1} \right) - u_y \left( \hat{\theta}_t^n; \theta_t^{n+1} \right) \frac{u_c \left( \theta_t^n \right)}{u_y \left( \theta_t^n \right)}}{1 - \frac{u_y \left( \hat{\theta}_t^n; \theta_t^{n+1} \right)}{u_y \left( \theta_t^n \right)}} \\
&> u_c \left( \hat{\theta}_t^n; \theta_t^{n+1} \right) - u_y \left( \hat{\theta}_t^n; \theta_t^{n+1} \right) \frac{u_c \left( \theta_t^n \right)}{u_y \left( \theta_t^n \right)} \\
&= u_c \left( \hat{\theta}_t^n; \theta_t^{n+1} \right) - u_y \left( \hat{\theta}_t^n; \theta_t^{n+1} \right) \frac{1}{(1 - \tau \left( \theta_t^n \right))}
\end{aligned} \tag{98}$$

If

$$u_c(\hat{\theta}_t^n; \theta_t^{n+1})(1 - \tau(\theta_t^n)) + u_y(\hat{\theta}_t^n; \theta_t^{n+1}) \xrightarrow{a.s.} \infty$$

then

$$u_c(\hat{\theta}_t^n; \theta_t^{n+1}) - u_y(\hat{\theta}_t^n; \theta_t^{n+1}) \frac{1}{(1 - \tau(\theta_t^n))} \xrightarrow{a.s.} \infty$$

must also hold, since  $(1 - \tau(\theta_t^n)) \in [0, 1]$  follows from the definition of  $\tau$  and Proposition 11. Hence we must also have

$$u_c(\theta_t^n) + u_y(\theta_t^n) \alpha(\theta_t^n) \xrightarrow{a.s.} \infty$$

This in turn implies  $\frac{1 - \alpha(\theta_t^n)}{u_c(\theta_t^n) + u_y(\theta_t^n) \alpha(\theta_t^n)}$  can only converge to a non-zero limit if  $|\alpha(\theta_t)|$  is itself always infinite at that limit. But since we know  $\alpha(\theta_t) = 0$  when  $\theta_t = \theta^N$  we can rule that out.

The alternative is that  $\tau(\theta_t^n) \xrightarrow{a.s.} 0$ . In this case we have  $u_c(\theta_t^n) = -u_y(\theta_t^n)$  at the limit, and so

$$\frac{1 - \alpha(\theta_t^n)}{u_c(\theta_t^n) + u_y(\theta_t^n) \alpha(\theta_t^n)} = \frac{1}{u_c(\theta_t^n)}$$

Hence the inverse of the marginal utility of consumption must be converging to a common value for all agents. But since  $u_c(\theta_t^n) = -u_y(\theta_t^n)$  the marginal disutility of production must also be converging to the *same* value across agents. Suppose this were a finite value. We have shown when analysing the first-best that if  $u_c$  is common across types and  $u_c = -u_y$  holds then utility must be *decreasing* in type. This is clearly inconsistent with incentive compatibility, which is enough to rule out  $\frac{1 - \alpha(\theta_t^n)}{u_c(\theta_t^n) + u_y(\theta_t^n) \alpha(\theta_t^n)}$  converging to a non-zero value in this case too. This completes the proof.