Applying perturbation analysis to dynamic optimal tax problems

Charles Brendon
University of Oxford

October 12, 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 incentive compatibility conditions are locally binding. 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.

1 Introduction

There is a growing interest among macroeconomists in methods for solving dynamic optimal policy problems in the presence of asymmetric information, including in particular multi-period optimal tax problems in the spirit of Mirrlees (1971) and Diamond and Mirrlees (1978). To date, two major approaches to these problems have emerged in the 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.

The second 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 augment 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 delivers them an expected lifetime utility consistent with incentive compatibility. This method has recently been applied to optimal dynamic tax policy by Sleet and Yeltekin (2010b) following the extension of earlier theory to settings with private information by the same authors (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 essental analytical character of the underlying solution from being completely clear. The aim of this paper is to show how a solution to one particular type of dynamic contracting problem – a Mirrleesian dynamic optimal income tax problem – can be obtained through the *analytical* derivation of a complete set of conditions necessary for an interior optimum. In this sense its relationship with the rest of the literature is similar to the relationship between the analysis of a simple consumption-savings problem through Euler equations and an alternative approach using dynamic programming. The former is based on the principle that if a plan is optimal it cannot be possible to perturb outcomes away from the optimum in a manner consistent with all constraints remaining satisfied, and obtain a resource surplus in the process. When combined with further binding restrictions – in the consumption-savings problem, a full intertemporal budget constraint – the associated first-order conditions are sufficient to characterise any interior solution. This logic has already been applied 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) through an 'inverse Euler condition', 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.

The key requirement in obtaining a result of this form is to define a perturbation to the optimal allocation that simultaneously satisfies two conditions: reversibility and local incentive compatibility. The first 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 Δ, 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 be at a corner solution if we are to state the consumption Euler equation with equality. The second requirement 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 would induce. Satisfying these two requirements for a broader set of perturbations than the intertemporal utility reallocations used in 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.

Through perturbation analysis we are able to show that the optimal allocation in a dynamic Mirrleesian economy with iid idiosyncratic shocks is characterised by a single *inter*temporal optimality condition, coupled with N *intra*temporal optimality conditions, where N is the cardinality of the set of possible productivity draws.[1] Together with binding incentive compatibility and resource constraints, this will in general be sufficient to characterise the optimum. The intertemporal optimality condition is the inverse Euler condition in the event that preferences are additively separable between consumption and labour. In the event that preferences are non-separable, we show how a natural generalisation of the inverse Euler condition can be obtained. Since the work of Thomas and Worrall (1990) 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 policy-

[1] We consider a model with a discrete number of possible shock realisations.

maker's discount factor.[2] We show that the long-run 'immiseration' results that are impled by the inverse Euler condition under usual Inada conditions extend only partially to the case of non-separable preferences. Specifically, if consumption and labour supply are (Edgeworth) substitutes then long-run immiseration in some form will obtain for all agents with probability one along the optimal path, though we cannot rule out the possibility that this occurs through the policymaker demanding an unbearably large quantity of output from agents rather than (or even alongside) consumption deprivation. But if consumption and labour supply are Edgeworth complements then we *additionally* cannot rule out the possibility of a more benign long-run outcome that is not associated with immiseration in any form. The asymmetry between the two cases derives from the possibility that the marginal cost of incentive-compatible utility provision may be zero even for non-extreme values of consumption and output in the case of complements, for reasons set out below.

In the more realistic case that idiosyncratic shocks are non-iid, we show that one extra intertemporal condition must be added to the inverse Euler equation. This condition ensures that there is no way to reduce the present value of the resource cost associated with preventing higher-productivity agents from mimicking lower-productivity types. Simultaneously, the number of intratemporal optimality conditions is reduced to $N-1$. The key intuition here is that in the non-iid case perturbations to outcomes across types in some time period $t+1$ will affect expected utility at time t differently depending on the probability measure applied − with this measure differing between mimickers and truth-tellers. Nonetheless, we show how to construct the $N-1$ remaining intratemporal conditions in a manner that ensures equivalent expected utility gains for both mimickers and truth-tellers, and it is of some interest that we lose the possibility of perturbations in one dimension only when doing so. Since the computational burden associated with solving models of this form via recursive methods − such as those proposed by Fernandes and Phelan (2000) − is substantial, it is hoped that analytical results regarding the solution's character will be of some use, and may suggest more effective computational methods.

2 Model setup

The basic framework that we use essentially follows the recent textbook treatment of Kocherlakota (2011), except that we allow for a more general specification of preferences. 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

[2] 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.

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:

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

where $u : \mathbb{R}_+^3 \to \mathbb{R}$. 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 θ_{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 finite set $\Theta \subset \mathbb{R}$, which is time-invariant. Expectations are taken across all stochastic variables relevant to the equilibrium evolution of the agent's utility (ultimately, drawings from Θ 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 Θ^∞, say θ^∞, according to some probability measure on Θ^∞, π_Θ. 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 θ^t is a t-length truncation of the vector θ^∞. 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.

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 θ. We additionally impose Inada conditions: $\lim_{c\to\infty} u_c\left(c, y; \theta\right) = 0$ and $\lim_{c\to 0} u_c\left(c, y; \theta\right) = \infty$ for all non-zero, finite (y, θ) pairs, and $\lim_{y\to\infty} u_y\left(c, y; \theta\right) = -\infty$ and $\lim_{y\to 0} u_y\left(c, y; \theta\right) = 0$ for all non-zero, finite (c, θ) pairs. The purpose of these conditions is to ensure an interior solution will obtain 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 θ 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:

$$u_{y\theta} - u_{yy}\frac{u_\theta}{u_y} > 0 \tag{2}$$

Similarly, if consumption and utility are constant then so too should be the marginal utility of consumption, implying:

$$u_{c\theta} - u_{cy}\frac{u_\theta}{u_y} = 0 \tag{3}$$

Finally, a usual Spence-Mirrlees single-crossing condition is imposed to ensure higher realisations of θ can naturally be associated with higher productivity:[3]

$$\left(\frac{u_{c\theta}}{u_c}-\frac{u_{y\theta}}{u_y}\right)\frac{u_y}{u_c}<0 \tag{4}$$

Condition (4) comes from differentiating the expression for the slope of a within-period indifference curve in output-consumption space ($\frac{dc}{dy}=-\frac{u_y}{u_c}$); the Spence-Mirrlees condition demands that this indifference curve should be flattening in θ. Since we are assuming the stochastic productivity parameter has discrete support it is useful also re-state the condition for non-marginal changes in θ: if $\theta'<\theta''$ then condition (4) implies for all (c,y) pairs:

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

The policymaker's role is to choose, at the start of the first time period, tax schedules for all future periods that will link an individual's consumption to their output level, conditional upon their history of actions to date. In treating consumption as a choice variable of the policymaker we are implicitly assuming that there are no opportunities for the individuals to engage in 'hidden' saving, so that the policymaker can behave as if taxing savings at a 100 per cent marginal rate if this is necessary to prevent 'unwanted' consumption deferral.[4] Since the revelation principle will apply in this setting (see, for instance, Golosov, Tsyvinski and Werning (2006)), we may equivalently restrict the policy choice to direct revelation mechanisms that simply deliver consumption and output levels to individuals as a function of their current and past productivity reports – deferring for now a consideration of decentralisation schemes. We denote by $\widehat{\theta}_t^i\left(\theta^\infty\right):\Theta^t\to\Theta$ individual i's report at time t as a function of their actual productivity (where this function is measurable with respect to θ^t), by $\widehat{\theta}^{i,t}\left(\theta^\infty\right):\Theta^t\to\Theta^t$ the history of all such reports up to time t, and by $\widehat{\theta}^{i,\infty}\left(\theta^\infty\right):\Theta^\infty\to\Theta^\infty$ a complete sequence of reports. We refer to $\widehat{\theta}^{i,t}\left(\cdot\right)$ as the t-truncation of $\widehat{\theta}^{i,\infty}\left(\cdot\right)$. The policymaker's primal choice problem is:

$$\max_{\{c_t(\theta^\infty),y_t(\theta^\infty)\}_{t=1}^\infty}\int_{\Theta^\infty}\sum_{t=1}^{\infty}\beta^{t-1}u\left(c_t\left(\widetilde{\theta}^\infty\right),y_t\left(\widetilde{\theta}^\infty\right);\widetilde{\theta}_t\right)d\pi_\Theta\left(\widetilde{\theta}^\infty\right) \tag{6}$$

subject to $c_t\left(\theta^\infty\right)$ and $y_t\left(\theta^\infty\right)$ being measurable with respect to θ^t, together

[3] This condition is in fact implied by our earlier assumptions, so long as consumption is a normal good and we are considering a consumption-output bundle associated with a weakly positive tax distortion. Normality is easily shown to imply $u_{cy}+u_{yy}<0$, whilst a tax wedge ensures $u_c+u_y>0$. Together these imply $u_{cy}-\frac{u_c}{u_y}u_{yy}<0$, and the single crossing condition then follows from simple applications of (2) and (3).

[4] 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.

with the incentive compatibility constraint:

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

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

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

where A_t is the quantity of real bonds that the policymaker purchases for time t, each paying R_t units of real income. Dynamic solvency requires that $\lim_{s\rightarrow\infty}\left[\left(\prod_{r=1}^{s} R_{t+r}^{-1}\right) A_{t+s}\right] = 0$.[5]

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 that it is sufficient to impose a binding restriction to prevent agents with histories $\left(\theta^{t-1}, \theta_t'\right)$ mimicking those with histories $\left(\theta^{t-1}, \theta_t''\right)$, where $\theta_t'' = \max\left\{\theta \in \Theta : \theta < \theta_t'\right\}$. The basic intuition for this result – that envy is always directed 'downwards' from one type to the next in equilibrium – is familiar from the analysis of static optimal tax models, articulated most clearly by Dasgupta (1982). To understand it, it is useful to start the analysis 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 can be neglected, with lump-sum taxation used to implement a first-best. We summarise the important features of this first-best in a series of propositions, the proofs of which are trivial with the exception of the fourth.

Proposition 1 *In the full information benchmark the optimal allocations $c_t(\theta^\infty)$ and $y_t(\theta^\infty)$ are measurable with respect to θ_t.*

Proof. History invariance follows immediately from the absence of any expectational constraints restricting the policymaker in the full information case. ■

[5] In what follows we will sometimes suppress the explicit dependence of c_t and y_t upon θ^∞, as well as indexing these functions with individual-specific superscripts where this is most natural.

This result is useful, as it implies we can consider which agents envy others in the full information case by focusing on within-period outcomes alone. If an agent of type θ_t' could somehow make the policymaker believe he or she was of type θ_t'', there would be no direct implications for that agent's expected outcomes from period $t+1$ onwards.

Proposition 2 *The following conditions hold for all $t \geq 1$ and all $i \in [0,1]$:*

$$u_c\left(c_t^i, y_t^i; \theta_t^i\right) = -u_y\left(c_t^i, y_t^i; \theta_t^i\right) \tag{9}$$

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

Proof. An interior optimum is guaranteed by the Inada conditions, and these conditions are then necessary by the usual arguments. ■

Unsurprisingly, these conditions imply that the first-best involves no distortion at the labour-consumption margin, and no intertemporal distortion. All redistribution takes place through lump-sum wealth transfers across agents.

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

$$u_c\left(c_t^i, y_t^i; \theta_t^i\right) = u_c\left(c_t^j, y_t^j; \theta_t^j\right) \tag{11}$$

Proof. Again, the Inada conditions ensure an interior optimum, and this condition must hold if the utilitarian policymaker is unable to improve outcomes by further redistribution. ■

In the event that consumption and labour are additively separable in the utility function we will have $u_{cy} = u_{c\theta} = 0$, and the preceeding condition implies equalised consumption across all agents (since $u_{cc} < 0$). Since agents who are more productive have a higher marginal product for a given quantity of labour, they will generally be required to work harder. For this reason utility is decreasing in type – a result that extends to the general preference specification under a minor additional restriction:

Proposition 4 *$\theta_t^i > \theta_t^j$ implies $u\left(c_t^i, y_t^i; \theta_t^i\right) < u\left(c_t^j, y_t^j; \theta_t^j\right)$, so long as $u_{cc} + u_{cy} < 0$.*

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

$$\frac{dc}{d\theta} = \frac{u_{yy}u_{c\theta} - u_{cy}u_{y\theta}}{u_{cy}^2 - u_{cc}u_{yy}}$$

$$\frac{dy}{d\theta} = \frac{u_{cc}u_{y\theta} - u_{cy}u_{c\theta}}{u_{cy}^2 - u_{cc}u_{yy}}$$

The overall effect on utility at the margin is:

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

(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 $u_{cy}^2 - u_{cc}u_{yy} < 0$,[6] and we have $u_y < 0$, so if $u_{cc} + u_{cy} < 0$ then condition (2) gives:

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

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

The requirement that $u_{cc} + u_{cy} < 0$ deserves brief comment here: it says that the marginal utility of consumption should not be increasing as an agent increases both output and consumption by a unit at the margin (holding type constant). Intuitively, if the converse were true then holding an agent's marginal utility of consumption constant as he or she works a greater number of hours would require an increase in consumption greater than any marginal increase in output, and this tends to increase utility when we start from an optimal point at which the marginal rate of substitution between counsumption and output is 1. But the inequality is a perfectly natural one to assume in this case, as our last proposition in this section makes clear.

Proposition 5 *Leisure is a normal good if and only if* $u_{cc} + u_{cy} < 0$.

Proof. The consumer's problem is:

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

subject to

$$c = y + \omega$$

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

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

[6] Consider movements in the vector $\begin{bmatrix} -\frac{u_{yc}}{u_{cc}} \\ 1 \end{bmatrix}$ here.

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

To summarise the main lessons of these five propositions, we know that the first-best solution for a utilitarian policymaker involves zero marginal distortions on savings and labour supply, as well as equalised marginal consumption utility (and output disutility) across agents each period. At the same time, so long as leisure is a normal good we know that the optimum sees higher-type agents experience a lower level of utility than those who are less productive than them – a result that derives from the higher marginal productivity of these agents in the event that consumption and labour supply levels are constant across agents.[7] Together with the fact that there is no history dependence in outcomes at the optimum, this result 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 can only verify agents' output levels, and not their types.

It does not follow directly from this analysis that incentive compatibility constraints will always bind 'downwards' at a constrained optimum, but the aim of considering the first-best is to show that we have do decent grounds for *assuming* this – even when preferences take a far more general form than is conventional in this setting. $u_\theta > 0$, and so even if the policymaker can provide no incentives to stop all agents receiving the same consumption-output bundle in a given time period, it must be the case that utility is *higher* for those who are more productive. The single crossing condition is well-known in the static model to ensure that a separating equilibrium will exist, in which each agent of type θ' envies the agent of type $\theta'' = \max\left\{\theta \in \Theta : \theta < \theta'\right\}$. The dynamic model introduces the possibility that incentives can be spread through time, but the fact that the first-best involves no history dependence again suggests that this is unlikely to change the subset of incentive compatibility constraints that is binding.

4 Applying perturbation analysis: the case of iid shocks

To keep the analysis manageable it is useful to focus first on the case in which productivity shocks are iid. This allows for a clearer understanding of the general perturbation techniques we apply, but it is also a case in which we can be a lot surer regarding the sufficiency of local incentive compatibility constraints binding. We discuss this matter first.

4.1 Incentive compatibility when shocks are iid

For all periods $t \geq 1$, define $\widehat{\theta}_{m,t}^{\infty}(\theta^\infty) : \Theta^\infty \to \Theta^\infty$ as the reporting strategy associated with truth-telling in all periods up to t, when the agent mimics a

[7] It is, indeed, a case of 'From each according to his means, to each according to his needs.'

type one lower and follows an optimal reporting strategy thereafter:

$$\widehat{\theta}^{\infty}_{m,t}\left(\theta'^{\infty}\right) = \left[\theta'_1, \theta'_2, ..., \theta'_{t-1}, \theta''_t, \theta''_{t+1}, ...\right]$$

where $\theta''_t = \max\left\{\theta \in \Theta : \theta < \theta'_t\right\}$ and $\left\{\theta''_{t+1}, \theta''_{t+2}, ...\right\}$ are then optimal choices conditional upon prior reports. If $\theta'_t = \min\left\{\theta \in \Theta\right\}$ then we normalise $\widehat{\theta}^{\infty}_{m,t}\left(\theta^{\infty}\right) = \theta^{\infty}$. If incentive compatabilty holds 'downwards', the following is true:

$$\int_{\Theta^{\infty}} \sum_{s=0}^{\infty} \beta^s u\left(c_{t+s}\left(\widetilde{\theta}^{\infty}\right), y_{t+s}\left(\widetilde{\theta}^{\infty}\right); \widetilde{\theta}_{t+s}\right) d\pi_{\Theta}\left(\widetilde{\theta}^{\infty}|\theta^t\right) \tag{12}$$

$$\geq \int_{\Theta^{\infty}} \sum_{s=0}^{\infty} \beta^s u\left(c_{t+s}\left(\widehat{\theta}^{\infty}_{m,t}\left(\widetilde{\theta}^{\infty}\right)\right), y_{t+s}\left(\widehat{\theta}^{\infty}_{m,t}\left(\widetilde{\theta}^{\infty}\right)\right); \widetilde{\theta}_{t+s}\right) d\pi_{\Theta}\left(\widetilde{\theta}^{\infty}|\theta^t\right)$$

When the condition binds strictly, the agent with history θ^t is just indifferent between reporting θ_t truthfully and mimicking a type one lower, provided θ_t is not itself the smallest element in Θ. It is simple to show that in this case agents with an *arbitrary* reporting strategy prior to t will also be indifferent between truth-telling and mimicking in the same manner.[8] We are interested in the conditions under which this restriction implies *global* incentive compatibility – that is, for an arbitrary reporting strategy at t, $\widehat{\theta}^{\infty}_{a,t}\left(\theta^{\infty}\right) : \Theta^{\infty} \rightarrow \Theta^{\infty}$, defined by:

$$\widehat{\theta}^{\infty}_{a,t}\left(\theta'^{\infty}\right) = \left[\theta'_1, \theta'_2, ..., \theta'_{t-1}, \theta''_t, \theta''_{t+1}, ...\right]$$

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

$$\int_{\Theta^{\infty}} \sum_{s=0}^{\infty} \beta^s u\left(c_{t+s}\left(\widetilde{\theta}^{\infty}\right), y_{t+s}\left(\widetilde{\theta}^{\infty}\right); \widetilde{\theta}_{t+s}\right) d\pi_{\Theta}\left(\widetilde{\theta}^{\infty}|\theta^t\right) \tag{13}$$

$$\geq \int_{\Theta^{\infty}} \sum_{s=0}^{\infty} \beta^s u\left(c_{t+s}\left(\widehat{\theta}^{\infty}_{a,t}\left(\widetilde{\theta}^{\infty}\right)\right), y_{t+s}\left(\widehat{\theta}^{\infty}_{a,t}\left(\widetilde{\theta}^{\infty}\right)\right); \widetilde{\theta}_{t+s}\right) d\pi_{\Theta}\left(\widetilde{\theta}^{\infty}|\theta^t\right)$$

We can show the following:

Proposition 6 *Suppose productivity shocks are iid and consumption is a normal good. Then any direct revelation mechanism that the policymaker chooses that satisfies (12) will additionally satisfy (13), provided it has the following properties:*

1. *The constraints in (12) bind with equality.*

[8] To see this, consider the 'arbitrary' reporting strategy $\widehat{\theta}^{\infty}_{a,t}\left(\theta^{\infty}\right) = \left[\theta'^{t-1}, \theta_t, \theta_{t+1}, ...\right]$ 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 θ'^{t-1}, and has current type θ_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 compatability constraint binds – implying it must also do so for the agent whose prior reports were arbitrary.

2. *Conditional upon* $\widehat{\theta}^{t-1}$, $c_t\left(\widehat{\theta}^{\infty}\right)$ *and* $y_t\left(\widehat{\theta}^{\infty}\right)$ *are strictly increasing in* $\widehat{\theta}_t$ *for all* $t \geq 1$ *and* $\widehat{\theta}^{t-1} \in \Theta^{t-1}$.

3. *Conditional upon* $\widehat{\theta}^{t-1}$, $\frac{c_t\left(\widehat{\theta}^{t-1},\widehat{\theta}'_t\right)-c_t\left(\widehat{\theta}^{t-1},\widehat{\theta}''_t\right)}{y_t\left(\widehat{\theta}^{t-1},\widehat{\theta}'_t\right)-y_t\left(\widehat{\theta}^{t-1},\widehat{\theta}''_t\right)} < 1$ *for all distinct* $\widehat{\theta}'_t, \widehat{\theta}''_t \in \Theta$.[9]

In addition, condition (13) will hold with strict inequality unless the strategy $\widehat{\theta}^{\infty}_{a,t}$ *involves reporting* $\widehat{\theta}_t \in \{\max\{\theta \in \Theta : \theta < \theta_t\}, \theta_t\}$

Proof. See appendix. ■

This is an extremely useful result, as it suggests that if we can find an optimal mechanism for the 'relaxed' problem, in which only the constraints defined by equation (12) are imposed, then we need only check the two conditions above to verify that it also satisfies 'global' incentive compatibility. Moreover, the conditions themselves will indeed be features of any optimum, though demonstrating this fully requires further development of the analysis. Intuitively, the fact that the marginal disutility of production is diminishing in θ makes it natural to screen higher types through higher consumption and output bundles, whilst the restriction that $\frac{c_t\left(\widehat{\theta}^{t-1},\widehat{\theta}'_t\right)-c_t\left(\widehat{\theta}^{t-1},\widehat{\theta}''_t\right)}{y_t\left(\widehat{\theta}^{t-1},\widehat{\theta}'_t\right)-y_t\left(\widehat{\theta}^{t-1},\widehat{\theta}''_t\right)} < 1$ is equivalent to asserting that higher types do not pay negative tax rates on their additional incomes relative to lower types. Since the marginal benefits of transferring income to higher types will be relatively low, this certainly ought to hold (at least for well-behaved preferences).

Finally, the fact that we can show incentive compatibility binds strictly for non-truthful reporting strategies other than 'downwards' mimicking is of great use in what follows, since it implies small enough perturbations to the optimal allocations cannot cause these constraints to be violated – and thus that the first-order conditions associated with these perturbations can be asserted with confidence.

4.2 Perturbation analysis

The remainder of this section is devoted to 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 in the iid case. The analysis is perhaps easiest to understand with the aid of an indifference curve map, linking output on the horizontal axis and consumption on the vertical, and the reader is actively encouraged to visualise matters in this form.[10]

[9] Recall that c_t and y_t are measurable with respect to $\widehat{\theta}^t$.

[10] The textbook treatment of the static optimal income tax problem by Atkinson and Stiglitz (1980) makes use of this geometrical presentation – see page 413.

Conditional upon a particular reporting history prior to the current time period t, denoted $\widehat{\theta}^{t-1}$, an agent's time t report-contingent consumption and output allocations under the optimal scheme can be described by an $N \times 2$ matrix $\Psi_t^* \left(\widehat{\theta}^{t-1} \right)$, with each row corresponding to a particular $\theta_t \in \Theta$,[11] and the columns listing, in turn, consumption and output levels for the given 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 $\Delta(\delta) : \mathbb{R} \to \mathbb{R}^{2N}$ for some relevant parameter δ (perhaps the consumption increment for the agent of the highest type), 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 reporting history $\widehat{\theta}^{t-1}$ changed at $t-1$ according to an analogous function $\Delta_{-1}(\delta) : \mathbb{R} \to \mathbb{R}^2$. We wish to construct these Δ and Δ_{-1}functions so that they satisfy the following three properties:

1. Incentive compatibility constraints that hold for each time period up to the tth when allocations for agents with reporting history $\widehat{\theta}^{t-1}$ are $\left(c_{t-1}^* \left(\widehat{\theta}^{t-1} \right), y_{t-1}^* \left(\widehat{\theta}^{t-1} \right) \right)$at $t-1$ and $\Psi_t^* \left(\widehat{\theta}^{t-1} \right)$ at t will continue to hold when allocations are $\left(c_{t-1}^* \left(\widehat{\theta}^{t-1} \right), y_{t-1}^* \left(\widehat{\theta}^{t-1} \right) \right) + \Delta_{-1}(\delta)$ at $t-1$ and $\Psi_t^* \left(\widehat{\theta}^{t-1} \right) + \Delta(\delta)$ at t, provided δ is small enough.

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 δ in the neighbourhood of $\delta = 0$.

Since we work under the assumption that incentive compatibility constraints bind only 'downwards' at the optimum, the first property is equivalent to requiring that any additional incentive that an agent of type θ_t^n may have to mimic an agent of type θ_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 θ_t^n receives from truthful reporting.[12] Symmetrically, we impose that a reduction in the incentives to mimic should be matched by an equal reduction in the utility from truthful reporting.

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

[11] We assume that these are ordered in ascending values for θ_t, with the lowest type's allocation in the first row of Ψ^* and the highest type's in the Nth row.

[12] We use superscripts here to index the agents' types within the set Θ, with θ_t^n increasing in $n \in \{1, ..., N\}$

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 a wedge between the two. This requirement provides a substantial obstacle relative to the firstl: 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. But perturbations of this form will only ever give us *in*equality restrictions – to the effect that the net marginal cost of changing outcomes thus 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)). Thus in all cases the policymaker will experience no direct loss or gain from the perturbation. A necessary condition for the original allocations $\left(c^*_{t-1}, y^*_{t-1}\right)$ and Ψ^*_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, contradicting the presumed optimality of the original allocation.

4.2.1 Deriving admissible perturbations

Perturbations at the top There is a very simple example of a perturbation that satisfies all three of the above requirements, and it is useful to start by considering it. This is a movement along the within-period indifference curve of the 'top' agent. 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.[13] 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 $u_c = -u_y$.[14] The logic generalises to the intertemporal model, as the following makes clear.

[13] When this set has unbounded support the result need no longer hold, as the influential work by Saez (2001) has emphasised.

[14] 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, since utility provision to higher types is a relatively wasteful use of resources at the optimum.

Proposition 7 *Suppose the optimal solution to the 'relaxed' iid problem in which only constraints (12) are imposed has the three properties described in Proposition 6. Then in all time periods $t \geq 1$ and (if $t > 1$) for all reporting histories $\widehat{\theta}^{t-1}$, the allocation (c_t^*, y_t^*) for the agent who reports $\widehat{\theta}_t'$ such that $\theta_t' = \max\{\theta \in \Theta\}$ satisfies $u_c\left(c_t^*, y_t^*; \theta_t'\right) = -u_y\left(c_t^*, y_t^*; \theta_t'\right)$.*

Proof. Consider a perturbation to the allocation $\Psi_t^*\left(\widehat{\theta}^{t-1}\right)$ given by the $N \times 2$ matrix of functions $\Delta(\delta) : \mathbb{R} \to \mathbb{R}^{2N}$ such that the nth row of Δ equals $(0, 0)$ for all $n \in \{1, ..., N-1\}$ and the Nth row equals $(\delta, f(\delta))$, with $f(\delta) : \mathbb{R} \to \mathbb{R}$ defined implicitly by:

$$u\left(c_t^* + \delta, y_t^* + f(\delta); \theta_t'\right) = u\left(c_t^*, y_t^*; \theta_t'\right) \tag{14}$$

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 $\widehat{\theta}_t'$ when not of type θ_t', but Proposition 6 implies all agents of a lower type than θ_t' *strictly* prefer their own allocation to (c_t^*, y_t^*) at the optimum, and so for values of δ close enough to zero their incentive compatibility constraints will remain satisfied. This then implies that the value of the policymaker's objective will remain unchanged as δ is varied in the neighbourhood of $\delta = 0$. The impact of the perturbation on the resources available to the policymaker in period t (in equilibrium) will be $\pi_\Theta\left(\theta_t'|\theta^{t-1}\right)\pi_\Theta\left(\theta^{t-1}\right)[f(\delta) - \delta]$. If the original allocation is optimal then the *marginal* impact on resources as δ moves away from zero must be zero, or else it would be possible to raise a surplus:

$$\pi_\Theta\left(\theta_t'|\theta^{t-1}\right)\pi_\Theta\left(\theta^{t-1}\right)[f'(0) - 1] = 0 \tag{15}$$

Probabilities are non-zero, so this implies:

$$f'(0) = 1 \tag{16}$$

Since utility for a truth-teller is unchanged by the perturbation we have:

$$u_c\left(c_t^*, y_t^*; \theta_t'\right) + u_y\left(c_t^*, y_t^*; \theta_t'\right) f'(0) = 0 \tag{17}$$

The result follows immediately. ■

This proposition – which generalises easily to the non-iid case, covered below – is interesting in its own right, since Kocherlakota (2011) has provided a computed example in which the optimal consumption-output distortion for 'top' agents is non-zero, conditional upon a particular past report. As argued in more detail below, it seems that this result derives from the author's assumption about the conditional distribution of shocks – specifically, that the *support* of this distribution depends upon past productivity draws, so that the highest productivity draw a previously low-type agent can hope for is less than the highest realised across all agents in the economy. When this sort of dependence is *not* permitted, the 'no distortion at the top' result familiar from the static optimal tax literature holds in the dynamic case too.

Uniform utility perturbations As already discussed, 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 provides the natural generalisation of this analysis to the non-separable case. 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 again throughout much of the subsequent analysis.

Before presenting the proposition we define the function $\alpha(c, y; \theta) : \mathbb{R}_+^2 \times \Theta \to \mathbb{R}$ as follows:

$$\alpha(c, y; \theta) = \frac{u_c(c, y; \theta) - u_c(c, y; \theta')}{u_y(c, y; \theta') - u_y(c, y; \theta)} \tag{18}$$

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

Proposition 8 *Suppose the optimal solution to the 'relaxed' iid problem in which only constraints (12) are imposed has the three properties described in Proposition 6. Then for all time periods $t \geq 1$ and for all reporting histories $\widehat{\theta}^t$, the allocations $\left(c_t^*\left(\widehat{\theta}^t\right), y_t^*\left(\widehat{\theta}^t\right)\right)$ and $\Psi_{t+1}^*\left(\widehat{\theta}^t\right)$ satisfy the following condition:*

$$\begin{aligned} & R_{t+1}\beta \frac{1 - \alpha(c_t^*, y_t^*; \theta_t)}{u_c(c_t^*, y_t^*; \theta_t) + u_y(c_t^*, y_t^*; \theta_t)\,\alpha(c_t^*, y_t^*; \theta_t)} \\ & = \sum_{\theta_{t+1} \in \Theta} \pi_\Theta(\theta_{t+1}|\theta_t) \frac{1 - \alpha\left(c_{t+1}^*, y_{t+1}^*; \theta_{t+1}\right)}{u_c\left(c_{t+1}^*, y_{t+1}^*; \theta_{t+1}\right) + u_y\left(c_{t+1}^*, y_{t+1}^*; \theta_{t+1}\right)\alpha\left(c_{t+1}^*, y_{t+1}^*; \theta_{t+1}\right)} \end{aligned} \tag{19}$$

where θ_t is the last entry in $\widehat{\theta}^t$ and c_{t+1}^* and y_{t+1}^* are given by the relevant entries in $\Psi_{t+1}^*\left(\widehat{\theta}^t\right)$.[15]

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 $\widehat{\theta}^t$ 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 δ 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 tth. The difficulty lies in constructing the perturbations in a way that will preserve incentive compatibility. Again, we exploit the supposition that the conditions set out in Proposition 6 apply. This implies that for small enough values of δ we need only concern ourselves with the $N-1$ constraints at $t+1$ that prevent mimicking by types 'one higher' than any given $\theta_{t+1} \in$

[15] We suppress dependence upon $\widehat{\theta}^t$ to keep the notation manageable.

Θ, and the similar t-dated constraint preventing mimicking of type θ_t by the immediately superior type.

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

$$\Delta_{-1}(\delta) = \left(\phi^c\left(\theta_t, -\beta\delta; c_t^*, y_t^*\right), \phi^y\left(\theta_t, -\beta\delta; c_t^*, y_t^*\right)\right) \tag{20}$$

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

$$u\left(c^* + \phi^c(\theta, k; c^*, y^*), y^* + \phi^y(\theta, k; c^*, y^*); \theta\right) = u\left(c^*, y^*; \theta\right) + k \tag{21}$$

$$u\left(c^* + \phi^c(\theta, k; c^*, y^*), y^* + \phi^y(\theta, k; c^*, y^*); \theta'\right) = u\left(c^*, y^*; \theta'\right) + k \tag{22}$$

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

$$u\left(c^* + \phi^c(\theta, k; c^*, y^*), y^*; \theta\right) = u\left(c^*, y^*; \theta\right) + k \tag{23}$$

$$\phi^y(\theta, k; c^*, y^*) = 0 \tag{24}$$

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, since the single crossing property holds. Similarly, the nth row of $\Delta(\delta)$is given by:

$$\left[\phi^c\left(\theta_{t+1}^n, \delta; c_{t+1}^*, y_{t+1}^*\right), \phi^y\left(\theta_{t+1}^n, \delta; c_{t+1}^*, y_{t+1}^*\right)\right] \tag{25}$$

where we index by type in the natural way. By construction this perturbation must preserve incentive compatibility at $t+1$ for small enough values of δ, since the within-period utility that any agent can gain from mimicking is being changed by exactly the same amount (δ) as the within-period utility from truth-telling (for the mimicking strategies that need concern us). It must also preserve incentive compatibility at t (for small enough δ), 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 δ units at $t+1$, discounted at rate β), both for agents of true type θ_t and for the potential mimickers whose type is one higher.[16] 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:

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

[16] Note that this assertion does not rely on the iid assumption, since utility is increased uniformly across all types at $t+1$.

We require for optimality that the derivative of this expression with respect to δ should equal zero when $\delta = 0$; otherwise the policymaker could use less resources in obtaining the same value for aggregate utility. This implies the optimality condition:

$$\begin{aligned} & \beta \left[\phi_2^c \left(\theta_t, 0; c_t^*, y_t^* \right) - \phi_2^y \left(\theta_t, 0; c_t^*, y_t^* \right) \right] \\ = \; & R_{t+1}^{-1} \sum_{\theta_{t+1} \in \Theta} \pi_\Theta \left(\theta_{t+1} | \theta^t \right) \left[\phi_2^c \left(\theta_{t+1}^n, 0; c_{t+1}^*, y_{t+1}^* \right) \right. \\ & \left. - \phi_2^y \left(\theta_{t+1}^n, \delta; c_{t+1}^*, y_{t+1}^* \right) \right] \end{aligned} \tag{26}$$

where ϕ_2^c denotes the derivative of ϕ^c with respect to its second argument. By total differentiation of conditions (21) to (24) with respect to k it is easy to show:

$$\begin{aligned} & \phi_2^c \left(\theta, 0; c^*, y^* \right) - \phi_2^y \left(\theta, 0; c^*, y^* \right) \\ = \; & \frac{1 - \alpha \left(c^*, y^*; \theta \right)}{u_c \left(c^*, y^*; \theta \right) + u_y \left(c^*, y^*; \theta \right) \alpha \left(c^*, y^*; \theta \right)} \end{aligned} \tag{27}$$

The result follows. ■

Again, this result is an interesting one in its own right. On a simple analytical level, together with its generalisation to the non-iid case (which is straightforward) it helps fill a significant gap in the existing theory. Golosov, Tsyvinski and Werning (2006) wrote 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 help go some way towards achieving this.

More importantly from an economic perspective, we are able to give some 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 sense that under some preference structures we cannot analytically rule out the inequality:

$$u_c \left(c_t^*, y_t^*; \theta_t \right) > \beta R_{t+1} \sum_{\theta_{t+1} \in \Theta} \pi_\Theta \left(\theta_{t+1} | \theta_t \right) u_c \left(c_{t+1}^*, y_{t+1}^*; \theta_{t+1} \right) \tag{28}$$

holding in certain time periods for certain realisations of θ^∞. 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. But to gain some insight on this wedge it is better first to provide some direct interpretation of equation (19).

The function $\alpha(c, y; \theta)$ gives the number of units by which the output level of an agent of type θ is increased from y at the margin per unit marginal increase in that agent's consumption from c, along the unique vector dimension that provides equal utility increments to both the agent of true type θ and the agent whose type is one higher. In the familiar case of separable preferences this dimension is orthogonal to agents' output levels: consumption utility is completely independent of one's labour supply level in producing a given quantity of output, and thus is also independent of one's type. In this case $\alpha(c, y; \theta) = 0$. More generally, it follows that the expression $\frac{1-\alpha(c,y;\theta)}{u_c(c,y;\theta)+u_y(c,y;\theta)\alpha(c,y;\theta)}$ gives the ratio between the marginal resource cost to the policymaker of providing the agent with one unit of consumption, and the marginal utility impact that doing so will have – given that every unit increase in consumption must be accompanied by an increase in output of $\alpha(c, y; \theta)$ units. Hence this expression can be read as the marginal resource cost of providing a unit of utility to the agent of type θ in the neighbourhood of (c, y). The optimality condition states that the marginal resource cost of providing β units of utility at time t to an agent with shock history θ^t should be the same in present value terms as the marginal cost of providing 1 unit to that agent across all productivity realisations at $t+1$. When $\alpha(c, y; \theta) = 0$, which holds in the case of separable preferences, it collapses to the familiar inverse Euler condition.

By Jensen's inequality we know that $\sum_{\theta_{t+1}\in\Theta} \pi_\Theta(\theta_{t+1}|\theta_t) \left[\frac{1}{u_c\left(c^*_{t+1}, y^*_{t+1}; \theta_{t+1}\right)}\right] \geq \left[\sum_{\theta_{t+1}\in\Theta} \pi_\Theta(\theta_{t+1}|\theta_t)\, u_c\left(c^*_{t+1}, y^*_{t+1}; \theta_{t+1}\right)\right]^{-1}$, with a strict inequality holding so long as the marginal utility of consumption varies in θ_{t+1}. From this it is easy to show in the case of separable preferences ($\alpha(c, y; \theta) = 0$) that savings are deterred:[17]

$$u_c\left(c^*_t, y^*_t; \theta_t\right) \leq \beta R_{t+1} \sum_{\theta_{t+1}\in\Theta} \pi_\Theta(\theta_{t+1}|\theta_t)\, u_c\left(c^*_{t+1}, y^*_{t+1}; \theta_{t+1}\right) \tag{29}$$

again with a strict inequality holding so long as the marginal utility of consumption varies in θ_{t+1}.

The reason the usual Euler condition (with an equality in the above relationship) does not hold in this environment is just that the policymaker does not have the option to engineer the sorts of perturbations that would imply it. The Euler condition need obtain only if it is possible to reduce (increase) consumption at time t by one unit, save the proceeds at the interest rate R_{t+1}, and increase (reduce) consumption across all states of the world at $t+1$ by R_{t+1} units. But since $u_{cc} < 0$, this perturbation will imply a lower utility gain (loss) for those $t+1$ types whose equilibrium consumption is high, relative to the utility gain (loss) received by those they are on the cusp of mimicking. This implies incentive compatibility will in general be violated at $t+1$ when the perturbation tends to increase savings.

The Euler condition states that spreading resources through time, with equal

[17] See, for instance, Golosov, Kocherlakota and Tsyvinski (2003) for a fuller treatment.

consumption increments across states at $t+1$, cannot raise utility. But our policymaker is limited to ensuring that spreading *utility* through time, with equal utility increments across states at $t+1$, cannot raise surplus *resources*. 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. In the more general case this marginal cost is the expression contained in equation (19). It turns out that inequality (29) does generalise a little beyond the separable case, but not much.

Proposition 9 *Suppose the optimal solution to the 'relaxed' iid problem in which only constraints (12) are imposed has the three properties described in Proposition 6, and additionally that in all time periods $s \geq 1$ and for all reporting histories $\widehat{\theta}^s$ the allocations $\left(c_s^*\left(\widehat{\theta}^s\right), y_s^*\left(\widehat{\theta}^s\right)\right)$ imply $u_c\left(c_s^*, y_s^*; \theta_s\right) \geq u_y\left(c_s^*, y_s^*; \theta_s\right)$. Then for all time periods $t \geq 1$ and for all reporting histories $\widehat{\theta}^t$, the allocations $\left(c_t^*\left(\widehat{\theta}^t\right), y_t^*\left(\widehat{\theta}^t\right)\right)$ and $\Psi_{t+1}^*\left(\widehat{\theta}^t\right)$ will satisfy inequality (29) 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 is well known in the event that condition 1 holds, and does not require the extra assumption that intratemporal wedges are weakly positive. With separability $u_c(c, y; \theta) = u_c(c, y; \theta')$ for all $\theta, \theta' \in \Theta$, and so $\alpha(c, y; \theta) = 0$ always holds. We can then apply Jensen's inequality to the inverse Euler condition in the manner set out above, noting that Proposition 6 assumes consumption to be strictly increasing in θ_{t+1} (implying $u_c\left(c_{t+1}^*, y_{t+1}^*; \theta_{t+1}\right)$ varies in θ_{t+1}), so the inequality can be stated strictly.

The definition of Edgeworth substitutes implies $u_{cy} < 0$ 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:

$$\frac{1-\alpha(\theta)}{u_c(\theta) + u_y(\theta)\alpha(\theta)} \geq \frac{1}{u_c(\theta)} \tag{30}$$

(where we now suppress dependence upon c and y to ease notation). Hence:

$$\begin{aligned} & \sum_{\theta_{t+1}\in\Theta} \pi_\Theta(\theta_{t+1}|\theta_t) \frac{1-\alpha(\theta_{t+1})}{u_c(\theta_{t+1}) + u_y(\theta_{t+1})\alpha(\theta_{t+1})} \\ \geq & \sum_{\theta_{t+1}\in\Theta} \pi_\Theta(\theta_{t+1}|\theta_t) \frac{1}{u_c(\theta_{t+1})} \\ \geq & \left[\sum_{\theta_{t+1}\in\Theta} \pi_\Theta(\theta_{t+1}|\theta_t) u_c(\theta_{t+1})\right]^{-1} \end{aligned} \tag{31}$$

where the last result uses Jensen's inequality, and will hold strictly provided the marginal utility of consumption varies in θ_{t+1}. If $\theta_t = \max\{\theta \in \Theta\}$ then Proposition 7 implies $u_c(c_t^*, y_t^*; \theta_t) = u_y(c_t^*, y_t^*; \theta_t)$, so we have:

$$\begin{aligned} R_{t+1}\beta\frac{1}{u_c(\theta_t)} &= R_{t+1}\beta\frac{1-\alpha(\theta_t)}{u_c(\theta_t)+u_y(\theta_t)\alpha(\theta_t)} \\ &= \sum_{\theta_{t+1}\in\Theta}\pi_\Theta(\theta_{t+1}|\theta_t)\frac{1-\alpha(\theta_{t+1})}{u_c(\theta_{t+1})+u_y(\theta_{t+1})\alpha(\theta_{t+1})} \\ &\geq \left[\sum_{\theta_{t+1}\in\Theta}\pi_\Theta(\theta_{t+1}|\theta_t)\,u_c(\theta_{t+1})\right]^{-1} \end{aligned} \tag{32}$$

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 15 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. Beyond this, though, it is hard to say much of specific relevance to the wedge present in the conventional Euler condition. But this condition isn't the *only* way to characterise an optimal savings decision in an economy free from taxes. For instance, in that setting optimality 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. And, indeed, 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 must not increase utility (for a movement in either direction) at an optimum. In particular, in a world with no taxation the following condition would hold:

$$\begin{aligned} &\frac{u_c(\theta_t)+u_y(\theta_t)\alpha(\theta_t)}{1-\alpha(\theta_t)} \\ =\;& \beta R_{t+1}\sum_{\theta_{t+1}\in\Theta}\pi_\Theta(\theta_{t+1}|\theta_t)\frac{u_c(\theta_{t+1})+u_y(\theta_{t+1})\alpha(\theta_{t+1})}{1-\alpha(\theta_{t+1})} \end{aligned} \tag{33}$$

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 of cost. The dynamic optimality 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 there is a sense in which the agent's saving levels are being distorted whenever equation (33) does not hold, with saving implicitly being discouraged whenever the left-hand side is less than the right. This saving distortion may well interact with concurrent distortions at the labour-consumption margin within a period, but there is nothing inherently correct about focusing on deviations from the traditional *consumption* Euler equation in assessing the extent of savings distortions. An Euler equation that states that the marginal rate of substitution between leisure in any two periods must equal the intertemporal price ratio (R_{t+1}), or some combination between this and the consumption Euler condition such as equation (33) is of equal validity in characterising savings behaviour under autarky.

The useful feature of equation (33) 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. We have the following.

Proposition 10 *Suppose the optimal solution to the 'relaxed' iid problem in which only constraints (12) are imposed has the three properties described in Proposition 6. Then for all time periods $t \geq 1$ and for all reporting histories $\widehat{\theta}^t$, 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 $\left(c_t^*\left(\widehat{\theta}^t\right), y_t^*\left(\widehat{\theta}^t\right)\right)$ and $\Psi_{t+1}^*\left(\widehat{\theta}^t\right)$ will satisfy the following condition:*

$$\begin{aligned} & \frac{u_c(\theta_t) + u_y(\theta_t)\,\alpha(\theta_t)}{1 - \alpha(\theta_t)} \\ \leq \quad & \beta R_{t+1} \sum_{\theta_{t+1} \in \Theta} \pi_\Theta(\theta_{t+1}|\theta_t) \frac{u_c(\theta_{t+1}) + u_y(\theta_{t+1})\,\alpha(\theta_{t+1})}{1 - \alpha(\theta_{t+1})} \end{aligned} \tag{34}$$

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:

$$\frac{u_c(\theta_t) + u_y(\theta_t)\,\alpha(\theta_t)}{1 - \alpha(\theta_t)} > 0 \tag{35}$$

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

$$\begin{aligned} & \sum_{\theta_{t+1} \in \Theta} \pi_\Theta(\theta_{t+1}|\theta_t) \left[\frac{u_c(\theta_{t+1}) + u_y(\theta_{t+1})\,\alpha(\theta_{t+1})}{1 - \alpha(\theta_{t+1})}\right]^{-1} \\ \geq \quad & \left[\sum_{\theta_{t+1} \in \Theta} \pi_\Theta(\theta_{t+1}|\theta_t) \frac{u_c(\theta_{t+1}) + u_y(\theta_{t+1})\,\alpha(\theta_{t+1})}{1 - \alpha(\theta_{t+1})}\right]^{-1} \end{aligned} \tag{36}$$

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 θ_{t+1}. The left-hand side of (36) is also the right-hand side of equation (19); the inequality in the Proposition then follows from using that equation in (36). ■

Note that this result has been stated irrespective of the manner and extent to which income is being taxed *within* periods t and $t+1$: in particular, 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. Moreover, note also that we are *not* able to state a similar 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 θ_{t+1}, preventing us from applying Jensen's inequality.[18]

Moving away from its implications for marginal tax wedges, it will also be interesting to consider what Proposition 8 implies for the 'immiseration'-type results that emerge in the special case that $R_t = \beta^{-1}$ for all t. In that case equation (19) 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.

Intermediate cases We have shown 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 local movements in either direction along the within-period indifference curve of the highest-type agent are always incentive-compatible 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 appled. 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 δ units for all realisations of the productivity parameter at $t+1$, in both cases in a manner that is locally incentive conpatible and feasible. These perturbations will raise

[18] In fact subsequent results (specifically Proposition 28) 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.

within-period utility for all contemporary realisations of θ by δ units in the period that the relevant $\Delta(\delta)$ schedule is applied. This sub-section shows that it is also possible to find a set of perturbations that is 'intermediate' between these two extremes, in the sense that these can raise the within-period utility of the top n types by δ units in the period that the relevant $\Delta(\delta)$ schedule is applied whilst holding constant the utility of all other types, for all $n \in \{1, ..., N-1\}$.

The intuition that we exploit is the following. Suppose one were to perturb the within-period allocation at time t of an agent with prior reporting history $\widehat{\theta}^{t-1}$ and whose current report is $\widehat{\theta}_t^n$ (indexing by ranking in Θ), in a manner that keeps the utility of this type constant – that is, by a movement along this agent's within-period indifference curve in consumption-output space. If the conditions set out in Proposition 6 hold then we can preserve *within-period* incentive compatibility for small such movements provided we simultaneously change the utility at t of all *higher-type* agents with the same prior reporting history $\widehat{\theta}^{t-1}$, from the $n+1$th to the Nth, by an amount equal to the change in the utility that the $n+1$th agent can obtain by mimicking the nth. 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 – so in the case of separability, for instance, via consumption increments alone. Coupled with these changes must be a perturbation in period $t-1$ to the utility of an agent with reporting history $\widehat{\theta}^{t-1}$, so as to keep the present value of reporting $\widehat{\theta}^{t-1}$ constant in that period – with this perturbation again being constructed to ensure equal utility increments for truth-tellers and mimickers.

For each potential choice of n in the above argument, the set of time-t perturbations contained in the relevant $\Delta(\delta)$ function will hold equilibrium consumption and output levels constant for all agents of type θ_t^{n-1} and lower, whilst changing them for type θ_t^n and higher. This implies the marginal impact on that $\Delta(\delta)$ function of moving δ away from zero must differ in the choice of n, which in turn implies we will have a further set of $N-1$ linearly independent dimensions along which the optimal allocation can be perturbed at the margin whilst preserving incentive compatibility locally – providing $N-1$ distinct optimality conditions. The next proposition sets out the argument in analytical form. First, it is useful to define $\tau(c, y, \theta) \equiv 1 + \frac{u_y(c,y,\theta)}{u_c(c,y,\theta)}$, which is the implicit within-period marginal income tax rate faced by an agent of type θ receiving an allocation (c, y). Then we have:

Proposition 11 *Suppose the optimal solution to the 'relaxed' iid problem in which only constraints (12) are imposed has the three properties described in Proposition 6. Then for all time periods $t \geq 1$, all reporting histories $\widehat{\theta}^t$, and all $\theta_{t+1}^n \in \Theta : \theta_{t+1}^n \neq \max\left\{\theta' \in \Theta\right\}$ (so $n < N$), the allocations $\left(c_t^*\left(\widehat{\theta}^t\right), y_t^*\left(\widehat{\theta}^t\right)\right)$*

and $\Psi^*_{t+1}\left(\widehat{\theta}^t\right)$ *satisfy the following condition:*[19]

$$\begin{aligned} & -\pi_\Theta\left(\theta^n_{t+1}|\theta_t\right)\frac{\tau\left(\theta^n_{t+1}\right)}{u_c\left(\widehat{\theta}^n_{t+1};\theta^{n+1}_{t+1}\right)\left(1-\tau\left(\theta^n_{t+1}\right)\right)+u_y\left(\widehat{\theta}^n_{t+1};\theta^{n+1}_{t+1}\right)} \quad (37)\\ & +\sum_{m=n+1}^{N}\pi_\Theta\left(\theta^m_{t+1}|\theta_t\right)\frac{1-\alpha\left(\theta^m_{t+1}\right)}{u_c\left(\theta^m_{t+1}\right)+u_y\left(\theta^m_{t+1}\right)\alpha\left(\theta^m_{t+1}\right)} \\ = & \ \beta R_{t+1}\pi_\Theta\left(\theta_{t+1}>\theta^n_{t+1}|\theta_t\right)\frac{1-\alpha\left(\theta_t\right)}{u_c\left(\theta_t\right)+u_y\left(\theta_t\right)\alpha\left(\theta_t\right)}\end{aligned}$$

Proof. 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 δ. The nth row is then constructed to equal $\left(\varphi^c\left(\theta^n_{t+1},\delta;c^*_{t+1},y^*_{t+1}\right),\varphi^y\left(\theta^n_{t+1},\delta;c^*_{t+1},y^*_{t+1}\right)\right)$, where these values are defined implicitly (and uniquely) by the following two equations:

$$u\left(c+\varphi^c\left(\theta,k;c,y\right),y+\varphi^y\left(\theta,k;c,y\right);\theta\right)=u\left(c,y;\theta\right) \quad (38)$$

$$u\left(c+\varphi^c\left(\theta,k;c,y\right),y+\varphi^y\left(\theta,k;c,y\right);\theta'\right)=u\left(c,y;\theta'\right)+k \quad (39)$$

with $\theta'=\min\left\{\theta''\in\Theta:\theta''>\theta\right\}$. (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 θ, whilst increasing it by k units for an agent whose type is one higher.) The mth row in $\Delta(\delta)$ is given for $m\in\{n+1,...,N\}$ by $\left(\phi^c\left(\theta^m_{t+1},\delta;c^*_{t+1},y^*_{t+1}\right),\phi^y\left(\theta^m_{t+1},\delta;c^*_{t+1},y^*_{t+1}\right)\right)$, where these functions are defined in the proof of Proposition 8. We additionally assume a time-t perturbation schedule $\Delta_{-1}(\delta)$ given by:

$$\begin{aligned} & \Delta_{-1}(\delta) \quad (40)\\ = & \left(\phi^c\left(\theta_t,-\beta\delta\pi_\Theta\left(\theta_{t+1}>\theta^n_{t+1}|\theta_t\right);c^*_t,y^*_t\right),\phi^y\left(\theta_t,-\beta\delta\pi_\Theta\left(\theta_{t+1}>\theta^n_{t+1}|\theta_t\right);c^*_t,y^*_t\right)\right)\end{aligned}$$

For values of δ close enough to zero these perturbations will preserve local incentive compatibility at $t+1$. For the nth 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 nth see no change to their allocations. For agents whose types are higher than the nth, 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 other constraints will remain strictly satisfied for small enough values of δ.

The perturbations will also preserve incentive compatibility at t for small enough values of δ. Given the iid assumption, the discounted value of the $\Delta(\delta)$

[19] We suppress the dependence of τ, α, u_c and u_y on equilibrium allocations to keep the notation manageable; $u\left(\widehat{\theta}^n_{t+1};\theta^{n+1}_{t+1}\right)$ is then used to denote the utility an agent of type θ^{n+1}_{t+1} can obtain by mimicking an agent of type θ^n_{t+1}. Superscripts on productivity parameters denote their ranking within Θ as before.

perturbation at time t is $\beta\delta\pi_\Theta\left(\theta_{t+1} > \theta^n_{t+1}|\theta_t\right)$ to *both* the agent of type θ_t and the agent whose type is one higher and chooses to mimic (that is, δ units of extra utility received if and only if one's type exceeds θ^n_{t+1}). By perturbing the t-dated utility received by both truth-teller and mimicker by $-\beta\delta\pi_\Theta\left(\theta_{t+1} > \theta^n_{t+1}|\theta_t\right)$ 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, for small enough δ.

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:

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

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

$$\begin{aligned}
&R_t^{-1}\left\{\pi_\Theta\left(\theta^n_{t+1}|\theta_t\right)\left[\varphi^c_2\left(\theta^n_{t+1},0\right) - \varphi^y_2\left(\theta^n_{t+1},0\right)\right]\right. \\
&\left.+\sum_{m=n+1}^{N}\pi_\Theta\left(\theta^m_{t+1}|\theta_t\right)\left[\phi^c_2\left(\theta^m_{t+1},0\right) - \phi^y_2\left(\theta^m_{t+1},0\right)\right]\right\} \\
&-\beta\pi_\Theta\left(\theta_{t+1} > \theta^n_{t+1}|\theta_t\right)\left[\phi^c_2\left(\theta_t,0\right) - \phi^y\left(\theta_t,0\right)\right] \\
=\;& 0
\end{aligned} \tag{41}$$

Again, differentiating the equations defining ϕ we can show:

$$\phi^c_2(\theta,0) - \phi^y_2(\theta,0) = \frac{1-\alpha(\theta)}{u_c(\theta) + u_y(\theta)\alpha(\theta)} \tag{42}$$

and total differentiation of (38) and (39) gives:

$$\begin{aligned}
\varphi^c_2(\theta,0) - \varphi^y_2(\theta,0) &= \frac{1 + \frac{u_c(\theta)}{u_y(\theta)}}{u_c\left(\widehat{\theta};\theta'\right) - u_y\left(\widehat{\theta};\theta'\right)\frac{u_c(\theta)}{u_y(\theta)}} \\
&= -\frac{\tau(\theta)}{u_c\left(\widehat{\theta};\theta'\right)(1-\tau(\theta)) + u_y\left(\widehat{\theta};\theta'\right)}
\end{aligned} \tag{43}$$

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

$$-\pi_{\Theta}\left(\theta_{t+1}^{n}|\theta_{t}\right)\frac{\tau\left(\theta_{t+1}^{n}\right)}{u_{c}\left(\widehat{\theta}_{t+1}^{n};\theta_{t+1}^{n+1}\right)\left(1-\tau\left(\theta_{t+1}^{n}\right)\right)+u_{y}\left(\widehat{\theta}_{t+1}^{n};\theta_{t+1}^{n+1}\right)} \tag{44}$$

$$+\sum_{m=n+1}^{N}\pi_{\Theta}\left(\theta_{t+1}^{m}|\theta_{t}\right)\frac{1-\alpha\left(\theta_{t+1}^{m}\right)}{u_{c}\left(\theta_{t+1}^{m}\right)+u_{y}\left(\theta_{t+1}^{m}\right)\alpha\left(\theta_{t+1}^{m}\right)}$$

$$= \beta R_{t+1}\pi_{\Theta}\left(\theta_{t+1}>\theta_{t+1}^{n}|\theta_{t}\right)\frac{1-\alpha\left(\theta_{t}\right)}{u_{c}\left(\theta_{t}\right)+u_{y}\left(\theta_{t}\right)\alpha\left(\theta_{t}\right)}$$

QED. ■

An interesting feature of this condition is that it nests our earler two as extreme cases. Suppose we set $n=0$, and extend the set Θ to include some arbitrary element, denoted θ^{0}, which is strictly less than all other elements of Θ and whose probability in all periods is zero under the π_{Θ} measure. Then $\pi_{\Theta}\left(\theta_{t+1}^{0}|\theta_{t}\right)=0$ and $\pi_{\Theta}\left(\theta_{t+1}>\theta_{t+1}^{0}|\theta_{t}\right)=1$ by definition, and condition (37) reduces to the generalised version of the inverse Euler equation, (19). If, on the other hand, we set $n=N$ then the last two terms drop, and the condition reduces to a requirement that $\tau\left(\theta_{t+1}^{N}\right)=0$ – the 'zero distortion at the top' result from Proposition 7

It proves useful for the analysis of the non-iid case to elaborate on this further. If we define $\varphi^{c}\left(\theta',k;c,y\right)$ and $\varphi^{y}\left(\theta',k;c,y\right)$ for $\theta'=\max\left\{\theta\in\Theta\right\}$ by:

$$\varphi^{c}\left(\theta',k;c,y\right)=k \tag{45}$$

$$u\left(c+\varphi^{c}\left(\theta',k;c,y\right),y+\varphi^{y}\left(\theta',k;c,y\right);\theta'\right)=u\left(c,y;\theta'\right) \tag{46}$$

and φ_{2}^{c} and φ_{2}^{y} correspondingly, then the set of $\Delta\left(\delta\right)$ perturbations that we consider at $t+1$ is given by the $N+1$ matrices of the following form:

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

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

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

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

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

where $n \in \{1, ..., N+1\}$ 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 $\Delta_{-1}^{N+1}(\delta)$:

$$\Delta_{-1}^{N+1}(\delta) = [0, 0] \tag{47}$$

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

$$\begin{aligned} & \Delta_{-1}^N(\delta) \\ = \; & \left[\phi^c\left(\theta_t, -\beta\delta\pi_\Theta\left(\theta_{t+1} > \theta_{t+1}^{N-1}|\theta_t\right)\right), \phi^y\left(\theta_t, -\beta\delta\pi_\Theta\left(\theta_{t+1} > \theta_{t+1}^{N-1}|\theta_t\right)\right)\right] \end{aligned} \tag{48}$$

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

$$\Delta_{-1}^1(\delta) = \left[\phi^c\left(\theta_t, -\beta\delta\right), \phi^y\left(\theta_t, -\beta\delta\right)\right] \tag{49}$$

The marginal effects on allocations of moving δ 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 relevant to $t+1$, which we may denote $\Delta^{n\prime}(0)$ for $n \in \{1, ..., N+1\}$, are linearly independent from one another, and 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 11 are 'intermediate' between those of Propositions 7 and 8 is now apparent. A final useful 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 $\Delta^{n\prime}(0)$ for $n \in \{1, ..., N+1\}$. We define this matrix $\widehat{J}$:

$$\widehat{J} = \begin{bmatrix} 1 & 1 & ... & 1 & 1 \\ 0 & 1 & ... & 1 & 1 \\ ... & ... & ... & ... & ... \\ 0 & 0 & ... & 1 & 1 \\ 0 & 0 & ... & 0 & 1 \\ 0 & 0 & ... & 0 & 0 \end{bmatrix} \tag{50}$$

Element $\widehat{J}_{n,m}$ is then the marginal impact on the within-period utility of an agent of type θ_{t+1}^{m} caused by a marginal move in δ 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 δ in each case.

Now, note that a trivial implication of Propositions 8 and 11 together is that we should expect the following condition to hold at the optimum, for all $n \in \{1, ..., N\}$:[20]

$$\begin{aligned} & -\pi_{\Theta}\left(\theta_{t+1}^{n}|\theta_{t}\right)\frac{\tau\left(\theta_{t+1}^{n}\right)}{u_{c}\left(\widehat{\theta}_{t+1}^{n};\theta_{t+1}^{n+1}\right)\left(1-\tau\left(\theta_{t+1}^{n}\right)\right)+u_{y}\left(\widehat{\theta}_{t+1}^{n};\theta_{t+1}^{n+1}\right)} \\ & +\sum_{m=n+1}^{N}\pi_{\Theta}\left(\theta_{t+1}^{m}|\theta_{t}\right)\frac{1-\alpha\left(\theta_{t+1}^{m}\right)}{u_{c}\left(\theta_{t+1}^{m}\right)+u_{y}\left(\theta_{t+1}^{m}\right)\alpha\left(\theta_{t+1}^{m}\right)} \\ = \; & \pi_{\Theta}\left(\theta_{t+1}>\theta_{t+1}^{n}|\theta_{t}\right)\sum_{\theta_{t+1}\in\Theta}\pi_{\Theta}\left(\theta_{t+1}|\theta_{t}\right)\frac{1-\alpha\left(\theta_{t+1}\right)}{u_{c}\left(\theta_{t+1}\right)+u_{y}\left(\theta_{t+1}\right)\alpha\left(\theta_{t+1}\right)} \end{aligned} \tag{51}$$

Equation (51) is an extremely useful expression for analytical purposes, since it gives a relationship that must hold at the optimum between variables that are *entirely particular to one time period*.[21] In this way we are able to divide up the set of $N+1$ optimality conditions that we have derived into N *intra*temporal conditions, and just one *inter*temporal condition. Indeed, the precise set of N intratemporal optimality conditions need not be stated in the manner above. Denote by π_{Θ}^{vec} the N-dimensional vector stacking the probabilities $\pi_{\Theta}\left(\theta_{t+1}^{n}|\theta_{t}\right)$ in order from $n=1$ to $n=N$. Let γ be any $N \times 1$ vector with the following property:

$$\gamma' J \pi_{\Theta}^{vec} = 0 \tag{52}$$

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

Proposition 12 *Suppose the optimal solution to the 'relaxed' iid problem in which only constraints (12) are imposed has the three properties described in Proposition 6. Then for all time periods $t \geq 1$ and all reporting histories $\widehat{\theta}^{t}$, the matrices $\{\Delta^{n\prime}(0)\}_{n=1}^{N}$ associated with the optimal allocation matrix $\Psi_{t+1}^{*}\left(\widehat{\theta}^{t}\right)$ satisfy the following condition:*

$$\left(\pi_{\Theta}^{vec}\right)'\left[\sum_{n=1}^{N}\gamma_{n}\Delta^{n\prime}(0)\right]k=0 \tag{53}$$

where k is here defined as the 2×1 vector $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and γ_n is the nth element of any vector γ that satisfies equation (52).

[20] The case of $n=N$ reduces to the 'no distortion at the top' result.
[21] The iid assumption ensures $\pi_{\Theta}\left(\theta_{t+1}^{n}|\theta_{t}\right)$ is independent of θ_{t}.

Proof. Re-writing in vector notation the result of Proposition 11, each $\Delta^{n\prime}(0)$ must satisfy the condition:

$$\left(\pi_{\Theta}^{vec}\right)' \Delta^{n\prime}(0)\, k = c\lambda^{n\prime} J \pi_{\Theta}^{vec} \tag{54}$$

where λ^n is an $N \times 1$ selection vector, with zeroes in all entries except for a one in the nth, and c is a scalar constant (for the given $\widehat{\theta}^t$) equal to the left-hand side of (19). (This is just a restatement in vector notation of the result of Proposition 11.) Hence:

$$\gamma_n \left(\pi_{\Theta}^{vec}\right)' \Delta^{n\prime}(0)\, k = c\gamma_n \lambda^{n\prime} J \pi_{\Theta}^{vec} \tag{55}$$

$$\sum_{n=1}^{N} \gamma_n \left(\pi_{\Theta}^{vec}\right)' \Delta^{n\prime}(0)\, k = c\gamma' J \pi_{\Theta}^{vec} \tag{56}$$

The result follows from the definition of γ in equation (52). ■

Notice that there will, in general, exist $N-1$ linearly independent γ vectors satisfying equation (52). Equation (51) gives the $N-1$ possibilities for which $\gamma_1 = 1$, $\gamma_n = -1$ for some $n \in \{2, ..., N\}$, and there are zero entries elsewhere. When we move to the non-iid case the set of admissible γ vectors is reduced in an important way, and this matrix representation proves invaluable in characterising composite movements that *are* still possible in that case.

Proposition 12 has really just manipulated linear combinations of equation (37) to obtain its within-period set of optimality conditions. That equation was, in turn, derived by applying the marginal perturbations Δ^n and Δ^n_{-1} in periods $t+1$ and t respectively (for some $t \geq 1$ and any $n \in \{1, ..., N-1\}$). But in the non-iid case we will find that it is not necessarily possible to apply this particular pair of perturbations, since incentive compatibility in period t becomes a much more demanding requirement. For this reason it is useful to show that the same result can be obtained by *directly* considering a 'composite' perturbation that changes utility across types at $t+1$ according to a vector that differs from the rows of the matrix $\widehat{J}$.

Proposition 13 *Suppose the optimal solution to the 'relaxed' iid problem in which only constraints (12) are imposed has the three properties described in Proposition 6. Then for any vector $\nu \in \mathbb{R}^N$ (whose nth element is denoted ν_n), all time periods $t \geq 1$ and all reporting histories $\widehat{\theta}^t$, it is possible to perturb the optimal allocations $\left(c_t^*\left(\widehat{\theta}^t\right), y_t^*\left(\widehat{\theta}^t\right)\right)$ and $\Psi_{t+1}^*\left(\widehat{\theta}^t\right)$ in a manner that will preserve incentive compatibility in all periods whilst raising the within-period utility of an agent of type θ_{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 δ satisfying $|\delta| < \varepsilon$ for some $\varepsilon > 0$. Additionally, considering time 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 Ψ_1^* in a manner that will preserve within-period incentive compatibility whilst raising the within-period utility of an agent of type θ_1^n by an amount $\nu_n\delta$ in period 1, again for any scalar δ satisfying $|\delta| < \varepsilon$ for some $\varepsilon > 0$.*

Proof. By Proposition 6 we can ignore global incentive compatibility restrictions for perturbations that are sufficiently small. We need only then 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 δ in an open neighbourhood of 0. This requires that the following two conditions are satisfied at $t+1$ for all $n \in \{1, ..., N\}$:

$$u\left(c^*_{n,t+1} + \delta^c_n(\delta), y^*_{n,t+1} + \delta^y_n(\delta); \theta^n_{t+1}\right) = u\left(c^*_{n,t+1}, y^*_{n,t+1}; \theta^n_{t+1}\right) + \nu_n \delta \quad (57)$$

$$u\left(c^*_{n,t+1} + \delta^c_n(\delta), y^*_{n,t+1} + \delta^y_n(\delta); \theta^{n+1}_{t+1}\right) = u\left(c^*_{n,t+1}, y^*_{n,t+1}; \theta^{n+1}_{t+1}\right) + \nu_{n+1} \delta \quad (58)$$

where $\delta^c_n(\delta)$ and $\delta^y_n(\delta)$ are the perturbations to the nth agent's consumption and output levels respectively. For the Nth agent we just need:

$$u\left(c^*_{N,t+1} + \delta^c_N(\delta), y^*_{N,t+1}; \theta^N_{t+1}\right) = u\left(c^*_{N,t+1}, y^*_{N,t+1}; \theta^N_{t+1}\right) + \nu_N \delta \quad (59)$$

and we normalise $\delta^y_N(\delta) = 0$.[22]

Equations (57) and (59) 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 (58) 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 δ these equations must solve for unique values of $\delta^c_n(\delta)$ and $\delta^y_n(\delta)$ for all n. 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 $\widehat{\theta}^t$ is to increase it by an amount $\beta\nu'\pi^{vec}_\Theta\delta$. If $\nu'\pi^{vec}_\Theta = 0$ 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 θ_t and a mimicker whose type is one higher. This can be done through the perturbation $[\phi^c(\theta_t, -\beta\nu'\pi^{vec}_\Theta\delta), \phi^y(\theta_t, -\beta\nu'\pi^{vec}_\Theta\delta)]$, as already established. ■

This result is useful for two reasons. First, because it implies a set of 'intratemporal' 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.[23] Second, because for each 'utility increment' vector ν there will always be an equivalent γ, satisfying:

$$\gamma = \left(J^{-1}\right)' \nu \quad (60)$$

[22] This is analogous to the normalisation $\phi^y(\theta, k; c^*, y^*) = 0$ in equation (24).
[23] This is the purpose of stating separately the final sentence in the Proposition.

If one solves for the marginal effect on the policymaker's surplus at $t+1$ associated with the perturbation that changes within-period utility away from the optimum according to the vector ν (per unit increase in δ), we can show that this is equal to the effect on the policymaker's surplus of an additive combination of the $\Delta^{n\prime}(0)$ marginal perturbations for $n = \{1, ..., N\}$, in proportions corresponding to the entries in the associated γ.[24] Together with movements along the Nth 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 ν), the complete set of $\Delta^{n\prime}(0)$ marginal perturbations for $n \in \{1, ..., N+1\}$ will therefore span the entire set of $(N+1)$ dimensions along which the consumption and output of all agents can be perturbed at the margin from the optimal allocation Ψ^*_{t+1} without undermining within-period incentive compatibility at $t+1$. We summarise this result in the final Proposition of this section, the proof of which is relegated to the appendix.

Proposition 14 *Suppose the optimal solution to the 'relaxed' iid problem in which only constraints (12) are imposed has the three properties described in Proposition 6, and consider perturbations of the form oulined in Proposition 13, 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 $\Delta^{n\prime}(0)$ matrices for $n \in \{1, ..., N\}$:*

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

where γ_n is the nth element in the vector γ, which in turn satisfies:

$$\gamma = \left(J^{-1}\right)'\nu \tag{61}$$

Moreover, any incentive-compatible perturbation that changes allocations at t and $t+1$ across agents with a common reporting history to t must have marginal effects on allocations at $t+1$ that are expressible as a linear combination of the $\Delta^{n\prime}(0)$ marginal effects alone.

Proof. See appendix. ■

4.2.2 Optimal income tax rates

The results of this section allow us to demonstrate a further result with important economic implications.

Proposition 15 *Suppose the optimal solution to the 'relaxed' iid problem in which only constraints (12) are imposed has the three properties described in Proposition 6. Then for all time periods $t \geq 1$, all reporting histories $\widehat{\theta}^{t-1}$ and all $\theta^n_t \in \Theta$ the implicit marginal tax rate $\tau\left(\theta^n_t\right)$ satisfies $\tau\left(\theta^n_t\right) \geq 0$.*

[24] One needs to be a bit careful about the indexing here: the element γ_n would give the number of units of $\Delta^{n-1\prime}(0)$ applied, given the way we have defined $\Delta^n(\delta)$.

Proof. Consider the perturbation given by applying just the nth row of the schedule $\Delta^{n+1}(\delta)$ at time t – that is, a movement along the within-period indifference curve of the nth agent. For negative values of δ (only) in the neighbourhood of $\delta = 0$ this will preserve incentive compatibility at t, since the net impact on the utility obtainable from reporting $\widehat{\theta}^t$ at t is zero for truth-tellers and strictly negative for mimickers. Hence the marginal cost of such a perturbation as δ is increased marginally *below* zero must be weakly positive at any optimum. From our earlier results, this implies:

$$-\left[\varphi_2^c\left(\theta_{t+1}^n, 0\right) - \varphi_2^y\left(\theta_{t+1}^n, 0\right)\right] \geq 0 \tag{62}$$

The proof of Proposition 11 shows:

$$\varphi_2^c\left(\theta_t^n, 0\right) - \varphi_2^y\left(\theta_t^n, 0\right) = \frac{\tau\left(\theta_t^n\right)}{u_c\left(\widehat{\theta}_t^n; \theta_t^{n+1}\right)\left(1 - \tau\left(\theta_t^n\right)\right) + u_y\left(\widehat{\theta}_t^n; \theta_t^{n+1}\right)} \tag{63}$$

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

$$\frac{\tau\left(\theta_{t+1}^n\right)}{u_c\left(\widehat{\theta}_{t+1}^n; \theta_{t+1}^{n+1}\right)\left(1 - \tau\left(\theta_{t+1}^n\right)\right) + u_y\left(\widehat{\theta}_{t+1}^n; \theta_{t+1}^{n+1}\right)} \geq 0 \tag{64}$$

We have:

$$\begin{aligned} \left(1 - \tau\left(\theta_{t+1}^n\right)\right) &= -\frac{u_y\left(\theta_{t+1}^n\right)}{u_c\left(\theta_{t+1}^n\right)} \\ &> -\frac{u_y\left(\widehat{\theta}_{t+1}^n; \theta_{t+1}^{n+1}\right)}{u_c\left(\widehat{\theta}_{t+1}^n; \theta_{t+1}^{n+1}\right)} \end{aligned} \tag{65}$$

where the last inequality is an application of the single-crossing condition. Hence the denominator in condition (64) will be strictly positive, and the result follows. ■

So unlike the capital distortion the direction of the intratemporal distortion on production is unambiguous: in the iid case at least, the optimal effective marginal income tax rate is never negative.[25] In a sense this 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 result 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. 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 indfference curve has a slope $\left(\frac{dc}{dy}\right)$ that is less than 1.

[25] The result generalises straightforwardly to the Markov case, as shown below.

5 The non-iid case

5.1 Incentive compatibility when shocks are Markov

In the more realistic case that agent productivities are Markov rather than iid, the foregoing linear algebra proves very useful. But first we need to augment the arguments justifying a focus on local incentive compatibility constraints alone. Again, we want to know when condition (12), stipulating that incentive compatibility holds 'downwards', is sufficient for global incentive compatibility, condition (13). Recall Proposition 6 proved that if the optimum to the 'restricted' problem in the iid case satisfied two quite plausible conditions then incentive compatibility held with *strict* inequality away from 'downwards' comparisons between adjacent pairs. This implies small enough movements away from the iid assumption will not be problematic, but it is useful to say something a bit more substantial. To this end, we define $V_t\left(\widehat{\theta}^{t-1},\theta_t\right)$ as the maximum discounted lifetime utility that an agent with reporting history $\widehat{\theta}^{t-1}$ and current type θ_t can obtain,[26] and $V_t^{vec}\left(\widehat{\theta}^{t-1}\right)$ as an $N\times 1$ vector stacking the values of $V_t\left(\widehat{\theta}^{t-1},\theta_t\right)$, ordered in ascending values of θ_t. Similarly, let $\pi_{\Theta}^{vec}(\theta)$ be the vector of probabilities across Θ in a given time period for an agent whose type was θ in the previous time period. Then we can show the following:

Proposition 16 *Suppose productivity shocks follow a Markov transition process and consumption is a normal good. Then any direct revelation mechanism that the policymaker chooses that satisfies (12) will additionally satisfy (13), provided it has the following properties:*

1. *The constraints in (12) bind with equality.*
2. *Conditional upon $\widehat{\theta}^{t-1}$, $c_t\left(\widehat{\theta}^{\infty}\right)$ and $y_t\left(\widehat{\theta}^{\infty}\right)$ are strictly increasing in $\widehat{\theta}_t$ for all $t\geq 1$ and $\widehat{\theta}^{t-1}\in\Theta^{t-1}$.*
3. *Conditional upon $\widehat{\theta}^{t-1}$, $\frac{c_t\left(\widehat{\theta}^{t-1},\widehat{\theta}_t'\right)-c_t\left(\widehat{\theta}^{t-1},\widehat{\theta}_t''\right)}{y_t\left(\widehat{\theta}^{t-1},\widehat{\theta}_t'\right)-y_t\left(\widehat{\theta}^{t-1},\widehat{\theta}_t''\right)}<1$ for all distinct $\widehat{\theta}_t',\widehat{\theta}_t''\in\Theta$.*
4. *Conditional upon $\widehat{\theta}^{t-1}$, $\left[\pi_{\Theta}^{vec}(\theta_t^n)-\pi_{\Theta}^{vec}(\theta_t^m)\right]'\left[V_{t+1}^{vec}\left(\widehat{\theta}^{t-1},\widehat{\theta}_t^k\right)-V_{t+1}^{vec}\left(\widehat{\theta}^{t-1},\widehat{\theta}_t^{k-1}\right)\right]\geq 0$ holds for all $n\in\{2,...,N\}$, $m\in\{1,...,n\}$ and $k\in\{2,...,N\}$.*

In addition, condition (13) will hold with strict inequality unless the strategy $\widehat{\theta}_{a,t}^{\infty}$ involves reporting $\widehat{\theta}_t\in\{\max\{\theta\in\Theta:\theta<\theta_t\},\theta_t\}$

Proof. See appendix. ■

[26] Note that this value is defined independently of whether the agent's past reports were truthful.

The extra condition that we need here relative to the iid case (i.e., the fourth) is less natural than the first three. It states that an increase in an agent's true type at time t must increase the value of changing that agent's report at time t from $\widehat{\theta}_t^{k-1}$ to $\widehat{\theta}_t^{k}$. That is, relative to $\pi_{\Theta}^{vec}(\theta_t^m)$ the probability measure $\pi_{\Theta}^{vec}(\theta_t^n)$ must put disproportionately high weight on those states at $t+1$ for which the difference in value is greatest between those who reported $\widehat{\theta}_t^{k}$ at t and those who reported $\widehat{\theta}_t^{k-1}$. Note that if the entries in the vector $V_{t+1}^{vec}\left(\widehat{\theta}^{t-1}, \widehat{\theta}_t^{k}\right) - V_{t+1}^{vec}\left(\widehat{\theta}^{t-1}, \widehat{\theta}_t^{k-1}\right)$ were all equal to one another we would satisfy the restriction with equality: the entries in the vector $\pi_{\Theta}^{vec}(\theta_t^n) - \pi_{\Theta}^{vec}(\theta_t^m)$ must sum to zero. So this is a requirement of the proabability structure that higher types at t can expect to do *disproportionately* better from higher reports. This has clear echoes of the within-period single crossing condition, which ensures higher types do disproportionately better from production-consumption bundles that are higher, so seen in this light it is not too remarkable.

It should also be stressed that this is only a *sufficient* set of conditions for the solution to the restricted problem to be a solution to the unrestricted one, and alternatives may well be available. Again, the conditions of the Proposition can be checked after establishing the solution to the restricted problem, and the perturbation analysis that follows derives conditions that must be satisfied in the neighbourhood of this (restricted problem's) solution – which will also be necessary features of the unrestricted problem *if* the conditions of the Proposition are indeed satisfied.

5.2 Perturbation analysis

When shocks are Markov rather than iid we are faced with an extra dimension of complication. For agents with a given reporting history $\widehat{\theta}^{t-1}$ 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 $\widehat{\theta}^{t-1}$, with a type at time $t-1$ that is one higher than the relevant entry in $\widehat{\theta}^{t-1}$, 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 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 $\Delta(\delta): \mathbb{R} \to \mathbb{R}^{2N}$ and $\Delta_{-1}(\delta): \mathbb{R} \to \mathbb{R}^2$ that can be used to perturb the consumption and output allocations across all agents with a given reporting history $\widehat{\theta}^{t-1}$, at t and $t-1$ respectively, subject to these functions satisfying the three conditions set out at the start of Section 4.2: the preservation of incentive compatibility, continuous differentiability in δ 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 Δ^n and Δ^n_{-1} schedules defined above. But in certain regards the earlier analysis can go through essentially unchanged. We focus on these similarities with the iid problem before turning to the differences.

5.2.1 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 and (possibly) $t-1$ of an agent who was not envied at $t-1$. This could either be because $t=1$ or because the agent was of the highest possible type (and thus, by our maintained focus on the 'restricted problem', not envied) at $t-1$. We can state the following.

Proposition 17 *Suppose the optimal solution to the 'relaxed' Markov problem in which only constraints (12) are imposed has the four properties described in Proposition 16. Then for all time periods $t \geq 1$ and any reporting history $\widehat{\theta}^t$ whose terminal entry $\widehat{\theta}_t$ is the maximal element of Θ, θ^N, the matrices $\{\Delta^{n\prime}(0)\}_{n=1}^N$ associated with the optimal allocation matrix $\Psi^*_{t+1}\left(\widehat{\theta}^t\right)$ satisfy the following condition:*

$$\left(\pi^{vec}_{\Theta}\left(\theta^N_t\right)\right)' \left[\sum_{n=1}^{N} \gamma_n \Delta^{n\prime}(0)\right] k = 0 \tag{66}$$

where k is here defined as the 2×1 vector $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and γ_n is the nth element of any vector γ that satisfies the equation $\gamma' J \pi^{vec}_{\Theta}\left(\theta^N_t\right) = 0$.

Similarly in period 1 the matrices $\{\Delta^{n\prime}(0)\}_{n=1}^N$ associated with the initial optimal allocation matrix Ψ^*_1 satisfy the following condition:

$$\left(\pi^{vec}_{\Theta}\right)' \left[\sum_{n=1}^{N} \gamma_n \Delta^{n\prime}(0)\right] k = 0 \tag{67}$$

γ_n is the nth element of any vector γ that satisfies the equation $\gamma' J \pi^{vec}_{\Theta} = 0$ and π^{vec}_{Θ} is the initial (unconditional) probability vector across productivity draws.

The proof of these claims merely repeats the logic contained in Propositions 11 and 12 through 14, 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 either 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. With little difficulty we can additionally show that the two 'extreme' perturbations – those associated with the first and last rows of the $\widetilde{J}$ matrix above – carry over for all histories.

Proposition 18 *Suppose the optimal solution to the 'relaxed' Markov problem in which only constraints (12) are imposed has the four properties described in Proposition 16. Then in all time periods $t \geq 1$ and (if $t > 1$) for all reporting histories $\widehat{\theta}^{t-1}$, the optimal allocation (c_t^*, y_t^*) for the agent who reports $\widehat{\theta}_t'$ 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. Small movements along the indifference curve of the 'top' agent in period t have no impact on expected utility in previous periods from reporting $\widehat{\theta}^{t-1}$, irrespective of the probability distribution under which these expectations are taken. Hence the proof of Proposition 7 goes through as before. ■

Proposition 19 *Suppose the optimal solution to the 'relaxed' Markov problem in which only constraints (12) are imposed has the four properties described in Proposition 16. Then for all time periods $t \geq 1$ and for all reporting histories $\widehat{\theta}^t$, the allocations $\left(c_t^*\left(\widehat{\theta}^t\right), y_t^*\left(\widehat{\theta}^t\right)\right)$ and $\Psi_{t+1}^*\left(\widehat{\theta}^t\right)$ satisfy the generalised 'inverse Euler' condition, (19).*

Proof. Utility increments that are uniform across types at $t+1$ generate a change to *ex ante* expected utility for $t+1$ that are independent of the probability distribution under which these expectations are taken. Hence the dependence of this distribution on an agent's true type at t does not affect the marginal consequences to mimickers or truth-tellers of the perturbation outlined in Proposition 8, and the logic of that Proposition goes through as before. ■

So even in the Markov problem we have two canonical results: no distortion at the top, and an inverse Euler condition, with the latter generalised for non-separable preferences. As mentioned briefly above, the latter has some interesting implications in the event that $R = \beta^{-1}$, in which case martingale convergence results can be applied to it; we explore these after completing the characterisation of optimality. The generalised inverse Euler condition can again be used to obtain the same results on the optimal level of capital taxation to implement. We state these in the following two Propositions – the proofs of which are omitted, since they are identical to their counterparts from the iid case.

Proposition 20 *Suppose the optimal solution to the 'relaxed' Markov problem in which only constraints (12) are imposed has the four properties described in*

Proposition 16, and additionally that in all time periods $s \geq 1$ and for all reporting histories $\widehat{\theta}^s$ the allocations $\left(c_s^*\left(\widehat{\theta}^s\right), y_s^*\left(\widehat{\theta}^s\right)\right)$ imply $u_c(\theta_s) \geq u_y(\theta_s)$. Then for all time periods $t \geq 1$ and for all reporting histories $\widehat{\theta}^t$, the allocations $\left(c_t^*\left(\widehat{\theta}^t\right), y_t^*\left(\widehat{\theta}^t\right)\right)$ and $\Psi_{t+1}^*\left(\widehat{\theta}^t\right)$ will satisfy inequality (29) 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* $\theta_t = \max\{\theta \in \Theta\}$.

Proposition 21 *Suppose the optimal solution to the 'relaxed' Markov problem in which only constraints (12) are imposed has the four properties described in Proposition 16. Then for all time periods $t \geq 1$ and for all reporting histories $\widehat{\theta}^t$, 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 $\left(c_t^*\left(\widehat{\theta}^t\right), y_t^*\left(\widehat{\theta}^t\right)\right)$ and $\Psi_{t+1}^*\left(\widehat{\theta}^t\right)$ will satisfy inequality (34), 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 $\theta_{t+1} \in \Theta$.*

We confirm in Proposition 26 below that the condition $u_c(\theta_s) \geq u_y(\theta_s)$ will indeed be satisfied at any solution to the restricted problem, so the extra assumption made in the first of these Propositions is again not an onerous one.

As mentioned briefly above, the 'no distortion at the top' result appears on the surface to contrast with a numerical simulation by Kocherlakota (2011),[27] in which 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 Kocherlakota assumes. In the first period of his model, young agents may be either type θ_H (high type) of θ_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 $\theta_L\theta'_H$ 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, θ'_H, do not have the same preference structure. 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, then, influenced

[27] See Chapter 6 of that book.

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 θ_H (a point noted by the author). If we were to map from his setting to ours, the appropriate specification of Θ would be a time-varying set:[28] $\Theta = \{\theta_L, \theta_H\}$ in the first period, and $\Theta = \left\{\theta_L\theta'_L, \theta_L\theta'_H, \theta_H\theta'_L, \theta_H\theta'_H\right\}$ 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 Θ in the second period. This result (together with zero marginal rates for those of type θ_H) is indeed reported by Kocherlakota.

5.2.2 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 *intra*temporal 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 *inter*temporal condition arises, ensuring that the cost to the policymaker of preventing mimicking is spread optimally through time. We explain these points in turn.

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

[28] Nothing in the analysis above has precluded the possibility of a time-varying Θ. 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 Θ for some or all agents that put zero weight on some of its elements.

weight vector γ jointly satisfies *two* conditions:

$$\gamma' J \pi_{\Theta}^{vec}\left(\theta_t^n\right) = \gamma' J \pi_{\Theta}^{vec}\left(\theta_t^{n+1}\right) = 0 \tag{68}$$

In general one can always find $N-2$ linearly independent γ vectors for which this condition is satisfied: 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 γ vector will again be an equivalent ν vector that directly lists the marginal utility effects on agents at $t+1$, given by $\nu = J'\gamma$. Proposition 13 can then be easily adjusted to cover intratemporal perturbations in the Markov case:

Proposition 22 *Suppose the optimal solution to the 'relaxed' Markov problem in which only constraints (12) are imposed has the four properties described in Proposition 16. Then for all time periods $t \geq 1$, all reporting histories $\widehat{\theta}^t$ such that $\widehat{\theta}_t = \theta_t^n \neq \theta_t^N$, and any vector ν that satisfies $\nu' \pi_{\Theta}^{vec}\left(\theta_t^n\right) = \nu' \pi_{\Theta}^{vec}\left(\theta_t^{n+1}\right) = 0$ it is possible to perturb the optimal allocations $\Psi_{t+1}^*\left(\widehat{\theta}^t\right)$ in a manner that will preserve incentive compatibility in all periods whilst raising the within-period utility of an agent of type θ_{t+1}^n by an amount $\nu_n \delta$ at $t+1$ for any δ 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 Proposition 13, 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 in earlier periods (even though the basic perturbations $\left\{\Delta^n\left(\delta\right)\right\}_{n=1}^N$ and $\left\{\Delta_{-1}^n\left(\delta\right)\right\}_{n=1}^N$ may no longer be admissible), and thus that the *marginal* utility effects associated with them are implementable, and (since equal to zero in *ex ante* expectation) must come at zero marginal resource cost.

Note also that the proof of Proposition 14 will go through essentially unchanged when we restrict attention to the subset of the possible utility increment vectors ν 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 $\widehat{\theta}^t = \theta^t$ (such that $\widehat{\theta}_t = \theta_t^n$) will again be:

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

So even though weighted *pairwise* sums of the basic perturbations $\left\{\Delta^n\left(\delta\right)\right\}_{n=1}^N$ may no longer be compatible with incentive compatibility,[29] any composite perturbations that still *are* incentive compatible in the Markov case can also still

[29] For instance, the iid condition (51) is a pairwise sum of the marginal effects of the Δ^1 and Δ^n perturbations for some $n \in \{2, ..., N\}$, with weights $-\pi_{\Theta}\left(\theta_{t+1} > \theta_{t+1}^n | \theta_t\right)$ and 1 respectively.

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 follow:

Proposition 23 *Suppose the optimal solution to the 'relaxed' Markov problem in which only constraints (12) are imposed has the four properties described in Proposition 16. Then for all time periods* $t \geq 1$ *and any reporting history* $\widehat{\theta}^t$ *such that* $\widehat{\theta}_t = \theta_t^n \neq \theta_t^N$, *the matrices* $\{\Delta^{n\prime}(0)\}_{n=1}^N$ *associated with the optimal allocation matrix* $\Psi_{t+1}^*\left(\widehat{\theta}^t\right)$ *satisfy the following condition:*

$$(\pi_\Theta^{vec}(\theta_t^n))' \left[\sum_{m=1}^N \gamma_m \Delta^{m\prime}(0)\right] k = 0 \tag{69}$$

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

The proof again follows directly from earlier arguments so is omitted here. Together with the 'no distortion at the top' condition, it implies we have $N-1$ linearly independent optimality conditions that must hold across types within each time period. The generalised inverse Euler condition gives a further condition (in all periods except the first),[30] 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 after the first and for any reporting history that did not feature the maximal element of Θ in the preceeding period. The final step in our characterisation is to provide this missing equation.

Recall again the basic problem faced by the policymaker in Mirrleesian optimal income tax models. 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 simultaneously along a within-period indifference curve – one can ensure that the marginal benefits to higher types from mimicking are reduced (appealing to the single crossing condition), and thus that so too are the utility rents that they can extract from the policymaker. Seen in this light, the problem is one of resolving the trade-off between the provision of

[30] An intertemporal resource constraint can be thought of as substituting for a dynamic optimality condition in the first time period.

wasteful amounts of utility to higher types, and the use of wasteful tax wedges impeding the production of lower types.

When productivity shocks are Markov there is a third alternative available to the policymaker. Instead of reducing the (wasteful) quantity of utility provided to higher types by 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 $\widehat{\theta}^t$ such that $\widehat{\theta}_t = \theta_t^n \neq \theta_t^N$, 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 θ_t^{n+1} see a reduction in their expected utility from mimicking under the measure $\pi_\Theta\left(\cdot|\theta_t^{n+1}\right)$, whilst expected utility under the measure $\pi_\Theta\left(\cdot|\theta_t^{n}\right)$ remains unchanged. A simple application of the theory of the second best suggests there will be net benefits to distorting $t+1$ allocations to this end. First, it is useful to state the following:

Proposition 24 *Suppose the optimal solution to the 'relaxed' Markov problem in which only constraints (12) are imposed has the four properties described in Proposition 16. Then for all time periods $t \geq 1$, all reporting histories $\widehat{\theta}^t$ such that $\widehat{\theta}_t = \theta_t^n \neq \theta_t^N$, and any vector ν that satisfies $\nu' \pi_\Theta^{vec}\left(\theta_t^n\right) = 0$ and $\nu' \pi_\Theta^{vec}\left(\theta_t^{n+1}\right) = 1$ it is possible to perturb the optimal allocations $\left(c_t^*\left(\widehat{\theta}^t\right), y_t^*\left(\widehat{\theta}^t\right)\right)$ and $\Psi_{t+1}^*\left(\widehat{\theta}^t\right)$ in a manner that will preserve incentive compatibility in all periods whilst raising the within-period utility of an agent of type θ_{t+1}^n by an amount $\nu_n \delta$ at $t+1$ for any δ satisfying $|\delta| < \varepsilon$ for some $\varepsilon > 0$ and leaving utility in all other periods constant.*

Proof. As ever, we can ignore global incentive compatibility restrictions for perturbations that are sufficiently small, since the conditions of Proposition 16 are assumed to be satisfied. We need only then ensure that '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 $m \in \{1, ..., N\}$:

$$u\left(c_{m,t+1}^* + \delta_{m,t+1}^c(\delta), y_{m,t+1}^* + \delta_{m,t+1}^y(\delta); \theta_{t+1}^m\right) = u\left(c_{m,t+1}^*, y_{m,t+1}^*; \theta_{t+1}^m\right) + \nu_m \delta \tag{70}$$

$$u\left(c_{m,t+1}^* + \delta_{m,t+1}^c(\delta), y_{m,t+1}^* + \delta_{m,t+1}^y(\delta); \theta_{t+1}^{m+1}\right) = u\left(c_{m,t+1}^*, y_{m,t+1}^*; \theta_{t+1}^{m+1}\right) + \nu_{m+1} \delta \tag{71}$$

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

$$u\left(c_{N,t+1}^* + \delta_{N,t+1}^c(\delta), y_{N,t+1}^*; \theta_{t+1}^N\right) = u\left(c_{N,t+1}^*, y_{N,t+1}^*; \theta_{t+1}^N\right) + \nu_N \delta \tag{72}$$

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

The proof of Proposition 13 shows that these conditions can indeed be satisfied by appropriate choice of $\delta^c_{m,t+1}(\delta)$ and $\delta^y_{m,t+1}(\delta)$ schedules. There remains the problem of incentive compatibility 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 nth agent at t to such an extent that a mimicker's utility is reduced by an offsetting amount (whils, by definition, leaving the utility of a truth-teller unaffected in this period also). That requires $\delta^c_{n,t}(\delta)$ and $\delta^y_{n,t}(\delta)$ schedules that satisfy:

$$u\left(c^*_{n,t}+\delta^c_{n,t}(\delta), y^*_{n,t}+\delta^y_{n,t}(\delta); \theta^n_t\right) = u\left(c^*_{n,t}, y^*_{n,t}; \theta^n_t\right) \tag{73}$$

$$u\left(c^*_{n,t}+\delta^c_{n,t}(\delta), y^*_{n,t}+\delta^y_{n,t}(\delta); \theta^{n+1}_t\right) = u\left(c^*_{n,t}, y^*_{n,t}; \theta^{n+1}_t\right) - \beta\delta \tag{74}$$

Again, by the single crossing condition the utility of the agent of type θ^{n+1}_t changes monotonically as one moves along a lower-type agent's indifference curve, so for small enough δ in an open neighbourhood of $\delta = 0$ this is always possible. ■

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

Proposition 25 *Suppose the optimal solution to the 'relaxed' Markov problem in which only constraints (12) are imposed has the four properties described in Proposition 16. Then for all time periods* $t \geq 1$ *and any reporting history* $\widehat{\theta}^t$ *such that* $\widehat{\theta}_t = \theta^n_t \neq \theta^N_t$, *the marginal perturbation matrices* $\{\Delta^{n\prime}(0)\}^N_{n=1}$ *associated with the optimal* $t+1$ *allocation matrix* $\Psi^*_{t+1}\left(\widehat{\theta}^t\right)$ *together with the optimal* t *allocation pair* $\left(c^*_t\left(\widehat{\theta}^t\right), y^*_t\left(\widehat{\theta}^t\right)\right)$ *must satisfy the following condition:*

$$\begin{aligned} &\beta R_{t+1}\frac{\tau(\theta^n_t)}{u_c\left(\widehat{\theta}^n_t;\theta^{n+1}_t\right)(1-\tau(\theta^n_t)) + u_y\left(\widehat{\theta}^n_t;\theta^{n+1}_t\right)} \\ &= \left(\pi^{vec}_\Theta(\theta^n_t)\right)'\left[\sum_{m=1}^N \gamma_m \Delta^{m\prime}(0)\right]k \end{aligned} \tag{75}$$

where k *is the* 2×1 *vector* $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ *and* γ_m *is the* m*th element of any vector* γ *that satisfies the two restrictions* $\gamma' J \pi^{vec}_\Theta(\theta^n_t) = 0$ *and* $\gamma' J \pi^{vec}_\Theta\left(\theta^{n+1}_t\right) = -1$.

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 θ^m_{t+1} by an amount $\nu_m\delta$ at $t+1$, where ν_m is the mth entry of the vector $\nu = J'\gamma$. By earlier arguments (c.f. proof of Proposition 14), the marginal cost of this $\Delta(\delta)$ perturbation as δ is moved away from 0, assessed from the perspective of time t, will be:

$$R^{-1}_{t+1}\pi_\Theta\left(\theta^t\right)\left(\pi^{vec}_\Theta(\theta^n_t)\right)'\left[\sum_{m=1}^N \gamma_m \Delta^{m\prime}(0)\right]k$$

By Proposition 24 we know incentive compatibility will be preserved for movements in the neighbourhood of $\delta = 0$ if allocations at t are perturbed by movements along the indifference curve of the relevant truth-telling agent, by an amount sufficient to increase the utility of a mimicker by $\beta\delta$ units. Retaining earlier definitions of the functions φ^c and φ^y, the cost of this perturbation, assessed at time t, will be:

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

and so the marginal cost as δ is moved away from zero is:

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

which we have already established (c.f. proof of Proposition 11) is equal to:

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

The result then follows from the fact that the total present value of the marginal cost of the perturbation must be zero at an interior 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 (75) explains 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.[31] 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 17 shows, if $t = 1$ there would be no incentives to set a value for $\left(\pi_\Theta^{vec}\left(\theta_{t-1}^n\right)\right)'\left[\sum_{m=1}^N \gamma_m \Delta^{m\prime}(0)\right] k$ different from zero (given $\gamma' J \pi_\Theta^{vec}\left(\theta_{t-1}^n\right) = 0$), 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 suggest greater equality at $t+1$ is desirable the higher is the marginal tax rate for an agent at t. This is because, as just discussed, higher marginal rates are really a

[31] Kapička (2010) makes a similar observation when using an efficient ('first-order') value function method to study a specific example of a dynamic Mirrleesian model.

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, for instance, would be the case if a monotone likelihood ratio existed between the pair of functions $\pi_{\Theta}\left(\cdot|\theta_t^n\right)$ and $\pi_{\Theta}\left(\cdot|\theta_t^{n+1}\right)$), 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 the higher are initial tax rates (and thus the greater is the distortion the policymaker is willing to accept).

5.2.3 Optimal income tax rates in the Markov case

It is easy to show that the optimal within-period 'income tax' wedge between the marginal utility of consumption and marginal disutility of production, $\tau\left(\theta_t^n\right)$, must again be weakly positive for all agents in the Markov case.

Proposition 26 *Suppose the optimal solution to the 'relaxed' Markov problem in which only constraints (12) are imposed has the four properties described in Proposition 16. Then for all time periods* $t \geq 1$, *all reporting histories* $\widehat{\theta}^{t-1}$ *and all* $\theta_t^n \in \Theta$ *the implicit marginal tax rate* $\tau\left(\theta_t^n\right)$ *satisfies* $\tau\left(\theta_t^n\right) \geq 0$.

The proof is omitted, as it is identical to the iid case (c.f. Proposition 15) – deriving from the possibility of unidirectional movements down within-period indifference curves that induce a *strict* preference against mimicking on the part of higher types.

In keeping with many of the other results from this section, this Proposition is interesting chiefly because it highlights the limits to which a movement away from iid transition probabilities affects optimal policy. Quantitatively, marginal tax rates are bound to be different when the process driving productivities is changed; but we have established here that the main *qualitative* result relating to within-period marginal income tax rates endures. In a similar vein, we have shown that the optimal marginal rate remains zero at the top, and that optimal savings wedges must satisfy an identical optimality condition in the iid and Markov cases.

6 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 β. Then the generalised inverse Euler equation can be written as:

$$\begin{aligned}&\frac{1-\alpha\left(\theta_{t}\right)}{u_{c}\left(\theta_{t}\right)+u_{y}\left(\theta_{t}\right)\alpha\left(\theta_{t}\right)} \\ &=\sum_{\theta_{t+1}\in\Theta}\pi_{\Theta}\left(\theta_{t+1}|\theta_{t}\right)\frac{1-\alpha\left(\theta_{t+1}\right)}{u_{c}\left(\theta_{t+1}\right)+u_{y}\left(\theta_{t+1}\right)\alpha\left(\theta_{t+1}\right)}\end{aligned} \tag{76}$$

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, $\alpha\left(\theta_{t}\right)=0$ 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[32]), 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 (76) – the marginal cost of utility provision – for preference structures more general than the separable case. A first step is the following.

Lemma 27 *Under an optimal plan that solves the restricted problem,* $u_{c}\left(\theta_{t}\right)+u_{y}\left(\theta_{t}\right)\alpha\left(\theta_{t}\right)>0$ *always holds.*

Proof. By definition

$$\alpha\left(\theta_{t}^{n}\right)=\frac{u_{c}\left(\theta_{t}^{n}\right)-u_{c}\left(\widehat{\theta}_{t}^{n};\theta_{t}^{n+1}\right)}{u_{y}\left(\widehat{\theta}_{t}^{n};\theta_{t}^{n+1}\right)-u_{y}\left(\theta_{t}^{n}\right)} \tag{77}$$

for $n<N$, and $\alpha\left(\theta_{t}^{N}\right)=0$. In the latter case the result follows immediately from $u_{c}\left(\theta_{t}\right)>0$. In the former case we have from equation (5):

$$\frac{u_{y}\left(\widehat{\theta}_{t}^{n};\theta_{t}^{n+1}\right)}{u_{y}\left(\theta_{t}^{n}\right)}<\frac{u_{c}\left(\widehat{\theta}_{t}^{n};\theta_{t}^{n+1}\right)}{u_{c}\left(\theta_{t}^{n}\right)} \tag{78}$$

[32] 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.

Rewriting our object of interest, we have:

$$u_c(\theta_t) + u_y(\theta_t)\,\alpha(\theta_t) = \frac{u_c\left(\widehat{\theta}_t^n;\theta_t^{n+1}\right) - u_c(\theta_t^n)\,\frac{u_y\left(\widehat{\theta}_t^n;\theta_t^{n+1}\right)}{u_y(\theta_t^n)}}{1 - \frac{u_y\left(\widehat{\theta}_t^n;\theta_t^{n+1}\right)}{u_y(\theta_t^n)}} \tag{79}$$

The numerator of the right-hand side is clearly positive by the preceeding 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. When they are Edgeworth complements Proposition 26 becomes very useful. Taken together we have the following result.

Proposition 28 $\frac{1-\alpha(\theta_t)}{u_c(\theta_t)+u_y(\theta_t)\alpha(\theta_t)} > 0$ *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 the preceeding Lemma. 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 the marginal cost of utility provision, 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 θ^∞ 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 29 $\frac{1-\alpha(\theta_t)}{u_c(\theta_t)+u_y(\theta_t)\alpha(\theta_t)} \overset{a.s.}{\to} 0$ *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. 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 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 tth 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.[33] 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.[34]

Perhaps the more interesting 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 surprising fact we are able to confirm through a finite-horizon computed example, the details of which we now present.

6.1 Computed example

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

$$u(c, y; \theta) = \frac{c^{1-\varsigma}}{1-\varsigma} \exp\left\{(\varsigma - 1) v\left(\frac{y}{\theta}\right)\right\} \tag{80}$$

with the labour disutility schedule v defined by:

$$v(l) = \frac{l^{1+v}}{1+v} \tag{81}$$

This function implies that consumption and labour supply are Edgeworth complements provided $\varsigma > 1$, and are Edgeworth substitues 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

[33] This latter question is explored in detail by Farhi and Werning (2007).

[34] 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 polictically plausible than the former.

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 Θ 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): θ_L and θ_H, with $\theta_L < \theta_H$. Transition probabilities are denoted as follows:

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

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 ς,[35] 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 inital 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 1 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.[36] What is of more interest is that the marginal cost of utility provision (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 θ_L.

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

[35] High values of ς 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 or production on the part of truthtellers), 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.

[36] Roughly three fifths of agents draw the same type in all six periods.

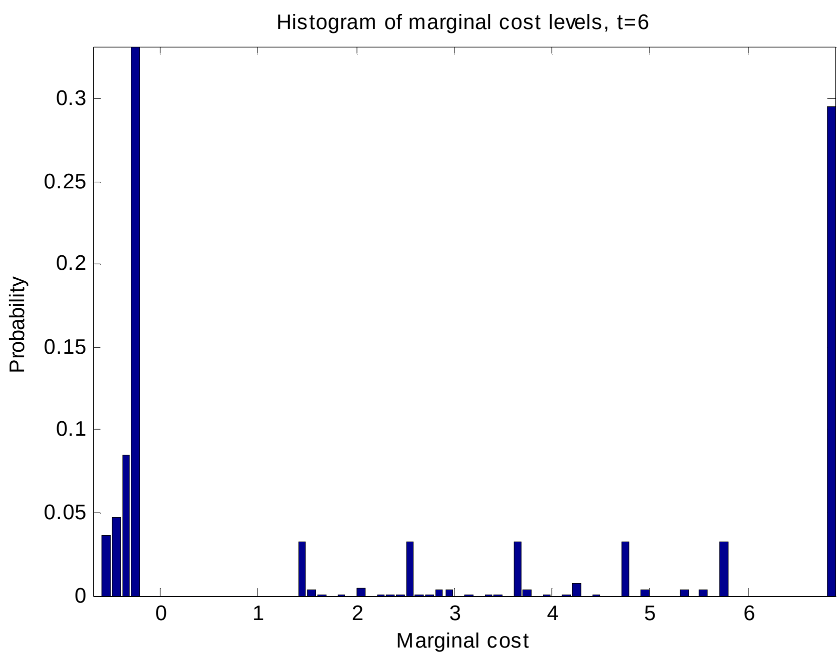

Figure 1: Distribution of marginal cost of utility provision in 6th period

level would benefit low types by more than (mimicking) high types, since the latter supply less labour to produce the given quantity of output. 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 low types than high (the former are already working longer hours, so their marginal disutility of effort is greater), offsetting the utility imbalance. 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 1 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 a 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 same 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 agents whost type was previously low. 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, 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.

7 A last word on savings wedges and immiseration

The results of the previous section – in particular Proposition 28 – 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 30 *Suppose the optimal solution to the 'relaxed' iid problem in which only constraints (12) are imposed has the three properties described in Proposition 6. Then for all time periods* $t \geq 1$ *and for all reporting histories* $\widehat{\theta}^t$*, if consumption and labour supply are Edgeworth complements then savings will be deterred at the optimum, in the sense that the allocations* $\left(c_t^*\left(\widehat{\theta}^t\right), y_t^*\left(\widehat{\theta}^t\right)\right)$ *and* $\Psi_{t+1}^*\left(\widehat{\theta}^t\right)$ *will satisfy inequality (34), 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 10, 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 know to be the case under complementarity and iid productivity draws by Proposition 28. 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 implicity choosing to 'front-load utility' in expectation. This is just a direct reading of inequality (34). 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.

8 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 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 higher-type agent mimicking the former would also receive one extra utility unit. 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 marginal 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 Markov.

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 β. 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 should note that the methods used in this paper can be applied more widely, albeit with some adaptation. For instance, a companion paper outlines a similar perturbation method applicable to dynamic Mirrlees problems in which the type space Θ 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] 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).

[5] Dasgupta, P. (1982), 'Utilitarianism, Information and Rights', in A. Sen and B. Williams (eds.), *Utilitarianism and Beyond*, Cambridge, Cambridge University Press.

[6] Diamond, P.A. and J.A. Mirrlees (1978), 'A Model of Social Insurance with Variable Retirement', *Journal of Public Economics*, 10, 295–336.

[7] Farhi, E. and I. Werning (2007), 'Inequality and Social Discounting', *Journal of Political Economy*, 115, 365–402.

[8] Fernandes, A. and C. Phelan (2000), 'A Recursive Formulation for Repeated Agency with History Dependence', *Journal of Economic Theory*, 91, 223–247.

[9] Golosov, M., N. Kocherlakota and A. Tsyvinski (2003), 'Optimal Indirect and Capital Taxation', *Review of Economic Studies*, 70, 569–587.

[10] Golosov, M., M. Troshkin and A. Tsyvinski (2011), 'Optimal Taxation: Merging Micro and Macro Approaches', *Journal of Money, Credit and Banking*, 43, 147–174.

[11] Golosov, A. Tsyvinski and I. Werning (2006), 'New Dynamic Public Finance: A User's Guide', in *NBER Macroeconomic Annual 2006*, Cambridge MA, MIT Press.

[12] 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*.

[13] Kocherlakota, N.R. (1996), 'Implications of Efficient Risk Sharing without Commitment', *Review of Economic Studies*, 63, 595–609.

[14] Kocherlakota, N.R. (2005), 'Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation', *Econometrica*, 73, 1587–1621.

[15] Kocherlakota, N.R. (2011), *The New Dynamic Public Finance*, Princeton, Princeton University Press.

[16] Marcet, A. and R, Marimon (1998), 'Recursive Contracts', Manuscript, Universitat Pompeu Fabra.

[17] Messner, M., N. Pavoni and C.M. Sleet (2011), 'Recursive Methods for Incentive Problems', Manuscript (available at SSRN: http://ssrn.com/abstract=1792816).

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

[19] Phelan, C. (2006), 'Opportunity and Social Mobility', *Review of Economic Studies*, 73, 487–505.

[20] Saez, E. (2001), 'Using Elasticities to Derive Optimal Income Tax Rates', *Review of Economic Studies*, 68, 205–229.

[21] 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).

[22] 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).

[23] 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 Proposition 6

We show first that equation (12) implies no agent will wish to mimic *any* type lower than their own. If shocks are iid the maximised value of utility from $t+1$ on for two agents with common reports to time t, $\widehat{\theta}^t$, must be identical, and independent of the agents' current types. We suppress dependence on past reports, and denote this utility $U_{t+1}\left(\widehat{\theta}_t\right)$. It is also useful to index the N elements of Θ by superscripts, with θ_t^n increasing in n. We wish to show:

$$\begin{aligned} & u\left(c_t\left(\widehat{\theta}_t^{n-m}\right), y_t\left(\widehat{\theta}_t^{n-m}\right); \theta_t^n\right) + \beta U_{t+1}\left(\widehat{\theta}_t^{n-m}\right) \\ \leq \quad & u\left(c_t\left(\widehat{\theta}_t^{n}\right), y_t\left(\widehat{\theta}_t^{n}\right); \theta_t^n\right) + \beta U_{t+1}\left(\widehat{\theta}_t^{n}\right) \end{aligned} \tag{82}$$

for all $n \leq N$ and $1 \leq m < n$, with a strict inequality holding whenever $m > 1$ (consistent with the final claim in the proposition). Equation (12) imposed with equality and summed across intermediate types gives:

$$\begin{aligned} & \beta\left(U_{t+1}\left(\widehat{\theta}_t^{n-m}\right) - U_{t+1}\left(\widehat{\theta}_t^{n}\right)\right) \\ = \quad & \sum_{k=0}^{m-1}\left[u\left(c_t\left(\widehat{\theta}_t^{n-k}\right), y_t\left(\widehat{\theta}_t^{n-k}\right); \theta_t^{n-k}\right) - u\left(c_t\left(\widehat{\theta}_t^{n-k-1}\right), y_t\left(\widehat{\theta}_t^{n-k-1}\right); \theta_t^{n-k}\right)\right] \end{aligned} \tag{83}$$

This implies it is sufficient for us to show:

$$\begin{aligned}
& u\left(c_t\left(\widehat{\theta}_t^n\right), y_t\left(\widehat{\theta}_t^n\right); \theta_t^n\right) - u\left(c_t\left(\widehat{\theta}_t^{n-m}\right), y_t\left(\widehat{\theta}_t^{n-m}\right); \theta_t^n\right) \\
\geq & \sum_{k=0}^{m-1}\left[u\left(c_t\left(\widehat{\theta}_t^{n-k}\right), y_t\left(\widehat{\theta}_t^{n-k}\right); \theta_t^{n-k}\right) - u\left(c_t\left(\widehat{\theta}_t^{n-k-1}\right), y_t\left(\widehat{\theta}_t^{n-k-1}\right); \theta_t^{n-k}\right)\right]
\end{aligned} \tag{84}$$

or

$$\begin{aligned}
& u\left(c_t\left(\widehat{\theta}_t^{n-m}\right), y_t\left(\widehat{\theta}_t^{n-m}\right); \theta_t^n\right) - u\left(c_t\left(\widehat{\theta}_t^{n-m}\right), y_t\left(\widehat{\theta}_t^{n-m}\right); \theta_t^{n-m}\right) \\
\leq & \sum_{k=n-m}^{n-1}\left[u\left(c_t\left(\widehat{\theta}_t^k\right), y_t\left(\widehat{\theta}_t^k\right); \theta_t^{k+1}\right) - u\left(c_t\left(\widehat{\theta}_t^k\right), y_t\left(\widehat{\theta}_t^k\right); \theta_t^k\right)\right]
\end{aligned} \tag{85}$$

which says that the t-period utility gain associated with moving from an actual type of θ_t^{n-m} to a type of θ_t^n at the consumption-output bundle associated with a report $\widehat{\theta}_t^{n-m}$ must be less than the aggregate of the utility gains associated with moving from type θ_t^k to θ_t^{k+1} at the bundle obtained by reporting $\widehat{\theta}_t^k$, summed across all k between $n-m$ and $n-1$. This, in turn, requires:

$$\begin{aligned}
& \sum_{k=n-m}^{n-1}\int_{\theta_t^k}^{\theta_t^{k+1}}\left[u_\theta\left(c_t\left(\widehat{\theta}_t^k\right), y_t\left(\widehat{\theta}_t^k\right); \widetilde{\theta}\right)\right. \\
& \left. -u_\theta\left(c_t\left(\widehat{\theta}_t^{n-m}\right), y_t\left(\widehat{\theta}_t^{n-m}\right); \widetilde{\theta}\right)\right] d\widetilde{\theta} \\
\geq & \ 0
\end{aligned} \tag{86}$$

The second assumption in the proposition ensures $c_t\left(\widehat{\theta}_t^k\right) > c_t\left(\widehat{\theta}_t^{n-m}\right)$ and $y_t\left(\widehat{\theta}_t^k\right) > y_t\left(\widehat{\theta}_t^{n-m}\right)$ for all $n-m < k \leq N$, whilst the third ensures we can transit between the points $\left(c_t\left(\widehat{\theta}_t^{n-m}\right), y_t\left(\widehat{\theta}_t^{n-m}\right)\right)$ and $\left(c_t\left(\widehat{\theta}_t^k\right), y_t\left(\widehat{\theta}_t^k\right)\right)$ along a path for which $\frac{dy_t}{dc_t} > 1$ throughout. To demonstrate the inequality it is sufficient that $u_{c\theta} + u_{y\theta}\frac{dy_t}{dc_t} > 0$ along this path. Conditions (3) and (2) give:

$$u_{c\theta} + u_{y\theta}\frac{dy_t}{dc_t} > \frac{u_\theta}{u_y}\left(u_{cy} + u_{yy}\frac{dy_t}{dc_t}\right) \tag{87}$$

The right-hand side of this expression will be positive so long as $u_{cy}+u_{yy}\frac{dy_t}{dc_t} < 0$. But $u_{cy}+u_{yy} < 0$ is implied whenever consumption is a normal good, so $\frac{dy_t}{dc_t} > 1$ is enough to confirm the result (recalling that $u_{yy} < 0$). Note that the inequality in (86) will then be strict whenever $m > 1$, as required by the final claim in the proposition.

It remains to show that there will be (strictly) no incentives to mimic a higher type. That is, we want to show:

$$\begin{aligned}
& u\left(c_t\left(\widehat{\theta}_t^{n-m}\right), y_t\left(\widehat{\theta}_t^{n-m}\right); \theta_t^{n-m}\right) + \beta U_{t+1}\left(\widehat{\theta}_t^{n-m}\right) \\
> & \ u\left(c_t\left(\widehat{\theta}_t^n\right), y_t\left(\widehat{\theta}_t^n\right); \theta_t^{n-m}\right) + \beta U_{t+1}\left(\widehat{\theta}_t^n\right)
\end{aligned} \tag{88}$$

for all $n \leq N$ and $1 \leq m < n$. By an analogous argument, applying condition (12) with equality, it is enough to show:

$$\begin{aligned} & u\left(c_t\left(\widehat{\theta}_t^n\right), y_t\left(\widehat{\theta}_t^n\right); \theta_t^{n-m}\right) - u\left(c_t\left(\widehat{\theta}_t^{n-m}\right), y_t\left(\widehat{\theta}_t^{n-m}\right); \theta_t^{n-m}\right) \\ < & \sum_{k=0}^{m-1}\left[u\left(c_t\left(\widehat{\theta}_t^{n-k}\right), y_t\left(\widehat{\theta}_t^{n-k}\right); \theta_t^{n-k}\right) - u\left(c_t\left(\widehat{\theta}_t^{n-k-1}\right), y_t\left(\widehat{\theta}_t^{n-k-1}\right); \theta_t^{n-k}\right)\right] \end{aligned} \tag{89}$$

or

$$\begin{aligned} & u\left(c_t\left(\widehat{\theta}_t^n\right), y_t\left(\widehat{\theta}_t^n\right); \theta_t^{n}\right) - u\left(c_t\left(\widehat{\theta}_t^{n}\right), y_t\left(\widehat{\theta}_t^{n}\right); \theta_t^{n-m}\right) \\ > & \sum_{k=n-m}^{n-1}\left[u\left(c_t\left(\widehat{\theta}_t^{k}\right), y_t\left(\widehat{\theta}_t^{k}\right); \theta_t^{k+1}\right) - u\left(c_t\left(\widehat{\theta}_t^{k}\right), y_t\left(\widehat{\theta}_t^{k}\right); \theta_t^{k}\right)\right] \end{aligned} \tag{90}$$

This, in turn, will hold so long as we have:

$$\begin{aligned} & \sum_{k=n-m}^{n-1}\int_{\theta_t^k}^{\theta_t^{k+1}}\left[u_\theta\left(c_t\left(\widehat{\theta}_t^k\right), y_t\left(\widehat{\theta}_t^k\right); \widetilde{\theta}\right)\right. \\ & \left. -u_\theta\left(c_t\left(\widehat{\theta}_t^n\right), y_t\left(\widehat{\theta}_t^n\right); \widetilde{\theta}\right)\right] d\widetilde{\theta} \\ < \; & 0 \end{aligned} \tag{91}$$

By assumption we have $c_t\left(\widehat{\theta}_t^k\right) < c_t\left(\widehat{\theta}_t^n\right)$ and $y_t\left(\widehat{\theta}_t^k\right) < y_t\left(\widehat{\theta}_t^n\right)$ for all $k < n \leq N$, and we can transit between the points $\left(c_t\left(\widehat{\theta}_t^k\right), y_t\left(\widehat{\theta}_t^k\right)\right)$ and $\left(c_t\left(\widehat{\theta}_t^n\right), y_t\left(\widehat{\theta}_t^n\right)\right)$ along a path for which $\frac{dy_t}{dc_t} > 1$ throughout. To demonstrate the inequality it is again sufficient that $u_{c\theta} + u_{y\theta}\frac{dy_t}{dc_t} > 0$ along this path. This has already been established, so the proof is complete.

A.2 Proof of Proposition 14

Note first from the definition of γ that $\gamma_n = \nu_n - \nu_{n-1}$ for all $n \in \{1, ..., N\}$, where we define $\nu_0 = 0$. The marginal resource cost at $t+1$ of the perturbation set out in Proposition 13 is:

$$\pi_\Theta\left(\theta^t\right)\sum_{n=1}^{N}\pi_\Theta\left(\theta_{t+1}^n|\theta_t\right)\left(\frac{\delta_n^c(0)}{d\delta} - \frac{\delta_n^y(0)}{d\delta}\right) \tag{92}$$

where the $\delta_n^c(\delta)$ and $\delta_n^y(\delta)$ functions satisfy restrictions (57) to (59), and $\delta_N^y(\delta)$ is again set to zero. Totally differentiating those restrictions, it is straightforward to show:

$$\frac{\delta_n^c(0)}{d\delta} = \frac{\frac{u_y\left(\theta_{t+1}^n\right)}{u_c\left(\theta_{t+1}^n\right)}\frac{\nu_{n+1}}{u_c\left(\widehat{\theta}_{t+1}^n;\theta_{t+1}^{n+1}\right)} - \frac{u_y\left(\widehat{\theta}_{t+1}^n;\theta_{t+1}^{n+1}\right)}{u_c\left(\widehat{\theta}_{t+1}^n;\theta_{t+1}^{n+1}\right)}\frac{\nu_n}{u_c\left(\theta_{t+1}^n\right)}}{\frac{u_y\left(\theta_{t+1}^n\right)}{u_c\left(\theta_{t+1}^n\right)} - \frac{u_y\left(\widehat{\theta}_{t+1}^n;\theta_{t+1}^{n+1}\right)}{u_c\left(\widehat{\theta}_{t+1}^n;\theta_{t+1}^{n+1}\right)}} \tag{93}$$

$$\frac{\delta_n^y(0)}{d\delta} = \frac{\frac{\nu_n}{u_c\left(\theta_{t+1}^n\right)} - \frac{\nu_{n+1}}{u_c\left(\widehat{\theta}_{t+1}^n;\theta_{t+1}^{n+1}\right)}}{\frac{u_y\left(\theta_{t+1}^n\right)}{u_c\left(\theta_{t+1}^n\right)} - \frac{u_y\left(\widehat{\theta}_{t+1}^n;\theta_{t+1}^{n+1}\right)}{u_c\left(\widehat{\theta}_{t+1}^n;\theta_{t+1}^{n+1}\right)}} \tag{94}$$

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

$$\frac{\delta_N^c(0)}{d\delta} = \frac{\nu_N}{u_c\left(\theta_{t+1}^N\right)} \tag{95}$$

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

$$\begin{aligned} \frac{\delta_n^c(0)}{d\delta} - \frac{\delta_n^y(0)}{d\delta} &= \nu_n \frac{1-\alpha\left(\theta_{t+1}^n\right)}{u_c\left(\theta_{t+1}^n\right) + \alpha\left(\theta_{t+1}^n\right) u_y\left(\theta_{t+1}^n\right)} \\ &\quad - \left(\nu_{n+1} - \nu_n\right) \frac{\tau\left(\theta_{t+1}^n\right)}{u_c\left(\widehat{\theta}_{t+1}^n;\theta_{t+1}^{n+1}\right)\left(1-\tau\left(\theta_{t+1}^n\right)\right) + u_y\left(\widehat{\theta}_{t+1}^n;\theta_{t+1}^{n+1}\right)} \\ &= \nu_n\left(\phi_2^c\left(\theta_{t+1}^n, 0\right) - \phi_2^y\left(\theta_{t+1}^n, 0\right)\right) \\ &\quad - \left(\nu_{n+1} - \nu_n\right)\left(\varphi_2^c\left(\theta_{t+1}^n, 0\right) - \varphi_2^y\left(\theta_{t+1}^n, 0\right)\right) \end{aligned} \tag{96}$$

and

$$\begin{aligned} \frac{\delta_N^c(0)}{d\delta} - \frac{\delta_N^y(0)}{d\delta} &= \frac{\delta_N^c(0)}{d\delta} \\ &= \frac{\nu_N}{u_c\left(\theta_{t+1}^N\right)} \\ &= \nu_N\left(\phi_2^c\left(\theta_{t+1}^N, 0\right) - \phi_2^y\left(\theta_{t+1}^N, 0\right)\right) \end{aligned} \tag{97}$$

where we apply the earlier definitions of the ϕ^c, ϕ^y, φ^c and φ^y functions. Hence:

$$\begin{aligned} & \sum_{n=1}^{N} \pi_\Theta\left(\theta_{t+1}^n | \theta_t\right)\left(\frac{\delta_n^c(0)}{d\delta} - \frac{\delta_n^y(0)}{d\delta}\right) \\ = & \sum_{n=1}^{N} \pi_\Theta\left(\theta_{t+1}^n | \theta_t\right)\left[\nu_n\left(\phi_2^c\left(\theta_{t+1}^n, 0\right) - \phi_2^y\left(\theta_{t+1}^n, 0\right)\right)\right. \\ & \left. - \left(\nu_{n+1} - \nu_n\right)\left(\varphi_2^c\left(\theta_{t+1}^n, 0\right) - \varphi_2^y\left(\theta_{t+1}^n, 0\right)\right)\right] \\ = & \sum_{n=1}^{N-1} \pi_\Theta\left(\theta_{t+1}^n | \theta_t\right)\left[\sum_{m=1}^{n} \gamma_m\left(\phi_2^c\left(\theta_{t+1}^n, 0\right) - \phi_2^y\left(\theta_{t+1}^n, 0\right)\right)\right. \\ & \left. -\gamma_{n+1}\left(\varphi_2^c\left(\theta_{t+1}^n, 0\right) - \varphi_2^y\left(\theta_{t+1}^n, 0\right)\right)\right] \\ & + \pi_\Theta\left(\theta_{t+1}^N | \theta_t\right) \sum_{m=1}^{N} \gamma_m\left(\phi_2^c\left(\theta_{t+1}^N, 0\right) - \phi_2^y\left(\theta_{t+1}^N, 0\right)\right) \end{aligned} \tag{98}$$

This last expression can equivalently be written in matrix form:

$$\left(\pi_\Theta^{vec}\right)'\left[\sum_{n=1}^{N} \gamma_n \Delta^{n\prime}(0)\right] k \tag{99}$$

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 $\Delta^{n\prime}(0)$ matrix and preserve incentive compatibility. 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

$$u\left(c^*_{N,t+1} + \delta_N^c(\delta), y^*_{N,t+1} + \delta_N^y(\delta); \theta_{t+1}^N\right) = u\left(c^*_{N,t+1}, y^*_{N,t+1}; \theta_{t+1}^N\right) + \nu_N \delta \tag{100}$$

This gives:

$$u_c\left(\theta_{t+1}^N\right)\frac{\delta_N^c(0)}{d\delta} + u_y\left(\theta_{t+1}^N\right)\frac{\delta_N^y(0)}{d\delta} = \nu_N \tag{101}$$

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

$$\begin{aligned}\frac{\delta_N^c(0)}{d\delta} - \frac{\delta_N^y(0)}{d\delta} &= \frac{\nu_N}{u_c\left(\theta_{t+1}^N\right)} \\ &= \nu_N\left(\phi_2^c\left(\theta_{t+1}^N, 0\right) - \phi_2^y\left(\theta_{t+1}^N, 0\right)\right) \\ &= \nu_N\left(\phi_2^c\left(\theta_{t+1}^N, 0\right) - \phi_2^y\left(\theta_{t+1}^N, 0\right)\right) \\ &\quad +\kappa\left(\varphi_2^c\left(\theta_{t+1}^N, 0\right) - \varphi_2^y\left(\theta_{t+1}^N, 0\right)\right)\end{aligned} \tag{102}$$

for arbitrary κ. So if $\frac{\delta_N^y(0)}{d\delta}$ differs from zero then $\frac{\delta_N^c(0)}{d\delta}$ can differ from $\frac{\nu_N}{u_c\left(\theta_{t+1}^N\right)}$, 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 $\Delta^{N+1\prime}(0)$.

A.3 Proof of Proposition 16

The proof is very similar to that of Proposition 6, with adjustments needing to be made only to incorporate the more complex Markov structure of shocks. To confirm incentive compatibility is satisfied for all 'downwards' comparisons we now wish to show:

$$\begin{aligned} & u\left(c_t\left(\widehat{\theta}_t^{n-m}\right), y_t\left(\widehat{\theta}_t^{n-m}\right); \theta_t^n\right) + \beta U_{t+1}\left(\widehat{\theta}_t^{n-m}|\theta_t^n\right) \\ \leq\ & u\left(c_t\left(\widehat{\theta}_t^{n}\right), y_t\left(\widehat{\theta}_t^{n}\right); \theta_t^n\right) + \beta U_{t+1}\left(\widehat{\theta}_t^{n}|\theta_t^n\right)\end{aligned} \tag{103}$$

for all $n \in \{2, ..., N\}$ and $m \in \{1, ..., n-1\}$, where the the function $U_{t+1}\left(\widehat{\theta}_t^{n-m}|\theta_t^n\right)$ is the expected lifetime utility from time $t+1$ onwards of an agent with a given past reporting history (suppressed in the notation) who reports $\widehat{\theta}_t^{n-m}$ at time t and is of true type θ_t^n. Note that this is equal to $\pi_\Theta^{vec}\left(\theta_t^n\right)' V_{t+1}^{vec}\left(\widehat{\theta}^{t-1}, \widehat{\theta}_t^{n-m}\right)$ in terms of the notation used in the Proposition.

Condition (12), holding with equality and summed across pairs of types intermediate between θ^n and θ^{n-m}, gives:

$$\begin{aligned} & \sum_{k=0}^{m-1} \beta \left[U_{t+1}\left(\widehat{\theta}_t^{n-k} | \theta_t^{n-k}\right) - U_{t+1}\left(\widehat{\theta}_t^{n-k-1} | \theta_t^{n-k}\right)\right] \\ = & \sum_{k=0}^{m-1} \left[u\left(c_t\left(\widehat{\theta}_t^{n-k-1}\right), y_t\left(\widehat{\theta}_t^{n-k-1}\right); \theta_t^{n-k}\right) - u\left(c_t\left(\widehat{\theta}_t^{n-k}\right), y_t\left(\widehat{\theta}_t^{n-k}\right); \theta_t^{n-k}\right)\right] \end{aligned} \tag{104}$$

The fourth condition in the Proposition ensures that for all $k \in \{0, ..., m-1\}$ we have:

$$\begin{aligned} & U_{t+1}\left(\widehat{\theta}_t^{n-k} | \theta_t^{n-k}\right) - U_{t+1}\left(\widehat{\theta}_t^{n-k-1} | \theta_t^{n-k}\right) \\ \leq & U_{t+1}\left(\widehat{\theta}_t^{n-k} | \theta_t^{n}\right) - U_{t+1}\left(\widehat{\theta}_t^{n-k-1} | \theta_t^{n}\right) \end{aligned} \tag{105}$$

Summing this, it follows:

$$\begin{aligned} & \beta \left[U_{t+1}\left(\widehat{\theta}_t^{n} | \theta_t^{n}\right) - U_{t+1}\left(\widehat{\theta}_t^{n-m} | \theta_t^{n}\right)\right] \\ \geq & \sum_{k=0}^{m-1} \beta \left[U_{t+1}\left(\widehat{\theta}_t^{n-k} | \theta_t^{n-k}\right) - U_{t+1}\left(\widehat{\theta}_t^{n-k-1} | \theta_t^{n-k}\right)\right] \end{aligned} \tag{106}$$

So:

$$\begin{aligned} & \beta \left[U_{t+1}\left(\widehat{\theta}_t^{n} | \theta_t^{n}\right) - U_{t+1}\left(\widehat{\theta}_t^{n-m} | \theta_t^{n}\right)\right] \\ \geq & \sum_{k=0}^{m-1} \left[u\left(c_t\left(\widehat{\theta}_t^{n-k-1}\right), y_t\left(\widehat{\theta}_t^{n-k-1}\right); \theta_t^{n-k}\right) - u\left(c_t\left(\widehat{\theta}_t^{n-k}\right), y_t\left(\widehat{\theta}_t^{n-k}\right); \theta_t^{n-k}\right)\right] \end{aligned} \tag{107}$$

Using this in (103), it follows that it is sufficient once more for us to show:

$$\begin{aligned} & u\left(c_t\left(\widehat{\theta}_t^{n}\right), y_t\left(\widehat{\theta}_t^{n}\right); \theta_t^{n}\right) - u\left(c_t\left(\widehat{\theta}_t^{n-m}\right), y_t\left(\widehat{\theta}_t^{n-m}\right); \theta_t^{n}\right) \\ \geq & \sum_{k=0}^{m-1} \left[u\left(c_t\left(\widehat{\theta}_t^{n-k}\right), y_t\left(\widehat{\theta}_t^{n-k}\right); \theta_t^{n-k}\right) - u\left(c_t\left(\widehat{\theta}_t^{n-k-1}\right), y_t\left(\widehat{\theta}_t^{n-k-1}\right); \theta_t^{n-k}\right)\right] \end{aligned} \tag{108}$$

The argument can then proceed as in the proof of Proposition 6.

Similarly, to rule out 'upwards' mimicking we need:

$$\begin{aligned} & u\left(c_t\left(\widehat{\theta}_t^{n-m}\right), y_t\left(\widehat{\theta}_t^{n-m}\right); \theta_t^{n-m}\right) + \beta U_{t+1}\left(\widehat{\theta}_t^{n-m} | \theta_t^{n-m}\right) \\ > & u\left(c_t\left(\widehat{\theta}_t^{n}\right), y_t\left(\widehat{\theta}_t^{n}\right); \theta_t^{n-m}\right) + \beta U_{t+1}\left(\widehat{\theta}_t^{n} | \theta_t^{n-m}\right) \end{aligned} \tag{109}$$

for all $n \in \{2, ..., N\}$ and $m \in \{1, ..., n-1\}$. We again have:

$$\begin{aligned} & \sum_{k=0}^{m-1} \beta \left[U_{t+1}\left(\widehat{\theta}_t^{n-k-1} | \theta_t^{n-k}\right) - U_{t+1}\left(\widehat{\theta}_t^{n-k} | \theta_t^{n-k}\right)\right] \\ = & \sum_{k=0}^{m-1} \left[u\left(c_t\left(\widehat{\theta}_t^{n-k}\right), y_t\left(\widehat{\theta}_t^{n-k}\right); \theta_t^{n-k}\right) - u\left(c_t\left(\widehat{\theta}_t^{n-k-1}\right), y_t\left(\widehat{\theta}_t^{n-k-1}\right); \theta_t^{n-k}\right)\right] \end{aligned} \tag{110}$$

Moreover, note:

$$\begin{aligned} & U_{t+1}\left(\widehat{\theta}_t^{n-m}|\theta_t^{n-m}\right) - U_{t+1}\left(\widehat{\theta}_t^{n}|\theta_t^{n-m}\right) \\ = & \sum_{k=0}^{m-1}\left[U_{t+1}\left(\widehat{\theta}_t^{n-k-1}|\theta_t^{n-m}\right) - U_{t+1}\left(\widehat{\theta}_t^{n-k}|\theta_t^{n-m}\right)\right] \\ > & \sum_{k=0}^{m-1}\left[U_{t+1}\left(\widehat{\theta}_t^{n-k-1}|\theta_t^{n-k}\right) - U_{t+1}\left(\widehat{\theta}_t^{n-k}|\theta_t^{n-k}\right)\right] \end{aligned} \tag{111}$$

where we have used the fourth assumption in the Proposition to assert:

$$\begin{aligned} & U_{t+1}\left(\widehat{\theta}_t^{n-k-1}|\theta_t^{n-m}\right) - U_{t+1}\left(\widehat{\theta}_t^{n-k}|\theta_t^{n-m}\right) \\ > & U_{t+1}\left(\widehat{\theta}_t^{n-k-1}|\theta_t^{n-k}\right) - U_{t+1}\left(\widehat{\theta}_t^{n-k}|\theta_t^{n-k}\right) \end{aligned} \tag{112}$$

Using this result in (109), it will be sufficient for us to show:

$$\begin{aligned} & u\left(c_t\left(\widehat{\theta}_t^{n}\right), y_t\left(\widehat{\theta}_t^{n}\right); \theta_t^{n-m}\right) - u\left(c_t\left(\widehat{\theta}_t^{n-m}\right), y_t\left(\widehat{\theta}_t^{n-m}\right); \theta_t^{n-m}\right) \\ < & \sum_{k=0}^{m-1}\beta\left[U_{t+1}\left(\widehat{\theta}_t^{n-k-1}|\theta_t^{n-k}\right) - U_{t+1}\left(\widehat{\theta}_t^{n-k}|\theta_t^{n-k}\right)\right] \end{aligned} \tag{113}$$

or:

$$\begin{aligned} & u\left(c_t\left(\widehat{\theta}_t^{n}\right), y_t\left(\widehat{\theta}_t^{n}\right); \theta_t^{n-m}\right) - u\left(c_t\left(\widehat{\theta}_t^{n-m}\right), y_t\left(\widehat{\theta}_t^{n-m}\right); \theta_t^{n-m}\right) \\ < & \sum_{k=0}^{m-1}\left[u\left(c_t\left(\widehat{\theta}_t^{n-k}\right), y_t\left(\widehat{\theta}_t^{n-k}\right); \theta_t^{n-k}\right) - u\left(c_t\left(\widehat{\theta}_t^{n-k-1}\right), y_t\left(\widehat{\theta}_t^{n-k-1}\right); \theta_t^{n-k}\right)\right] \end{aligned} \tag{114}$$

The rest of the proof can then follow that of Proposition 6.

A.4 Proof of Proposition 28

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 θ_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:

$$\sum_{\theta_{t+1}\in\Theta}\pi_\Theta(\theta_{t+1}|\theta_t)\left[\frac{1-\alpha(\theta_{t+1})}{u_c(\theta_{t+1}) + u_y(\theta_{t+1})\alpha(\theta_{t+1})}\right] \geq 1 \tag{115}$$

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

$$\left[\sum_{n=1}^{N+1}\gamma_n\Delta^{n\prime}(0)\right]=\widetilde{\Delta} \tag{116}$$

for some $N\times 2$ $\widetilde{\Delta}$ 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 $\Delta^{n\prime}(0)$ 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 $\Delta^{m\prime}(0)$ matrices. So we can choose the entries of γ by first finding an additive combination of $\Delta^{1\prime}(0)$ and $\Delta^{2\prime}(0)$ sufficient to increase the consumption of the agent of type θ^1_{t+1} without changing that agent's output requirements, then add to this sufficient units of $\Delta^{3\prime}(0)$ for the output level of the agent of type θ^2_{t+1} to remain unchanged (this perturbation has no impact on type θ^1_{t+1}), then sufficient units of $\Delta^{4\prime}(0)$ for the output level of the agent of type θ^3_{t+1} to remain unchanged, and so on. Since consumption is increasing for the agent of type θ^1_{t+1}, it must also increase for all higher-type agents – otherwise we could not continue to satisfy incentive compatibility. This $\widetilde{\Delta}$ 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 $\Delta^{n\prime}(0)$ and $\widetilde{\Delta}$, with a positive coefficient on the former and a negative on the latter. But if $\widetilde{\Delta}$ is providing utility through consumption increments alone it must come at a positive cost, whilst we have from (115) that the cost of $\Delta^{n\prime}(0)$. Hence we raise a surplus, contradicting optimality.[37]

A.5 Proof of Proposition 29

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 31 $\frac{\tau(\theta^n_t)}{u_c\left(\widehat{\theta}^n_t;\theta^{n+1}_t\right)(1-\tau(\theta^n_t))+u_y\left(\widehat{\theta}^n_t;\theta^{n+1}_t\right)}\overset{a.s.}{\to}0$ *holds under an optimal plan that solves the restricted problem.*

[37] 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 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 permts do not span an entire half-space in the (c_{t+1},y_{t+1}) plane for each agent.

Proof. In the iid case this follows directly from equation (51):

$$\begin{aligned} & \lim_{t\to\infty}\left[-\pi_\Theta\left(\theta^n_{t+1}|\theta_t\right)\frac{\tau\left(\theta^n_{t+1}\right)}{u_c\left(\widehat{\theta}^n_{t+1};\theta^{n+1}_{t+1}\right)\left(1-\tau\left(\theta^n_{t+1}\right)\right)+u_y\left(\widehat{\theta}^n_{t+1};\theta^{n+1}_{t+1}\right)}\right] \quad (117) \\ = & -\sum_{m=n+1}^{N}\pi_\Theta\left(\theta^m_{t+1}|\theta_t\right)\lim_{t\to\infty}\left[\frac{1-\alpha\left(\theta^m_{t+1}\right)}{u_c\left(\theta^m_{t+1}\right)+u_y\left(\theta^m_{t+1}\right)\alpha\left(\theta^m_{t+1}\right)}\right] \\ & +\pi_\Theta\left(\theta_{t+1}>\theta^n_{t+1}|\theta_t\right)\sum_{\theta_{t+1}\in\Theta}\pi_\Theta\left(\theta_{t+1}|\theta_t\right)\lim_{t\to\infty}\left[\frac{1-\alpha\left(\theta_{t+1}\right)}{u_c\left(\theta_{t+1}\right)+u_y\left(\theta_{t+1}\right)\alpha\left(\theta_{t+1}\right)}\right] \\ = & \; 0 \end{aligned}$$

In the Markov case we know that equation (51) 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 Tth occasion on which $\theta=\theta^N$ has obtained along the given sample path, we must have:

$$\begin{aligned} & \lim_{T\to\infty}\left[-\pi_\Theta\left(\theta^n_{t(T)+1}|\theta_{t(T)}\right)\right. \quad (118) \\ & \left.\cdot\frac{\tau\left(\theta^n_{t(T)+1}\right)}{u_c\left(\widehat{\theta}^n_{t(T)+1};\theta^{n+1}_{t(T)+1}\right)\left(1-\tau\left(\theta^n_{t(T)+1}\right)\right)+u_y\left(\widehat{\theta}^n_{t(T)+1};\theta^{n+1}_{t(T)+1}\right)}\right] \\ = & -\sum_{m=n+1}^{N}\pi_\Theta\left(\theta^m_{t(T)+1}|\theta_{t(T)}\right)\lim_{T\to\infty}\left[\frac{1-\alpha\left(\theta^m_{t(T)+1}\right)}{u_c\left(\theta^m_{t(T)+1}\right)+u_y\left(\theta^m_{t(T)+1}\right)\alpha\left(\theta^m_{t(T)+1}\right)}\right] \\ & +\pi_\Theta\left(\theta_{t(T)+1}>\theta^n_{t(T)+1}|\theta_{t(T)}\right) \\ & \cdot\sum_{\theta_{t(T)+1}\in\Theta}\pi_\Theta\left(\theta_{t(T)+1}|\theta_{t(T)}\right)\lim_{T\to\infty}\left[\frac{1-\alpha\left(\theta_{t(T)+1}\right)}{u_c\left(\theta_{t(T)+1}\right)+u_y\left(\theta_{t(T)+1}\right)\alpha\left(\theta_{t(T)+1}\right)}\right] \\ = & \; 0 \end{aligned}$$

But if $\frac{\tau\left(\theta^n_{t(T)+1}\right)}{u_c\left(\widehat{\theta}^n_{t(T)+1};\theta^{n+1}_{t(T)+1}\right)\left(1-\tau\left(\theta^n_{t(T)+1}\right)\right)+u_y\left(\widehat{\theta}^n_{t(T)+1};\theta^{n+1}_{t(T)+1}\right)}=0$ 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 (75) and (69). 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 $u_c\left(\widehat{\theta}_t^n;\theta_t^{n+1}\right)\left(1-\tau\left(\theta_t^n\right)\right)+u_y\left(\widehat{\theta}_t^n;\theta_t^{n+1}\right) \xrightarrow{a.s.} \infty$. Suppose the latter were true. By equation (79) we have:

$$
\begin{aligned}
u_c\left(\theta_t^n\right)+u_y\left(\theta_t^n\right)\alpha\left(\theta_t^n\right) &= \frac{u_c\left(\widehat{\theta}_t^n;\theta_t^{n+1}\right)-u_y\left(\widehat{\theta}_t^n;\theta_t^{n+1}\right)\frac{u_c\left(\theta_t^n\right)}{u_y\left(\theta_t^n\right)}}{1-\frac{u_y\left(\widehat{\theta}_t^n;\theta_t^{n+1}\right)}{u_y\left(\theta_t^n\right)}} \qquad (119)\\
&> u_c\left(\widehat{\theta}_t^n;\theta_t^{n+1}\right)-u_y\left(\widehat{\theta}_t^n;\theta_t^{n+1}\right)\frac{u_c\left(\theta_t^n\right)}{u_y\left(\theta_t^n\right)}\\
&= u_c\left(\widehat{\theta}_t^n;\theta_t^{n+1}\right)-u_y\left(\widehat{\theta}_t^n;\theta_t^{n+1}\right)\frac{1}{\left(1-\tau\left(\theta_t^n\right)\right)}
\end{aligned}
$$

If $u_c\left(\widehat{\theta}_t^n;\theta_t^{n+1}\right)\left(1-\tau\left(\theta_t^n\right)\right)+u_y\left(\widehat{\theta}_t^n;\theta_t^{n+1}\right) \xrightarrow{a.s.} \infty$ then $u_c\left(\widehat{\theta}_t^n;\theta_t^{n+1}\right)-u_y\left(\widehat{\theta}_t^n;\theta_t^{n+1}\right)\frac{1}{\left(1-\tau\left(\theta_t^n\right)\right)} \xrightarrow{a.s.} \infty$ must also hold, since $\left(1-\tau\left(\theta_t^n\right)\right)\in[0,1]$ follows from the definition of τ and Proposition 26. Hence we must also have $u_c\left(\theta_t^n\right)+u_y\left(\theta_t^n\right)\alpha\left(\theta_t^n\right) \xrightarrow{a.s.} \infty$. This in turn implies $\frac{1-\alpha\left(\theta_t^n\right)}{u_c\left(\theta_t^n\right)+u_y\left(\theta_t^n\right)\alpha\left(\theta_t^n\right)}$ can only converge to a non-zero limit if $\left|\alpha\left(\theta_t\right)\right|$ is itself always infinite at that limit. But since we know $\alpha\left(\theta_t\right)=0$ when $\theta_t=\theta^N$ we can rule that out.

The alternative is that $\tau\left(\theta_t^n\right) \xrightarrow{a.s.} 0$. In this case we have $u_c\left(\theta_t^n\right)=-u_y\left(\theta_t^n\right)$ at the limit, and so $\frac{1-\alpha\left(\theta_t^n\right)}{u_c\left(\theta_t^n\right)+u_y\left(\theta_t^n\right)\alpha\left(\theta_t^n\right)}=\frac{1}{u_c\left(\theta_t^n\right)}$. Hence the inverse of the marginal utility of consumption must be converging to a common value for all agents. But since $u_c\left(\theta_t^n\right)=-u_y\left(\theta_t^n\right)$ 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 in Proposition 4 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 $\frac{1-\alpha\left(\theta_t^n\right)}{u_c\left(\theta_t^n\right)+u_y\left(\theta_t^n\right)\alpha\left(\theta_t^n\right)}$ converging to a non-zero value in this case too. This completes the proof.