Optimal taxation with persistent shocks

Marek Kapicka

University of California, Santa Barbara

June 25 2004

VERY PRELIMINARY!

Abstract

In this paper I study dynamic optimal taxation in a private information economy with continuum of individual productivity shocks that are persistent over time. I formulate the problem recursively and develop a first order approach in the spirit of Mirrlees (1971) to simplify it. The main advantage of the first order approach lies in the fact that it allows us to reduce the state space of the dynamic program dramatically. This allows numerical implementation of the problem. I solve quantitatively for the optimal capital income taxes in a simplified economy with persistent taste shocks. I find that whan the shocks follow random walk, the intertemporal wedge, roughly corresponding to the capital income taxes, is on average about three times lower than in the case of iid shocks.

1 Introduction

This paper studies optimal income taxation in a dynamic private information economy with continuum of idiosyncratic productivity shocks. I depart from the typical assumption of iid shocks and allow them to be persistent over time. This assumption implies tremendous burden on the dimensionality of the state space, as was shown by Fernandes and Phelan [6]. Therefore I develop a first order approach and show that it simplifies the state space to a manageable dimension. Thus, the advantage of the first order approach becomes twofold. First, it simplifies the incentive compatibility constraint, as is known from static models. Second, it reduces the state space of social planner's dynamic program.

This paper follows the Mirrleesian approach to the optimal taxation (see [11],[12],[13]). The advantage of this approach lies in its explicit modelling of private information structure. This

1

provides microfoundations for usage of nonlinear taxes. All tax instruments arise endogenously as part of the competitive equilibrium with benevolent government. This stands in contrast to the more traditional Ramsey approach where the tax instruments are exogenously given.

Originally, this approach studied static economies where it is optimal to put a wedge between marginal rate of substitution between consumption and leisure and marginal productivity of labor, due to the informational frictions. This gives rise to the nonlinear income tax in the optimum. Recently, this literature was extended to dynamic private information economies by Golosov, Kocherlakota, Tsyvinski [7], Kocherlakota [9] or Albanesi and Sleet [2]. The importance of this extension is twofold. First, it allows the government to improve upon static allocations by using history dependent taxes. Second, it allows us to analyze capital taxation. The general conclusion is that there is an additional wedge that arises in the optimum. It is an intertemporal wedge between marginal rate of intertemporal substitution and the interest rate.[1] This creates a role for capital income taxes.

The literature on dynamic optimal income taxation follows two approaches. It either focuses on quantitative implementation of solutions and simply assumes iid shocks (Albanesi and Sleet [2]) or focuses on very general qualitative conclusions and gives up on numerical implementation (Golosov, Kocherlakota, Tsyvinski [7], Kocherlakota [9]) This paper retains the ability to numerically compute optimal allocations but allows for reasonably realistic class of Markov shock processes, thus getting the best of both approaches. The assumption of shocks to be highly persistent appears to be rather a realistic one. Heathcote, Storesletten and Violante [8] find that the autocorrelation of wage shocks is about 0.94, and thus exhibits near random walk behavior.

Technically, the dynamic program presented in this paper builds on the work of Fernandes and Phelan [6]. They consider private information economies with Markov shocks but restrict the shocks to have only two possible values. They show how to formulate the program recursively. The extension of their approach to the case of continuum of shocks is quite straightforward, but not very useful per se. The reason is that, in contrast to the case of iid shocks, the continuation utility becomes a function. Thus, the state variable is also a function. This is where the extreme usefulness of the first order approach comes in. It helps to reduce the state space from a function to two numbers. In comparison with the two shock economy, the dynamic becomes no more complex with continuum of shocks.

---

[1]There are cases where this result does not hold and the intertemporal wedge is zero. See Shimer and Werning ([14]).

For the numerical exercise I consider a simple private economy with taste shocks. While this economy is not rich enough to study labor income taxes, it is convenient if one wants to focus on the intertemporal wedge, i.e. on the optimal capital income taxes. I consider the extreme case of shock persistence: the case when shocks follow a random walk. I compute the optimal intertemporal wedges for such economy and compare the results with an economy with iid shocks. I find that the intertemporal wedge is significantly reduced when we consider the economy with persistent shocks. It constitutes about a fourth of the intertemporal wedge in the economy with iid shocks. Based on the beforementioned research, the qualitative conclusion is to be expected. The contribution of the paper is rather the quantitative assessment.

To understand why the first order approach simplifies the dynamic program, we must first understand why the state space becomes so complicated without it. If shocks are persistent, the continuation utility depends not only on what the agent reports today, but also what her true shock is. The reason is that the probability distribution of future shocks depends on the true shock. To ensure incentive compatibility, next period has to deliver the continuation utility for all possible *true* shocks. I call this the continuation utility *function* (as a function of the true state). It follows, that the whole continuation utility function must be a state variable of the social planner's problem.[2] The first order approach, on the other hand, implies that only the marginal change of the continuation utility is what matters. Thus, we can replace the continuation utility function by marginal continuation utility. This is the gist of the usefulness of the first order approach. Besides that, we have the ordinary promised keeping constraint and therefore the state space boils down to a manageable set of two real numbers.

The main problem obviously lies in the justification of the first order approach. I will use the envelope theorem of Milgrom and Segal (2003), but the problem with the first order approach is more complicated. This is so because the utility of an agent consists of two parts: period utility function and the continuation utility function. The latter is endogenous to the social planner's problem and we cannot simply impose any properties on it. Yet, using the envelope theorem requires differentiability with respect to agent's type and some other technical properties. It turns out, that we can justify the required properties for the continuation utility function solely by imposing some structure on the probability distribution of shock. The envelope theorem can then be fully justified.

---

[2]In case of iid shocks, the continuation utility is independent of the current true shock, so we have only one value of the continuation utility, instead of a function. That's why models with iid. shocks are relatively easy to solve.

The paper is organized as follows. Next section lays out the physical environment of a general model. Section 3 formulates the social planner's problem in a sequence space. Section 4 then constructs recursive formulation of the social planner's problem and shows the equivalence between both solutions. First order approach is introduced and justified in section 5. Section 6 introduces a simple taste shock economy and shows that the problem can be further simplified. Section 7 presents numerical simulations. Section 8 concludes and discusses further extensions. Most of the proofs can be found in the Appendix.

2 The Model

The world begins at time $t = 1$. There is a continuum of agents, with a unit measure in this economy. The agents have period utility that depends positively on consumption $c \in \mathcal{R}_+$, negatively on labor supply $l \in [0, 1]$ and is given by $U : \mathcal{R}_+ \times [0, 1] \rightarrow \mathcal{R}$. We assume the utility function is bounded, concave and twice differentiable with respect to both arguments. It also satisfies that $U_{cl} \geq 0$.

At the beginning of each period, agents observe their current productivity shock $\theta \in \Theta \subseteq \mathcal{R}_+$. If the agent supplies l units of labor, her output is given by $y = \theta l$. We will follow a standard approach and substitute labor out of the problem. Thus, the period utility function depends on consumption, output and productivity shock in the following way: $U = U(c, \frac{y}{\theta})$.

The productivity shock follows first order Markov process $\pi(\theta'|\theta)$ It is assumed that the function π is twice differentiable with respect to both arguments. Moreover, we assume that the transition function is such that there exists an integrable function $\kappa : \Theta \rightarrow \mathcal{R}$ such that $\left| \frac{\pi_2(\theta'|\theta)}{\pi(\theta'|\theta)} \right| \leq \kappa(\theta)$ for all θ' and almost all θ.

We denote an invariant distribution by $\pi(\theta)$. We can construct probability of an arbitrary sequence θ^t that follows θ_0 and write it as $\Pi(\theta^t|\theta_0)$. Unconditional distribution of this sequence is denoted by $\Pi(\theta^t)$. We assume that the shock θ_0 is the same for everyone and is observed by the social planner.

Consumption and output are observed by the social planner, while the productivity shock is not. It is a private information of the agent. The agents are infinitely lived and discount future by factor β. Their objective is to maximize expected lifetime utility.

3 Sequence problem

For each period, the social planner designs a pair of consumption assignments $C_t : \Theta^t \to \mathcal{R}_+$ and output assignments $Y_t : \Theta^t \to \mathcal{R}_+$. Call the collection of these assignments for all period $\mathbf{C} = \{C_t\}_{t \geq 1}$ and $\mathbf{Y} = \{Y_t\}_{t \geq 1}$ an *allocation*.

At the beginning of period t, the agents report their current type to the social planner. The reporting strategy of an agent can be described by a collection of functions $\hat{\boldsymbol{\theta}} = \{\hat{\theta}_t\}_{t \geq 1}$ where $\hat{\theta}_t : \Theta^t \to \Theta$ is a report in period t. The history of reports up to period t is denoted by $\hat{\theta}^t \in \Theta^t$.

The agent's preferences over allocations are given by

$$u(\mathbf{C}, \mathbf{Y}, \theta_0) = \sum_{t \geq 1} \int_{\theta^t \in \Theta^t} \beta^{t-1} U(C_t(\theta^t), \frac{Y_t(\theta^t)}{\theta_t}) \Pi(\theta^t | \theta_0) d\theta^t$$

Since the shocks are private information to the agent, an allocation must satisfy the incentive compatibility requirement. If the agent chooses reporting strategy $\hat{\boldsymbol{\theta}}$ he receives consumption assignment $\mathbf{C} \circ \hat{\boldsymbol{\theta}} = \{C_t(\hat{\theta}^t(\theta^t))\}_{t \geq 1}$ and similarly with output assignment. If he reports truthfully, he just receives $\mathbf{C}$ and $\mathbf{Y}$. Thus, incentive compatibility constraint can be written as

$$u(\mathbf{C}, \mathbf{Y}, \theta_0) \geq u(\mathbf{C} \circ \hat{\boldsymbol{\theta}}, \mathbf{Y} \circ \hat{\boldsymbol{\theta}}, \theta_0) \qquad \forall \hat{\boldsymbol{\theta}} \text{ s.t. } \mathbf{Y} \circ \hat{\boldsymbol{\theta}} \leq \boldsymbol{\theta} \tag{1}$$

where the last inequality reflects the fact that the agent cannot choose unfeasible reports, i.e. reports that would result in labor supply greater than 1.

Social planner maximizes the expected utility of all agents by a choice of an allocation. In principle, the social planner could assign different Pareto weights to different agents, but since all agents are ex ante identical, I will assume that they all have equal weights. This assumption can be easily dropped. The social planner is constrained by the incentive compatibility constraint and by a sequence of period by period resource constraint. Thus, we can write the social planner's problem as follows:

$$\max_{\mathbf{C}, \mathbf{Y}} u(\mathbf{C}, \mathbf{Y}, \theta_0)$$

subject to incentive compatibility constraint (1) and a sequence of period resource constraints

$$\int_{\theta_t \in \Theta} [C_t(\theta^t) - Y_t(\theta^t)] \Pi(\theta^t | \theta_0) d\theta_t = 0 \qquad \forall \theta^{t-1} \in \Theta^{t-1} \tag{2}$$

I will refer to this problem as a sequence problem of the social planner. Denote the solution to this program by $\mathbf{C}^*, \mathbf{Y}^*$.

5

4 Recursive formulation

In this section I will write down a decentralized problem of minimizing the costs of delivering certain promised utility and show how this program is related to the sequence problem In principle, there are several intermediate steps between the sequence problem and the decentralized cost minimization problem. Since they are fairly standard, I present them in the Appendix.

Define a *recursive allocation* to be a triplet of functions $c : \Theta \to \mathcal{R}_+$, $y : \Theta \to \mathcal{R}_+$ and $w' : \Theta^2 \to \mathcal{W}$.[3] The first function corresponds to consumption, the second one to output and the last one to continuation utility. Also define $\{q_t\}_{t \geq 1}$ to be a sequence of intertemporal prices of consumption satisfying $\sum_{t \geq 1} \prod_{i=1}^{t} q_i \leq +\infty$.

Consider a social planner that minimizes costs of delivering promised utility $w = w(\theta_-)$ to an agent who incurred shock θ_- at period $t-1$. Call this planner a component planner. Define L^Θ to be a space of all functions $\Theta \to \mathcal{W}$. The claim I will prove is that, for $t > 1$, the following dynamic program of a component planner is an equivalent way of writing the sequence problem of a social planner. [4]

$$V_t(w(.), \theta_-) = \min_{c,y,w'} \int_\theta [c(\theta) - y(\theta) + q_t V_{t+1}(w'(\theta,.), \theta)] \pi(\theta|\theta_-) d\theta$$

$$s.t.$$

$$w(\hat{\theta}_-) = \int_\theta [U(c(\theta), \frac{y(\theta)}{\theta}) + \beta w'(\theta, \theta)] \pi(\theta|\hat{\theta}_-) d\theta \qquad \forall \hat{\theta}_- \qquad (3)$$

$$U(c(\theta), \frac{y(\theta)}{\theta}) + \beta w'(\theta, \theta) \geq U(c(\hat{\theta}), \frac{y(\hat{\theta})}{\theta}) + \beta w'(\hat{\theta}, \theta) \qquad \forall \hat{\theta} \text{ s.t. } y(\hat{\theta}) \leq \theta, \text{ all } \theta \qquad (4)$$

$$w'(\theta, .) \in B^*$$

The state space of the value function is given by $L^\Theta \times \Theta$. The first constraint (3) incorporates two distinct things. First, for $\hat{\theta}_- = \theta_-$, it is a promise keeping constraint. Second, for other $\hat{\theta}_-$ it is a threat keeping constraint, in terms of Fernandes and Phelan. The second constraint (4) is a temporary incentive compatibility constraint. It implies that any one period deviations are suboptimal. The last constraint ensures that it will indeed be possible to deliver the continuation

---

[3]We cna show by standard arguments that since utility is bounded, $\mathcal{W}$ can be restricted to be bounded as well.

[4]The value function depends on the whole sequence of intertemporal prices $\{q_t\}_{t \geq 1}$ but the dependence is kept implicit to simplify the notation.

utility function next period. The set B^* is defined as a fixed point of the following operator:

$$T(B) \quad = \quad \{w \in L^\Theta : \exists c : \Theta \to \mathcal{R}_+, y : \Theta \to \mathcal{R}_+, w' : \Theta^2 \to \mathcal{R} \qquad (5)$$

$$\text{such that (3) and (4) holds and}$$

$$w'(\theta, .) \in B\}$$

Standard arguments of Abreu, Pearce and Stacchetti [1] imply that the fixed point B^* is nonempty and compact.

The solution for the initial period is special, because the social planner is not bound by any promised utility to the agents who deviated the period before. Thus, a component planner in period 1 solves

$$\hat{V}_1(w_1, \theta_0) \quad = \quad \min_{w(.) \in L^\Theta} V_1(w(.), \theta_0)$$

$$\text{s.t.} \quad w(\theta_0) \quad = \quad w_1$$

where $V_1(w(.), \theta_0)$ is given by the Bellman equation above.

Denote the solution to the recursive program for $t > 1$ by optimal policy functions $c_t^* : L^\Theta \times \Theta^2 \to \mathcal{R}_+$, $y_t^* : L^\Theta \times \Theta^2 \to \mathcal{R}_+$ and $w_{t+1}^* : L^\Theta \times \Theta^3 \to \mathcal{W}$. Note that time 1 optimal policy functions have smaller domain, i.e. $c_1^* : \mathcal{W} \times \Theta^2 \to \mathcal{R}_+$, $y_1^* : \mathcal{W} \times \Theta^2 \to \mathcal{R}_+$ and $w_{t+1}^* : \mathcal{W} \times \Theta^3 \to \mathcal{W}$. Denote the whole collection of optimal policy functions by $\mathbf{c}^* = \{c_t^*\}_{t \geq 1}$, $\mathbf{y}^* = \{y_t^*\}_{t \geq 1}$, $\mathbf{w}'^* = \{w_{t+1}^*\}_{t \geq 1}$.

It is well known that in the case of iid shocks the state space of the component planner's problem is just (w, θ_-). Why is the state space now larger than that? The reason is that the last period shock now affects the probability distribution of current shocks and therefore the continuation value. The social planner observes only reported shocks and thus cannot distinguish between agents with identical reports but different shocks. However, he must still deliver promised utility to all such agents. Hence, different promised continuation value must be assigned, according to shock last period. Consistently with this intuition, notice that it is the $w'(\theta, .)$ section of continuation utility that becomes an argument of the value function, i.e. a section that keeps report constant (and truthful) and varies along the true shock.

The evolution of the distribution of continuation utility functions can be defined as follows Suppose φ_t is a distribution of continuation utility functions at time t and that $D \subseteq L^\Theta$. The distribution next period φ_{t+1} satisfies the difference equation

$$\varphi_{t+1}(D) = \int_M d\varphi_t d\theta_t d\theta_{t-1}$$

7

where $M(D) = \{(w(.), \theta_{t-1}, \theta_t) \in L^\Theta \times \Theta^2 : w'_{t+1}(w(.), \theta_{t-1}; \theta_t, \theta_t) \in D\}$. The first period distribution φ_1 is just a mass 1 on a function that solves period 1 component planner's problem.

We now make the relationship between the sequence problem and the component planner's recursive problem more precise. We first define an allocation that can be constructed from a recursive allocation. In particular, we are interested in allocation that is constructed from the optimal recursive allocation. Suppose $\tilde{W}_t(w_1, \hat{\theta}^{t-1}; \theta_{t-1})$[5] solve a difference equation $\tilde{W}_1(w_1, \hat{\theta}_1; \theta_1) = w'^*_1(w_1, \theta_0; \hat{\theta}_1, \theta_1)$ and $\tilde{W}_{t+1}(w_1, \hat{\theta}^t; \theta_t) = w'^*_{t+1}(\tilde{W}_t(w_1, \hat{\theta}^{t-1}; \hat{\theta}_{t-1}), \hat{\theta}_{t-1}; \hat{\theta}_t, \theta_t)$ for $t > 1$. Define an allocation $\tilde{\mathbf{C}}^*, \tilde{\mathbf{Y}}^*$ as

$$\tilde{C}^*_t(w_1, \hat{\theta}^t) = c^*_t(\tilde{W}_t(w_1, \hat{\theta}^{t-1}; \hat{\theta}_{t-1}), \hat{\theta}_{t-1}; \hat{\theta}_t)$$

$$\tilde{Y}^*_t(w_1, \hat{\theta}^t) = y^*_t(\tilde{W}_t(w_1, \hat{\theta}^{t-1}; \hat{\theta}_{t-1}), \hat{\theta}_{t-1}; \hat{\theta}_t)$$

Call $\tilde{\mathbf{C}}^*, \tilde{\mathbf{Y}}^*$ the allocation generated by the optimal recursive allocation. We have the following result:

Theorem 1 *Suppose* $\mathbf{c}^*, \mathbf{y}^*, \mathbf{w}'^*$ *solves the component planner's problem. Suppose also that* $\mathbf{C}^*, \mathbf{Y}^*$ *solves the sequence problem of the social planner. If there is a sequence of prices* $\{q_t\}_{t \geq 1}$ *and* w_1 *such that, for all t*

$$\int_{\theta_t \in \Theta} \int_{\theta_{t-1} \in \Theta} \int_{w(.) \in L^\Theta} [c^*_t(w(.), \theta_{t-1}, \theta_t) - y^*_t(w(.), \theta_{t-1}, \theta_t)] d\varphi_t d\theta_{t-1} d\theta_t = 0 \qquad (6)$$

Then

i) $u(\mathbf{C}^*, \mathbf{Y}^*, \theta_0) = w_1$

ii) $\tilde{\mathbf{C}}^* = \mathbf{C}^*$ *and* $\tilde{\mathbf{Y}}^* = \mathbf{Y}^*$.

Proof. See the Appendix. ∎

The theorem makes the following claim: Suppose we solve both the recursive problem and the sequence problem. If the prices are such that the resource constrain in the recursive problem happens to be cleared, then first, the lifetime utility delivered in the sequence problem is the same as the promised utility in the recursive problem and second, the consumption and output assignments are, loosely speaking, identical.

There is several distinct steps that lead to the proof of Theorem (1). First, it is shown, that there is a cost minimization problem for the social planner such that if the resource constraint holds for every period, then the optimum is the same as in the sequence problem (Lemma 7).

---

[5]The last argument of $\hat{W}_t$ is the true shock last period while the middle argument is the history of reports.

The core of the proof lies in the second step which shows that the cost minimization problem can be decentralized into a series of component planner's problems (Lemma 8). More precisely, if a recursive allocation solves the component planner's problem, then there will be an allocation defined in the sequence space such that the costs will be the same - namely the allocation generated by the recursive allocation. The main obstacle in proving this result is to show that the incentive compatibility constraint in the sequence problem can be related to the temporary incentive compatibility in the recursive component planner's problem. The incentive compatibility constraint involves checking all possible deviations while the temporary incentive compatibility constraint involves only one period deviations. The bridge between these two is built on the separability result of Phelan and Fernandes which is proved in Lemma 6 in the Appendix. Incentive compatibility constraint implies that the agent will prefer to tell the truth even if he deviated in the past. Thus, multiple deviation cannot make the agent better off unless there is also a one period deviation.

The proof of Theorem 1 itself then builds on the fact that the resource constraint (6) implies that a resource constraint (2) holds for the generated allocation. By virtue of Lemma 7, the generated allocation is identical to the allocation that maximizes the sequence problem, hence it delivers the same utility as the recursive allocation, which is w_1. Second part of the theorem states the most important implication: If we solve the recursive component planner's problem, we can construct an allocation which will solve the sequence problem of the social planner.

5 First Order Approach

The use of first order approach is complicated by the fact that the value function of the agent consists from a sum of two terms: $U(c(\hat{\theta}), \frac{y(\hat{\theta})}{\theta})$ and $\beta w'(\hat{\theta}, \theta)$. The first term is just period utility function and we know its properties, namely differentiability with respect to θ. This is not true about the other term, $\beta w'(\hat{\theta}, \theta)$. This function is endogenous to the social planner's problem and we do not know its properties immediately.

Fortunately, it turns out that we can pin down the properties of interest quite easily. The reason is that, if we fix the report $\hat{\theta}$, the term $w'(\hat{\theta}, \theta)$ is just an expected utility from a fixed allocation.[6] Thus, its dependence on θ is solely driven by the dependence of this expectations on θ, i.e. by the dependence of $\pi(., \theta)$ on θ. If this function is differentiable, then $w'(\hat{\theta}, \theta)$ is also differentiable with respect to θ. More precisely, the social planner will never be able to choose a non-differentiable

---

[6]This is because current report is fixed and all future reports are truthful.

function since he couldn't satisfy his constraints next period. The set B^* contains only differentiable functions. Lemma (9) in the Appendix shows this result more formally.

Next theorem derives necessary and sufficient conditions for a recursive allocation to be incentive compatible. It also derives an envelope condition that is necessary for an allocation to be incentive compatible. The derivation relies heavily on previously discussed result that the continuation utility function is differentiable in agent's type. The proof draws upon the envelope theorem of Milgrom and Segal ([10]), Theorems 1 and 2.[7]

Theorem 2 *The allocation is incentive compatible if and only if*

i)

$$\frac{dy(\hat{\theta})}{d\hat{\theta}} \frac{U_l(c(\hat{\theta}), \frac{y(\hat{\theta})}{\theta})}{U_c(c(\hat{\theta}), \frac{y(\hat{\theta})}{\theta})} \frac{1}{\theta} + \beta \frac{\frac{\partial}{\partial \hat{\theta}} w'(\hat{\theta}, \theta)}{U_c(c(\hat{\theta}), \frac{y(\hat{\theta})}{\theta})}$$

is increasing in θ for all $\hat{\theta}$ and

ii)

$$U(c(\theta), \frac{y(\theta)}{\theta}) + \beta w'(\theta) = \int_0^\theta [-U_l(\tilde{\theta}) \frac{y(\tilde{\theta})}{\tilde{\theta}^2} + \beta \frac{\partial}{\partial \theta} w'(\tilde{\theta}|\tilde{\theta})] d\tilde{\theta} + U_0$$

Proof. See the Appendix. ∎

The condition in the first part of the theorem is a necessary and sufficient condition for an allocation to be incentive compatible. It is more complicated than in case of iid shocks because of the second term on the right hand side. In case of iid shocks, necessary and sufficient condition boils down to a requirement that $y(\theta)$ is increasing (provided that the utility function satisfies Spence-Mirrlees condition). In this case, however, this is neither sufficient nor necessary. In principle, we can have $y(\theta)$ decreasing, if $\frac{\partial}{\partial \hat{\theta}} w'$ increases sufficiently fast in θ. We will have to check this condition to determine if it holds for particular solution.

The validity of the first order approach indicates that not all the continuation utility function is needed for the recursive formulation. In particular, everything except for marginal continuation utility is irrelevant for the incentive compatibility - hence it is irrelevant for the dynamic program.

Define $g(\theta) = \frac{\partial}{\partial \theta} w'(\theta, \theta)$ We can then write the dynamic program of a component planner as follows:

---

[7]The proof of second order conditions relies on differentiability of the policy functions. I will later generalize the result.

$$V(w, g, \theta_-) = \min_{c,y,w',g'} \int_\theta [c(\theta) - y(\theta) + \beta V(w'(\theta), g'(\theta), \theta)] \pi(\theta|\theta_-) d\theta$$

$$s.t.$$

$$w = \int_\theta [U(c(\theta), \frac{y(\theta)}{\theta}) + \beta w'(\theta)] \pi(\theta|\theta_-) d\theta \tag{7}$$

$$g = \int_\theta [U(c(\theta), \frac{y(\theta)}{\theta}) + \beta w'(\theta)] \pi_2(\theta|\theta_-) d\theta \tag{8}$$

$$U(c(\theta), \frac{y(\theta)}{\theta}) + \beta w'(\theta) = \int_0^\theta [-U_l(\tilde{\theta}) \frac{y(\tilde{\theta})}{\tilde{\theta}^2} + \beta g'(\tilde{\theta})] d\tilde{\theta} + U_0 \tag{9}$$

$$(w'(\theta), g'(\theta)) \in b^*(\theta)$$

where $b^*(\theta)$ is a fixed point of the following operator:

$$T(b)(\theta_-) = \{(w, g) \in \mathcal{W} \times \mathcal{R}_+ : \exists c : \Theta \to \mathcal{R}_+, y : \Theta \to \mathcal{R}_+, w' : \Theta^2 \to \mathcal{R}$$

$$\text{such that (7), (8) and (9) holds and}$$

$$(w'(\theta), g'(\theta)) \in b(\theta)\}$$

The second constraint, (8) is a new feature of the dynamic program. It is a marginal version of the threat keeping constraint. For an agent with last period shock θ_- the social planner to restricted to increase the marginal continuation value for at rate g.

The significance of the first order approach is that it does not increase the dimensionality of the dynamic program as compared to Fernandes and Phelan, but allows for continuum of shocks.

6 Capital Income Taxation

To study capital income taxes I now consider a very simplified economy: An economy with taste shocks and logarithmic utility. Formally, the mapping to the previous formulation is given by the following specification:

$$U(c, \frac{y}{\theta}) = \theta \ln c$$

Since this economy is not concerned with the determination of output, it says nothing about the intratemporal wedge and hence about labor taxes. Nevertheless, a taste shock economy is a convenient tool if one wants to focus on capital income taxation.

The nature of the intertemporal wedge is rather different than in an economy with productivity shocks and labor taxes. Golosov Kocherlakota and Tsyvinski show that for a very general set of conditions we should expect the intertemporal wedge to be strictly positive. In other words, in an economy with zero capital taxes, individuals would tend to oversave. This result holds if utility is additively separable between consumption and the underlying shock. This is not the case in the taste shock economy, however. In fact, the intuition is quite the opposite here. High shock individuals would like to borrow from the future since they would like today to consume more than what the social planner offers them. Thus, we should expect that the intertemporal wedge should be negative for some fraction of high shock individuals. We will see that this intuition is confirmed.

I provide results for partial equilibrium where the intertemporal price is fixed at a value equal to the discount factor: $q = \beta$. In later versions of the paper I will generalize the results for the general equilibrium case where the intertemporal price of consumption is endogenous.

6.1   IID shocks

It is instructive to begin with the case of iid taste shocks: $\pi(\theta|\theta_-) = \pi(\theta)$. This section also provides some insights into the solution procedure which will become more complicated in the next section. This section is closely related to the results of Atkeson and Lucas [3], the main difference being the assumption of fixed intertemporal price and continuum of shocks.

We make the following change of variable: $u = \ln(c)$. It is easy to show that in this case the incentive compatibility constraint becomes

$$\theta u(\theta) + \beta w'(\theta) = \int_0^\theta u(\tilde{\theta})d\tilde{\theta} + \beta w_0 \tag{10}$$

The social planner responsible for delivering promised utility w allocates consumption according to agent's reports. In the light of the new notation the social planner minimizes its cost

$$V(w) = \min_{u,w'} \int [e^{u(\theta)} + \beta V(w'(\theta))]\pi(\theta)d\theta$$

subject to the incentive compatibility constraint (10) and the promise keeping constraint

$$w = \int [\theta u(\theta) + \beta w'(\theta)]\pi(\theta)d\theta$$

The following Lemma simplifies the computations tremendously as it allows us to normalize the state variable w :

Lemma 3 *Let $\gamma = E(\theta)$. The solution to the social planner's problem is given by the value function*

$$V(w) = Ae^{\frac{1-\beta}{\gamma}w}$$

and the policy functions

$$
\begin{aligned}
u^*(w,\theta) &= \frac{1-\beta}{\gamma}w + \bar{u}(\theta) \\
w'^*(w,\theta) &= w + \bar{w}'(\theta)
\end{aligned}
$$

for some functions $u(\theta)$ and $w'(\theta)$ and some constant A. Moreover, the function $w'(\theta)$ satisfies $Ew' < 0$.

The proof of this theorem is not provided as it is a special case of similar Lemma (4) in the next section. The importance of this result is, however, hard to overstate. We can solve the whole model as a simple static-looking problem and easily recover the policy functions u and w back.

In solving the static-looking problem I follow Mirrlees [11] in using variational approach to obtain necessary conditions for the optimum. We can write the necessary conditions as a set of differential equations, with appropriate boundary conditions.

The following change of variables becomes convenient: set $t = \ln(\theta)$, $x = e^{(1-\beta)\gamma\bar{w}'}$ and $y = (\frac{e^{\bar{u}}}{\theta} - \lambda x)$. Then it is easy to show that the optimum follows the following differential equations in x and y:

$$
\begin{aligned}
\frac{dx}{dt} &= \frac{\sigma y + \lambda(1 - 2x)}{\lambda + e^{-t}\frac{\beta}{(1-\beta)\gamma}(\frac{y}{x} + \lambda)} \\
\frac{dy}{dt} &= \lambda(x-1) - y(1+\sigma)
\end{aligned}
$$

where $\sigma(\theta) = \frac{\theta f'}{f}$ and λ is the Lagrange multiplier on the resource constraint. The boundary conditions for the two differential equations are given by $y_0 = y_\infty = 0$.

I will defer the discussion of quantitative results for later section.

6.2 Persistent shocks

I now focus on a case where any taste shock that an individual incurs persist over time. Consider again the change of variable $u = \ln(c)$. Consistently with the general results of section (5), the social planner that is allocated to an agent with promised utility w, promised marginal continuation utility

g and last period shock θ_-solves the following problem[8]:

$$V(w, g, \theta_-) = \min_{u, w', g'} \int [e^{U(\theta)} + \beta V(W'(\theta), G'(\theta), \theta)] \pi(\theta) d\theta$$

and he is constrained by the promise keeping and marginal threat keeping constraints

$$w = \int [\theta U(\theta) + \beta W'(\theta)] \pi(\theta | \theta_-) d\theta$$

$$g = \int [\theta U(\theta) + \beta W'(\theta)] \pi_\theta(\theta | \theta_-) d\theta$$

and the incentive compatibility constraint which takes the form

$$\theta U(\theta) + \beta W'(\theta) = \int\limits_0^\theta [U(\varepsilon) + \beta G'(\varepsilon)] d\varepsilon + \beta W_0$$

Although the problem is now considerably more complicated, we can still show that promised utility can be usefully normalized. The following Lemma parallels Lemma 3:

Lemma 4 *Let $\gamma_{\theta_-} = E(\theta | \theta_-)$ and $\varphi(\theta_-) = \frac{1-\beta}{\gamma_{\theta_-}} \frac{d}{d\theta_-} \gamma_{\theta_-}$. Suppose that $\hat{g} = g - \varphi(\theta_-) w$. Then the solution to the social planner's problem is given by a value function*

$$V(w, g, \theta_-) = v(\tilde{g}, \theta_-) e^{\frac{1-\beta}{\gamma_{\theta_-}} w}$$

and the policy functions

$$U(w, g, \theta_-, \theta) = \frac{1-\beta}{\gamma_{\theta_-}} w + u(\tilde{g}, \theta_-, \theta)$$

$$W'(w, g, \theta_-, \theta) = w + w'(\tilde{g}, \theta_-, \theta)$$

$$G'(w, g, \theta_-, \theta) = g'(\tilde{g}, \theta_-, \theta)$$

for some functions $u(\tilde{g}, \theta_-, \theta)$, $w'(\tilde{g}, \theta_-, \theta)$ and $g'(\tilde{g}, \theta_-, \theta)$.

Proof. See the Appendix. ∎

Normalization of promised utility to 0 thus yields the following Bellman equation:

$$v(g, \theta_-) = \min_{u, w', z} \int [e^{u(\theta)} + \beta e^{\frac{1-\beta}{\gamma_{\theta_-}} w'(\theta)} v(z(\theta), \theta)] \pi(\theta | \theta_-) d\theta$$

---

[8]I do not formulate explicitly the feasibility set for the social planner. It is assumed that the solution lies in the interior.

14

s.t.

$$0 = \int [\theta u(\theta) + \beta w'(\theta)] f(\theta|\theta_-) d\theta$$

$$g = \int [\theta u(\theta) + \beta w'(\theta)] f_{\theta_-}(\theta|\theta_-) d\theta$$

$$\theta u(\theta) + \beta w'(\theta) = \int_0^\theta [u(\varepsilon) + \beta z(\varepsilon) + \beta \varphi(\varepsilon) w'(\varepsilon)] d\varepsilon + \beta w_0'$$

where we use the fact that if $g'(\theta)$ is the marginal continuation value then, by virtue of our rescaling, $z(\theta) = g'(\theta) - \varphi(\theta) w'(\theta)$ enters the normalized value function.

6.3 Lognormal distribution

Even with this result in hand, the solution is not a simple one. Fortunately, imposing an additional restriction on the stochastic process for shocks yields additional simplifications. Assume that the taste shock can be written as $\theta = e^t$ where

$$t' = (1 - \rho)\mu + \rho t + \zeta.$$

The innovation ζ is normally distributed iid random variable with mean zero and variance ν^2. The unconditional mean of the taste shock is given by parameter μ and the persistence is driven by ρ.

A crucial implication of lognormality is, however, that we can normalize one state variable - the last period shock - arbitrarily and therefore reduce the dimension of the dynamic program to one state variable only - promised marginal utility.

Lemma 5 *The value function satisfies $v(g, \theta_-) = v(\theta_-^{1-\rho} g, 1)$. The policy functions satisfy $u(g, \theta, \theta_-) = u(\theta_-^{1-\rho} g, \frac{\theta}{\theta_-^\rho}, 1)$, $w'(g, \theta, \theta_-) = \theta_-^\rho w'(\theta_-^{1-\rho} g, \frac{\theta}{\theta_-^\rho}, 1)$ and $z(g, \theta, \theta_-) = z(\theta_-^{1-\rho} g, \frac{\theta}{\theta_-^\rho}, 1)$.*

Proof. See the Appendix. ∎

To economize on notation, we write $u(g, \theta, 1) = u(g, \theta)$, $w'(g, \theta, 1) = w'(g, \theta)$ and $z(g, \theta, 1) = z(g, \theta)$. The value function is written concisely as $v(g, 1) = v(g)$.

Using the new notation, we can write the dynamic program as

$$v(g) = \min_{u, w', h} \int [e^{u(\theta)} + \beta e^{\frac{1-\beta}{\gamma} w'(\theta)} v(h(\theta))] \pi(\theta) d\theta$$

15

s.t.

$$0 = \int [\theta u(\theta) + \beta w'(\theta)]\pi(\theta)d\theta$$

$$g = \int [\theta u(\theta) + \beta w'(\theta)]\pi_{\theta_-}(\theta)d\theta$$

$$\theta u(\theta) + \beta w'(\theta) = \int_0^\theta [u(\varepsilon) + \beta\frac{h(\varepsilon)}{\varepsilon^{1-\rho}} + \beta\rho(1-\beta)\frac{w'(\varepsilon)}{\varepsilon}]d\varepsilon + \beta w'_0$$

where $h(g,\theta) = \theta^{1-\rho}z(g,\theta)$ and $\gamma = E(\theta)$. In addition, lognormal distribution implies that $\pi_{\theta_-}(\theta) = \rho[\ln\theta - (1-\rho)\mu]\pi(\theta)$.

It becomes useful to partition the problem in two parts. Introduce again new variables $t = \ln(\theta)$, Define $x = e^{\frac{1-\beta}{\gamma}w'}$ and $y = \frac{\frac{e^u}{\theta} - \frac{1-\beta}{\gamma}v'x}{1-\rho(1-\beta)}$. Denote the Lagrange multipliers on the promise keeping and marginal threat keeping constraints as λ_w and λ_g. We can show that for some fixed functions $v(g)$ and $h(g,\theta)$ the functions $x(t)$ and $y(t)$ satisfy

$$\frac{dx}{dt} = \frac{\eta(\sigma + \eta - 1)y + \eta(\lambda_w + r\lambda_g) + \beta(\eta y + \frac{1-\beta}{\gamma}v'x)(e^{-(1-\rho)t}h + e^{-t}\rho\gamma\ln x) - (1+\eta)\frac{1-\beta}{\gamma}v'x - ye^{\rho t}\frac{dh}{dt}}{\frac{1-\beta}{\gamma}v' + \frac{\beta\gamma}{1-\beta}e^{-t}(\eta\frac{y}{x} + \frac{1-\beta}{\gamma}v')}$$

$$\frac{dy}{dt} = -\lambda_w - r\lambda_g + \frac{1-\beta}{\gamma}v'x - y(\sigma + 2 - \eta)$$

where $\eta = 1 - \rho(1-\beta)$. The boundary conditions are the same as in the iid. case, i.e. $y_0 = y_\infty = 0$. Here the term $r = \frac{\pi_{\theta_-}(\theta)}{\pi(\theta)}$ represents the relative change in the density with respect to the last period shock and the statistics σ is defined as before. Thus, we can easily solve for the function x and y in dependence on the rest of the policy functions and then provide more time expensive search for functions v and h. The numerical results are presented in the next section.

7 Numerical simulations

I look at the case of random walk with $\rho = 1$. Although the lognormal distribution is unbounded, I impose an upper bound and discretize the space of shocks on a grid with 500 points.[9]

Figure 1 gives the intertemporal wedge for the iid case. The intuition from section 6 is confirmed. The intertemporal wedge is positive for low shock agents as they would like to save more but becomes negative for high shock ones, to deter them from borrowing from future. the intertemporal wedge is quite sizable, ranging to almost 40% subsidy for the high taste agents.

---

[9]The discretization must be rather fine in order to allow to discretize the differential equations without much precision loss.
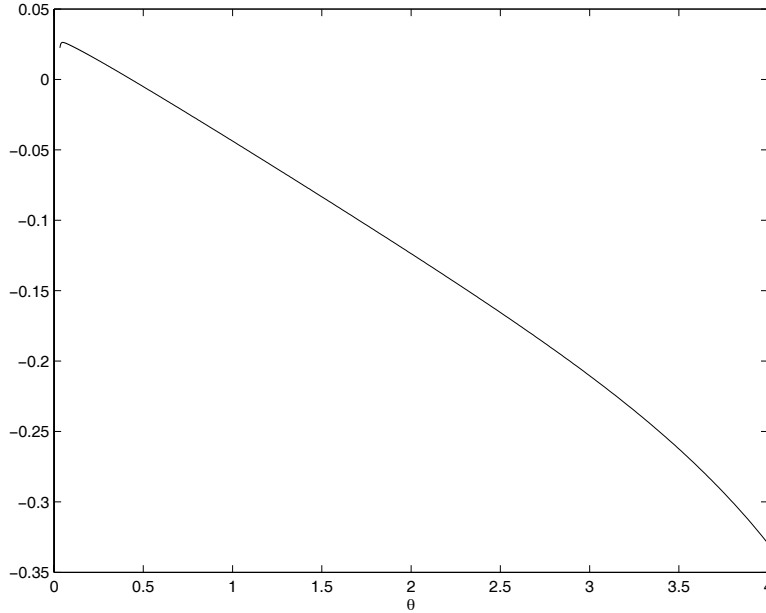
Figure 1: Intertemporal wedge, IID shocks

For the case of persistent shocks, Figure 2 illustrates the intertemporal wedge for various values of marginal continuation value. We can see that while the pattern of the intertemporal wedge is essentially unchanged (positive for low shock agents and negative for high shock agents) the size of the effect is significantly smaller. Maximal value of the intertemporal wedge reaches 15%, three times less than in the iid case.

8 Conclusions

The contribution of this paper is twofold. First, It develops a method of solving dynamic private information models with persistent shocks which is simple enough to be solved numerically. Second, it investigates the nature of optimal capital income taxes under realistic assumption of high shock persistence. The finding is that the intertemporal wedge is possibly as much as three times reduced when shocks are persistent.

Future versions of this paper will consider several extensions. First, one need to look at general equilibrium framework and assess the whole distribution of shocks and other state variables to get more meaningful results. Second, I plan to return the labor supply back to the model and investigate the effect of shock persistence on labor income taxes. Another unresolved question is the problem of implementation. While the intertemporal wedge has been found, the question of
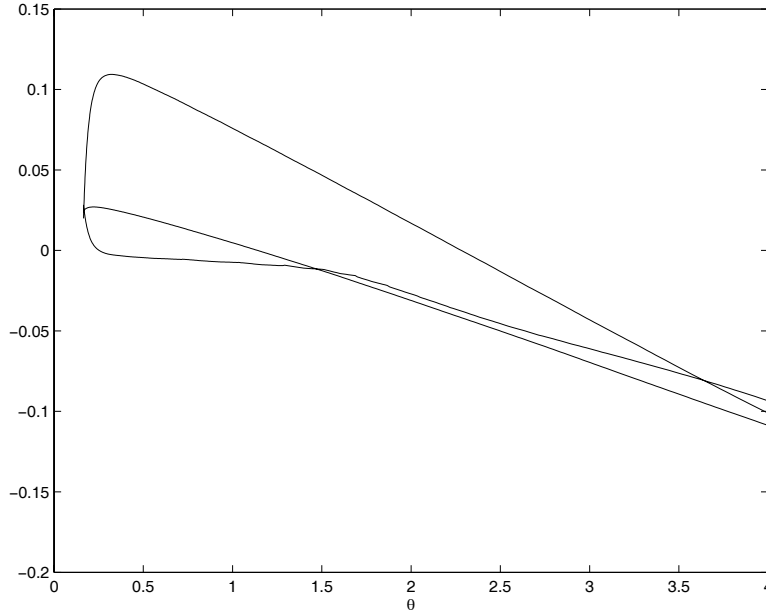
Figure 2: Intertemporal wedge, persistent shocks

how to implement it remains. It follows from Lemma 3 and 4 that capital income taxes will be independent of individual's assets. They will crucially depend on consumption. The exact nature will appear in future versions of the paper.

More generally, one may want dispose the assumption that productivity shocks are exogenous and model persistence of productivity shocks as partly endogenous. Human capital is a natural candidate through which income persistence is generated. Such framework allows us to analyze richer set of policies, including policies that promote human capital accumulation directly, not only through income or capital income taxes. Bohacek and Kapicka [4] proceed in this direction.

References

[1] Abreu, Dilip, David Pearce and Ennio Staccheti, " Toward a Theory of Discounted Repeated Games with Imperfect Information", Econometrica, Vol. 58, No. 5. (Sep., 1990), pp. 1041-1063

[2] Albanesi, Stefania and Christopher Sleet, "Dynamic optimal Taxation with Private Information", Discussion Paper 140, Federal Reserve Bank of Minneapolis (2003)

[3] Andrew Atkeson and Robert E. Lucas, Jr, "On Efficient Distribution With Private Information", The Review of Economic Studies, Vol. 59, No. 3. (Jul., 1992), pp. 427-453

[4] Bohacek Radim and Marek Kapicka, "Optimal Human Capital Policies", mimeo (2004)

[5] Paolo Costantini and Ferruccio Fontanella, "Shape-preserving Bivariate Interpolation", SIAM Journal on Numerical Analysis, Vol. 27, No. 2. (Apr., 1990), pp. 488-506

[6] Fernandes, Ana and Chris Phelan, "A recursive formulation for repeated agency with history dependence" Journal of Economic Theory **91,** 223-247 (2000)

[7] Golosov, Mikhail, Narayana Kocherlakota and Aleh Tsyvinski, "Optimal Indirect and Capital Taxation", Staff Report 273, Federal Reserve Bank of Minneapolis (2001)

[8] Heathcote, Jonathan, Kjetil Storesletten and Giovanni L. Violante, "The Macroeconomic Implications of Rising Wage Inequality in the United States", mimeo (2004)

[9] Kocherlakota, Narayana, "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation", Stanford University, mimeo (2003)

[10] Milgrom, Paul and Ilya Segal, "Envelope Theorems for Arbitrary choice Sets", Econometrica, **70**, 583-601 (2002)

[11] Mirrlees, James A., "An Exploration in the Theory of Optimum Income Taxation", The Review of Economic Studies **38**, 175-208 (1971)

[12] Mirrlees, James A., "Optimal Tax Theory: A Synthesis", Journal of Public Economics, **6** 327-358 (1976)

[13] Mirrlees, James A., "The Theory of Optimal Taxation", in K.J. Arrow and M.D. Intriligator, eds., Handbook of Mathematical Economics, vol. III, chapter 24 (1986)

[14] Shimer, Robert and Ivan Werning, "Optimal Unemployment Insurance with Sequential Search", University of Chicago mimeo (2003)

[15] Werning, Ivan, "Dynamic Optimal Taxation", University of Chicago, mimeo (2001)

9   Appendix

The following Lemma shows that the incentive compatibility constraint implies that after any history (not necessarily a truthful one) the agent still prefers to tell the truth.

Lemma 6 *Consider an allocation $\mathbf{C}, \mathbf{Y}$ and any θ_0. Define $\hat{\boldsymbol{\theta}}_t^\infty = \{\hat{\theta}_j\}_{j \geq t}$ to be a reporting strategy from period t on. If $\mathbf{C}, \mathbf{Y}$ is incentive compatible then for any history of reports $\hat{\theta}^{t-1}$, any last period shock θ_{t-1} and for any reporting strategy $\hat{\boldsymbol{\theta}}_t^\infty$*

$$\sum_{j \geq 0} \int_{\theta_t^{t+j} \in \Theta^j} \beta^j U(C_{t+j}(w_1, \hat{\theta}^{t-1}, \theta_t^{t+j}), \frac{Y_{t+j}(w_1, \hat{\theta}^{t-1}, \theta_t^{t+j})}{\theta_{t+j}}) \Pi(\theta_t^{t+j} | \theta_{t-1}) d\theta^t$$

$$\geq \sum_{j \geq 0} \int_{\theta_t^{t+j} \in \Theta^j} \beta^{j-1} U(C_{t+j}(w_1, \hat{\theta}^{t-1}, \hat{\theta}_t^{t+j}(\theta_t^{t+j})), \frac{Y_{t+j}(w_1, \hat{\theta}^{t-1}, \hat{\theta}_t^{t+j}(\theta_t^{t+j}))}{\theta_{t+j}}) \Pi(\theta_t^{t+j} | \theta_{t-1}) d\theta^t$$

Proof. Define $F_t(w_1, \hat{\theta}^t, \theta_t) = U(C_t(w_1, \hat{\theta}^t), \frac{Y_t(w_1, \hat{\theta}^t)}{\theta_t})$ to simplify the notation.

The proof is trivial for $\hat{\theta}^{t-1} = \theta^{t-1}$. For $\hat{\theta}^{t-1} \neq \theta^{t-1}$ suppose that there is a reporting strategy $\tilde{\boldsymbol{\theta}}_t^\infty$ such that

$$\sum_{j \geq 0} \int_{\theta_t^{t+j} \in \Theta^j} \beta^j F_{t+j}(w_1, (\hat{\theta}^{t-1}, \theta_t^{t+j}), \theta_{t+j}) \Pi(\theta_t^{t+j} | \theta_{t-1}) d\theta^t$$

$$< \sum_{j \geq 0} \int_{\theta_t^{t+j} \in \Theta^j} \beta^j F_{t+j}(w_1, (\hat{\theta}^{t-1}, \tilde{\theta}_t^{t+j}(\theta_t^{t+j})), \theta_{t+j}) \Pi(\theta_t^{t+j} | \theta_{t-1}) d\theta^t$$

For this equation to hold, the reporting strategy must improve upon the truthful strategy for a set $D \in \Theta$ of nonzero measure. Thus, for all $\theta_t \in D$,

$$F_t(w_1, (\hat{\theta}^{t-1}, \theta_t), \theta_t) + \sum_{j \geq 1} \int_{\theta_{t+1}^{t+j} \in \Theta^j} \beta^j F_{t+j}(w_1, (\hat{\theta}^{t-1}, \theta_t^{t+j}), \theta_{t+j}) \Pi(\theta_{t+1}^{t+j} | \theta_t) d\theta^t$$

$$< F_t(w_1, (\hat{\theta}^{t-1}, \tilde{\theta}_t(\theta_t)), \theta_t) + \sum_{j \geq 1} \int_{\theta_{t+1}^{t+j} \in \Theta^j} \beta^j F_{t+j}(w_1, (\hat{\theta}^{t-1}, \tilde{\theta}_t^{t+j}(\theta_t^{t+j})), \theta_{t+j}) \Pi(\theta_{t+1}^{t+j} | \theta_t) d\theta^t \ (11)$$

Define another strategy $\bar{\boldsymbol{\theta}}_t^\infty$ that is identical with $\tilde{\boldsymbol{\theta}}_t^\infty$ if $\theta_t \in D$ and involves truthtelling if $\theta_t \notin D$. By inequality (11), this strategy also dominates the truthtelling strategy.

Consider now an agent with last period shock $\hat{\theta}_{t-1}$. We have

$$\sum_{j \geq 0} \int_{\theta_t^{t+j} \in \Theta^j} \beta^j F_{t+j}(w_1, (\hat{\theta}^{t-1}, \theta_t^{t+j}), \theta_{t+j}) \Pi(\theta_t^{t+j} | \hat{\theta}_{t-1}) d\theta^t$$

$$= \int_{\theta_t \notin D} \sum_{j \geq 0} \int_{\theta_{t+1}^{t+j} \in \Theta^j} \beta^j F_{t+j}(w_1, (\hat{\theta}^{t-1}, \theta_t^{t+j}), \theta_{t+j}) \Pi(\theta_t^{t+j} | \hat{\theta}_{t-1}) d\theta^t$$

$$+ \int_{\theta_t \in D} \left\{ F_t(w_1, (\hat{\theta}^{t-1}, \theta_t), \theta_t) + \sum_{j \geq 1} \int_{\theta_{t+1}^{t+j}} \beta^j F_{t+j}(w_1, (\hat{\theta}^{t-1}, \theta_t^{t+j}), \theta_{t+j}) \Pi(\theta_t^{t+j} | \theta_t) d\theta^t \right\} \pi(\theta_t | \hat{\theta}_{t-1}) d\theta_t$$

$$< \int_{\theta_t \notin D} \sum_{j \geq 0} \int_{\theta_{t+1}^{t+j} \in \Theta^j} \beta^j F_{t+j}(w_1, (\hat{\theta}^{t-1}, \theta_t^{t+j}), \theta_{t+j}) \Pi(\theta_t^{t+j} | \hat{\theta}_{t-1}) d\theta^t$$

$$+ \int_{\theta_t \in D} \left\{ F_t(w_1, (\hat{\theta}^{t-1}, \bar{\theta}_t(\theta_t)), \theta_t) + \sum_{j \geq 1} \int_{\theta_{t+1}^{t+j}} \beta^j F_{t+j}(w_1, (\hat{\theta}^{t-1}, \bar{\theta}_t^{t+j}(\theta_t^{t+j})), \theta_{t+j}) \Pi(\theta_t^{t+j} | \theta_t) d\theta^t \right\} \pi(\theta_t | \hat{\theta}_{t-1}) d\theta_t$$

$$= \sum_{j \geq 0} \int_{\theta_t^{t+j} \in \Theta^j} \beta^j F_{t+j}(w_1, (\hat{\theta}^{t-1}, \bar{\theta}_t^{t+j}(\theta_t^{t+j})), \theta_{t+j}) \Pi(\theta_t^{t+j} | \hat{\theta}_{t-1}) d\theta^t.$$

The first equality follows from partitioning the space of current shocks to D and its complement. The inequality follows from the inequality (11), i.e. from the fact that for all $\theta_t \in D$, the allocation $\bar{\boldsymbol{\theta}}_t^{\infty}$ is preferred to truthtelling. The last equality follows from definition of $\bar{\boldsymbol{\theta}}_t^{\infty}$.

Thus, we have shown that there is a reporting strategy $\bar{\boldsymbol{\theta}}_t^{\infty}$ that improves upon truthtelling even after truthful history, which is a contradiction. ∎

As a first step toward the proof of Theorem 8 we start with constructing the following sequence cost minimization problem of the social planner: Suppose that there are intertemporal prices of consumption $\{q_t\}_{t \geq 1}$ satisfying $\sum_{t \geq 1} \prod_{i=1}^{t} q_i \leq +\infty$. The social planner minimizes cost of delivering lifetime utility w_1:

$$\Omega(w_1, \theta_0) = \min_{\mathbf{C}, \mathbf{Y}} \sum_{t \geq 1} \int_{\theta^t \in \Theta^t} (\prod_{i=1}^{t-1} q_i)[c_t(\theta^t) - y_t(\theta^t)] \Pi(\theta^t | \theta_0) d\theta^t$$

s.t.

$$u(\mathbf{C}, \mathbf{Y}, \theta_0) = w_1 \tag{A1}$$

$$u(\mathbf{C}, \mathbf{Y}, \theta_0) \geq u(\mathbf{C} \circ \hat{\boldsymbol{\theta}}, \mathbf{Y} \circ \hat{\boldsymbol{\theta}}, \theta_0) \qquad \forall \hat{\boldsymbol{\theta}} \text{ s.t. } \mathbf{Y} \circ \hat{\boldsymbol{\theta}} \leq \boldsymbol{\theta}$$

where the last inequality is just the incentive compatibility constraint (1). Next Lemma shows that this is just an equivalent way of writing the social planner's sequence problem.

Lemma 7 *Suppose that there exist an allocation* $\mathbf{C}^*, \mathbf{Y}^*$, *initial utility entitlement* w_1 *and prices* $\{q_t\}_{t \geq 1}$ *satisfying* $\sum_{t \geq 1} \prod_{i=1}^{t} q_i \leq +\infty$ *such that*

i) $\mathbf{C}^*, \mathbf{Y}^*$ *solves the sequence cost minimization problem of the social planner and*

ii) for all t, for all $\theta^{t-1} \in \Theta^{t-1}$

$$\int_{\theta_t \in \Theta} [C_t(\theta^t) - Y_t(\theta^t)]\Pi(\theta^t|\theta_0)d\theta_t = 0 \tag{A2}$$

Then $\mathbf{C}^*, \mathbf{Y}^*$ *solves the sequence problem of the social planner.*

Proof. First consider a relaxed problem with inequality $u(\mathbf{C}, \mathbf{Y}, \theta_0) \geq w_1$ instead of constraint (A1). Suppose that $\hat{\mathbf{C}}^*, \hat{\mathbf{Y}}^*$ solves the sequence problem. By construction, $\mathbf{C}^*, \mathbf{Y}^*$ satisfies the resource constraint for all periods and the incentive compatibility constraint. Thus, we must have $u(\hat{\mathbf{C}}^*, \hat{\mathbf{Y}}^*, \theta_0) \geq u(\mathbf{C}^*, \mathbf{Y}^*, \theta_0)$. This in turn implies that $\hat{\mathbf{C}}^*, \hat{\mathbf{Y}}^*$ satisfies all the constraint of the relaxed program and thus $u(\hat{\mathbf{C}}^*, \hat{\mathbf{Y}}^*, \theta_0) \leq u(\mathbf{C}^*, \mathbf{Y}^*, \theta_0)$ implying that $u(\hat{\mathbf{C}}^*, \hat{\mathbf{Y}}^*, \theta_0) = u(\mathbf{C}^*, \mathbf{Y}^*, \theta_0)$ and that $\hat{\mathbf{C}}^* = \mathbf{C}^*$ and $\hat{\mathbf{Y}}^* = \mathbf{Y}^*$. Finally, standard arguments imply that $u(\mathbf{C}, \mathbf{Y}, \theta_0) \geq w_1$ will hold as a strict equality. ∎

Next Lemma starts with the cost minimization problem and proves that the sequence cost minimization problem can be as well written as a dynamic program decentralized among different component planners. This Lemma thus shows that the optimal allocation can be decentralized.

Lemma 8

i) Suppose $\mathbf{c}, \mathbf{y}, \mathbf{w}'$ *is a recursive allocation and that* $\mathbf{C}^*, \mathbf{Y}^*$ *is an allocation generated by the recursive allocation. Then* $\mathbf{C}^*, \mathbf{Y}^*$ *satisfies incentive compatibility constraint 1 and resource constraint 2 and* $\Omega(w_1, \theta_0) \leq \hat{V}_1(w_1, \theta_0)$ *for all* w_1, θ_0

ii) Suppose $\mathbf{C}^*, \mathbf{Y}^*$ *is an allocation that solves the sequence cost minimization problem. Than there exists a recursive allocation* $\mathbf{c}, \mathbf{y}, \mathbf{w}'$ *such that the recursive allocation solves component planner's problem and* $\hat{V}_1(w_1, \theta_0) \leq \Omega(w_1, \theta_0)$ *for all* w_1, θ_0

iii) $\hat{V}_1(w_1, \theta_0) = \Omega(w_1, \theta_0)$

Proof of part i). It is immediate that $\mathbf{C}^*, \mathbf{Y}^*$ is an allocation. We need to show that it also delivers expected utility w_1 and that it is incentive compatible.

i.1 (promise keeping) We will first show that the promise keeping constraint holds. To simplify notation, let

$$
\begin{aligned}
f_t^*(w(.),\theta_-;\hat{\theta},\theta) &= U(c_t^*(w(.),\theta_-;\hat{\theta}),\frac{y_t^*(w(.),\theta_-;\hat{\theta})}{\theta}) \\
F_t^*(w_1,\hat{\theta}^t,\theta_t) &= U(C_t^*(w_1,\hat{\theta}^t),\frac{Y_t^*(w_1,\hat{\theta}^t)}{\theta_t})
\end{aligned}
$$

i.e f_t^* represents period utility resulting from the recursive allocation and F_t^* represents period utility resulting from the allocation. Recall from the definition of an allocation generated by the recursive formulation, that we have defined a continuation utility $W_t(w_1,\hat{\theta}^{t-1};\theta_{t-1})$ that satisfies initial condition $W_1(w_1,\varnothing;\theta_0) = w_1$ and solves a difference equation $\tilde{W}_2(w_1,\hat{\theta}_1;\theta_1) = w_1^*(w_1,\theta_0;\hat{\theta}_1,\theta_1)$ and $\tilde{W}_{t+1}(w_1,\hat{\theta}^t;\theta_t) = w_{t+1}^*(\tilde{W}_t(w_1,\hat{\theta}^{t-1},\hat{\theta}_{t-1}),\hat{\theta}_{t-1};\hat{\theta}_t,\theta_t)$ for $t > 1$.

Fix $w_1,\hat{\theta}^{t-1}$ and θ_{t-1}. The promise keeping constraint of the component planner implies that

$$
\begin{aligned}
&W_t(w_1,\hat{\theta}^{t-1},\theta_{t-1}) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\text{A3}) \\
&= \int_{\theta_t}[f_t^*(W_t(w_1,\hat{\theta}^{t-1};\theta_{t-1}),\theta_{t-1};\theta_t,\theta_t)) + \beta w_{t+1}'(W_t(w_1,\hat{\theta}^{t-1};\theta_{t-1}),\theta_{t-1};\theta_t,\theta_t)]\pi(\theta_t|\theta_{t-1})d\theta_t
\end{aligned}
$$

From the definition of W_t and F_t the right hand side can be as well written as

$$
\int_{\theta_t}[F_t^*(w_1,(\hat{\theta}^{t-1},\theta_t);\theta_t) + \beta W_{t+1}(w_1,(\hat{\theta}^{t-1},\theta_t);\theta_t)]\pi(\theta_t|\theta_{t-1})d\theta_t.
$$

Using component planner's promise keeping constraint again and expanding the last term, we have an expression for the right hand side

$$
\begin{aligned}
&\int_{\theta'}F_t^*(w_1,(\hat{\theta}^{t-1},\theta_t);\theta_t)\pi(\theta_t|\theta_{t-1})d\theta_t \\
&+\beta\int_{\theta_t,\theta_{t+1}}[f_{t+1}^*(W_{t+1}(w_1,(\hat{\theta}^{t-1},\theta_t);\theta_t),\theta_t;\theta_{t+1},\theta_{t+1})\pi(\theta_t,\theta_{t+1}|\theta_{t-1})d\theta_{t+1}d\theta_t \\
&+\beta^2\int_{\theta_t,\theta_{t+1}}w_{t+2}'(W_{t+1}(w_1,(\hat{\theta}^{t-1},\theta_t);\theta_t),\theta_t;\theta_{t+1},\theta_{t+1})]\pi(\theta_t,\theta_{t+1}|\theta_{t-1})d\theta_{t+1}d\theta_t.
\end{aligned}
$$

By definition of f^* and w' this equals to

$$
\begin{aligned}
&\int_{\theta'}F_t^*(w_1,(\hat{\theta}^{t-1},\theta_t),\theta_t)\pi(\theta_t|\theta_{t-1})d\theta_t + \beta\int_{\theta_t,\theta_{t+1}}F_{t+1}^*(w_1,(\hat{\theta}^{t-1},\theta_t,\theta_{t+1});\theta_{t+1})\pi(\theta_t,\theta_{t+1}|\theta_{t-1})d\theta_{t+1}d\theta_t \\
&+\beta^2\int_{\theta_t,\theta_{t+1}}W_{t+2}(w_1,(\hat{\theta}^{t-1},\theta_t,\theta_{t+1}),\theta_{t+1})]\pi(\theta_t,\theta_{t+1}|\theta_{t-1})d\theta_{t+1}d\theta_t
\end{aligned}
$$

By repeated substitution of the promise keeping constraint (TRANSVERSALITY - TO BE COMPLETED) we have

$$W_t(w_1, \hat{\theta}^{t-1}, \theta_{t-1}) = \sum_{j \geq 0} \int_{\theta_t^{t+j}} \beta^j F_{t+j}^*(w_1, (\hat{\theta}^{t-1}, \theta_t^{t+j}), \theta_{t+j}) \Pi(\theta_t^{t+j} | \theta_{t-1}) d\theta^t \tag{A4}$$

Setting $t = 1$ and $\theta_{t-1} = \theta_0$ we have proven (A1). This is because the right hand side is equal to $u(\mathbf{C}, \mathbf{Y}, \theta_0)$ and the initial condition of the difference equation implies that $W_1(w_1, \varnothing; \theta_0) = w_1$.

i.2 (incentive compatibility) To show incentive compatibility, take any (w_1, θ^{t-1}) pair. It follows from equations (A3) and (A4) that

$$\sum_{j \geq 0} \int_{\theta_t^{t+j}} \beta^j F_{t+j}^*(w_1, (\hat{\theta}^{t-1}, \theta_t^{t+j}), \theta_{t+j}) \Pi(\theta_t^{t+j} | \theta_{t-1}) d\theta^t$$

$$= \int_{\theta_t} [F_t^*(w_1, (\hat{\theta}^{t-1}, \theta_t), \theta_t) + \beta W_{t+1}(w_1, (\hat{\theta}^{t-1}, \theta_t); \theta_t)] \pi(\theta_t | \theta_{t-1}) d\theta_t$$

$$= \int_{\theta_t} [f_t^*(W_t(w_1, \hat{\theta}^{t-1}, \theta_{t-1}), \theta_{t-1}; \theta_t, \theta_t)) + \beta w_{t+1}'(W_t(w_1, \hat{\theta}^{t-1}, \theta_{t-1}), \theta_{t-1}; \theta_t, \theta_t)] \pi(\theta_t | \theta_{t-1}) d\theta_t$$

Consider any function $\hat{\theta}_t(.) : \Theta \to \Theta$ that represents a reporting strategy in period t.[10] By the incentive compatibility of the component planner's problem,

$$\int_{\theta_t} [f_t^*(W_t(w_1, \hat{\theta}^{t-1}, \theta_{t-1}), \theta_{t-1}; \theta_t, \theta_t)) + \beta w_{t+1}'(W_t(w_1, \hat{\theta}^{t-1}, \theta_{t-1}), \theta_{t-1}; \theta_t, \theta_t)] \pi(\theta_t | \theta_{t-1}) d\theta_t$$

$$\geq \int_{\theta_t} [f_t^*(W_t(w_1, \hat{\theta}^{t-1}, \theta_{t-1}), \theta_{t-1}; \hat{\theta}_t(\theta_t), \theta_t)) + \beta w_{t+1}'(W_t(w_1, \hat{\theta}^{t-1}, \theta_{t-1}), \theta_{t-1}; \hat{\theta}_t(\theta_t), \theta_t)] \pi(\theta_t | \theta_{t-1}) d\theta_t$$

Right hand side can be equivalently written as

$$\int_{\theta_t} F_t^*(w_1, (\hat{\theta}^{t-1}, \hat{\theta}_t(\theta_t)), \theta_{t-1}); \theta_t) d\theta_t + \beta \int_{\theta_t} W_{t+1}(w_1, (\theta^{t-1}, \hat{\theta}_t(\theta_t)); \theta_t) \pi(\theta_t | \theta_{t-1}) d\theta_t \tag{A5}$$

We have shown that a one period deviation is not optimal. To show that an arbitrary deviation is suboptimal, we must apply the incentive compatibility of the component planner repeatedly. The

---

[10]The history up to period t is soppressed to simplify notation.

expression (A5) can be written as

$$\int_{\theta_t} [F_t^*(w_1, (\hat{\theta}^{t-1}, \hat{\theta}_t(\theta_t)), \theta_{t-1}); \theta_t) d\theta_t + \beta \int_{\theta_t, \theta_{t+1}} [f_{t+1}^*(W_{t+1}(w_1, (\theta^{t-1}, \hat{\theta}_t(\theta_t)); \theta_t), \theta_t; \theta_{t+1}, \theta_{t+1}) d\theta_{t+1} d\theta_t$$

$$+\beta^2 \int_{\theta_t, \theta_{t+1}} w_{t+2}'(W_{t+1}(w_1, (\theta^{t-1}, \hat{\theta}_t(\theta_t)); \theta_t), \theta_t; \theta_{t+1}, \theta_{t+1}))] \pi(\theta_t, \theta_{t+1} | \theta_{t-1}) d\theta_{t+1} d\theta_t$$

$$\geq \int_{\theta_t} [F_t^*(w_1, (\hat{\theta}^{t-1}, \hat{\theta}_t(\theta_t)), \theta_{t-1}); \theta_t) d\theta_t + \beta \int_{\theta_t, \theta_{t+1}} [f_{t+1}^*(W_{t+1}(w_1, (\theta^{t-1}, \hat{\theta}_t(\theta_t)); \theta_t), \theta_t; \hat{\theta}_{t+1}(\theta_t, \theta_{t+1}), \theta_{t+1}) d\theta_{t+1} d\theta$$

$$+\beta^2 \int_{\theta_t, \theta_{t+1}} w_{t+2}'(W_{t+1}(w_1, (\theta^{t-1}, \hat{\theta}_t(\theta_t)); \theta_t), \theta_t; \hat{\theta}_{t+1}(\theta_t, \theta_{t+1}), \theta_{t+1}))] \pi(\theta_t, \theta_{t+1} | \theta_{t-1}) d\theta_{t+1} d\theta_t$$

where $\hat{\theta}_{t+1}(.) : \Theta^2 \to \Theta$ is an arbitrary reporting strategy in period $t + 1$ and the inequality follows from the component planner's incentive compatibility. If we apply incentive compatibility repeatedly, we get that

$$\sum_{j \geq 0} \int_{\theta_t^{t+j}} \beta^j F_{t+j}^*(w_1, (\hat{\theta}^{t-1}, \theta_t^{t+j}), \theta_{t+j}) \Pi(\theta_t^{t+j} | \theta_{t-1}) d\theta^t$$

$$\geq \sum_{j \geq 0} \int_{\theta_t^{t+j}} \beta^j F_{t+j}^*(w_1, (\hat{\theta}^{t-1}, \hat{\theta}^{t+j}(\theta^{t+j})), \theta_{t+j}) \Pi(\theta_t^{t+j} | \theta_{t-1}) d\theta^t$$

for any reporting strategy $\hat{\theta}_t^{\infty} = \{\hat{\theta}_j\}_{j \geq t}$ .Setting $t = 1$, proves incentive compatibility of $\tilde{\mathbf{C}}^*, \tilde{\mathbf{Y}}^*$.

Thus, the allocation $\tilde{\mathbf{C}}^*, \tilde{\mathbf{Y}}^*$ satisfies all the constraints of the sequential problem. It is easy to show by recursive substitution of the V_t function that the objective functions are the same and therefore $\Omega(w_1, \theta_0) = \hat{V}_1(w_1, \theta_0)$

Proof of part ii). We must first define a candidate recursive allocation. CONVEXITY - LATER. Let F be defined as before. Define also

$$W_t(w_1, \hat{\theta}^{t-1}, \theta_{t-1}) = \sum_{j \geq t} \int_{\theta_t^{t+j}} \beta^{j-1} F_{t+j-1}^*(w_1, (\hat{\theta}^{t-1}, \theta_t^{t+j-1}), \theta_{t+j}) \Pi(\theta_t^{t+j} | \theta_{t-1}) d\theta^t$$

For any function $w(.) \in L^{\Theta}$ and $\theta_{t-1} \in \Theta$ construct a set

$$H_t(w(.), \theta_{t-1}) = \{w_1, \hat{\theta}^{t-1} : W_t(w_1, \hat{\theta}^{t-1}, \theta_{t-1}) = w(\theta_{t-1})\}$$

It is the set of all histories and initial utility entitlements such that the promised utility function is $w(.)$ and last period shock was θ_{t-1}. If the set $H_t(w(.), \theta_{t-1})$ is empty, then set $c_t(w(.), \theta_{t-1}, \hat{\theta}_t) =$

$U^{-1}[(1-\beta)w(.)]$ and $w'_{t+1}(w(.), \theta_{t-1}, \hat{\theta}, \theta_t) = w(.)$. If not, let

$$c_t(w(.), \theta_{t-1}, \hat{\theta}_t) = C^*_t(w_1, \hat{\theta}^t) \text{ for all } w_1, \hat{\theta}^{t-1} \in H_t(w(.), \theta_{t-1})$$

$$y_t(w(.), \theta_{t-1}, \hat{\theta}_t) = Y^*_t(w_1, \hat{\theta}^t) \text{ for all } w_1, \hat{\theta}^{t-1} \in H_t(w(.), \theta_{t-1})$$

$$w'_{t+1}(w(.), \theta_{t-1}, \hat{\theta}_t, \theta_t) = W_{t+1}(w_1, \hat{\theta}^t, \theta_t) \text{ for all } w_1, \hat{\theta}^{t-1} \in H_t(w(.), \theta_{t-1})$$

To simplify the algebra, recall the the definition of period utility used before implies that

$$f_t(w(.), \hat{\theta}_{t-1}, \hat{\theta}_t, \theta_t) = F^*_t(w_1, \hat{\theta}^t, \theta_t) \text{ for all } w_1, \hat{\theta}^{t-1} \in H_t(w(.), \theta_{t-1})$$

We need to show that the recursive allocation satisfies temporary incentive compatibility, promise keeping and that it belongs to set B^*.

ii.1(incentive compatibility) We first show that this recursive mechanism satisfies temporary incentive compatibility. Take any $w_1, \hat{\theta}^{t-1} \in H_t(w(.), \theta_{t-1})$. By definition, we have

$$f_t(w(.), \theta_{t-1}, \theta_t, \theta_t) + \beta w'_{t+1}(w(.), \theta_{t-1}; \theta_t, \theta_t) = F^*_t(w_1, (\hat{\theta}^{t-1}, \theta_t), \theta_t) + \beta W_{t+1}(w_1, (\hat{\theta}^{t-1}, \theta_t), \theta_t).$$

We expand the term W_{t+1} on the right hand side and apply incentive compatibility and the separability result of Lemma (6). Since the allocation is incentive compatible, any reporting strategy $\hat{\boldsymbol{\theta}}^\infty_t = \{\hat{\theta}_j\}_{j \geq t}$ is suboptimal. Consider in particular a one period deviation strategy where the agent reports $\hat{\theta}_t$ after history $(\hat{\theta}^{t-1}, \theta_t)$ and tells the truth otherwise. Incentive compatibility implies that

$$F^*_t(w_1, (\hat{\theta}^{t-1}, \theta_t), \theta_t) + \beta \sum_{j \geq 1} \int_{\theta^{t+j}_{t+1}} \beta^{j-1} F^*_{t+j}(w_1, (\hat{\theta}^{t-1}, \theta_t, \theta^{t+j}_{t+1}), \theta_{t+j}) \Pi(\theta^{t+j}_{t+1}|\theta_t) d\theta^t$$

$$\geq F^*_t(w_1, (\hat{\theta}^{t-1}, \hat{\theta}_t), \theta_t) + \beta \sum_{j \geq 1} \int_{\theta^{t+j}_{t+1}} \beta^{j-1} F^*_{t+j}(\hat{\theta}^{t-1}, \hat{\theta}_t, \theta^{t+j}_{t+1}), \theta_{t+j}) \Pi(\theta^{t+j}_{t+1}|\theta_t) d\theta^t$$

But by definition of W_{t+1}, f_t and w'_{t+1} the right hand side is equal to

$$F^*_t(w_1, (\hat{\theta}^{t-1}, \hat{\theta}_t), \theta_t) + \beta W_{t+1}(w_1, \hat{\theta}^{t-1}; \hat{\theta}_t, \theta_t) = f_t(w(.), \theta_{t-1}; \hat{\theta}_t, \theta_t) + \beta w'_{t+1}(w(.), \theta_{t-1}; \hat{\theta}_t, \theta_t)$$

which proves that the recursive allocation is incentive compatible. ∎

ii.2(promise keeping) Next step is to show that the promise keeping constraint holds. We have, for all $w_1, \hat{\theta}^{t-1} \in H_t(w(.), \theta_{t-1})$,

$$\int_{\theta_t \in \Theta} [f_t(w(.), \theta_{t-1}, \theta_t, \theta_t) + \beta w'_{t+1}(w(.), \theta_{t-1}, \theta_t, \theta_t)]\pi(\theta_t|\theta_{t-1})d\theta_t$$

$$= \int_{\theta_t \in \Theta} F_t^*(w_1, (\hat{\theta}^{t-1}, \theta_t), \theta_t) + \beta W_{t+1}(w_1, (\hat{\theta}^{t-1}, \theta_t), \theta_t)\pi(\theta_t|\theta_{t-1})d\theta_t$$

$$= \sum_{j \geq t} \int_{\theta_t^{t+j}} \beta^{j-1} F_{t+j-1}^*(w_1, (\hat{\theta}^{t-1}, \theta_t^{t+j-1}), \theta_{t+j})\Pi(\theta_t^{t+j}|\theta_{t-1})d\theta^t$$

$$= W_t(w_1, \hat{\theta}^{t-1}, \theta_{t-1})$$

$$= w(.)$$

The equality are repeated uses of definitions and the last equality uses the fact that $w_1, \hat{\theta}^{t-1} \in H_t(w(.), \theta_{t-1})$. Hence, our candidate recursive allocation satisfies all the constraints of the dynamic program. We must have $\hat{V}_1(w_1, \theta_0) \leq \Omega(w_1, \theta_0)$.

ii.3 (inclusion in B^*) The last step to show is that $w'_{t+1}(w(.), \hat{\theta}_{t-1}, \theta_t, .) \in B^*$ for all t, all $w(.)$. Fix T. Let $B_T = L^\Theta$. Define for $t = 1..T$

$$B_{t-1}^T = \{w \in L^\Theta : \exists c : \Theta \rightarrow \mathcal{R}_+, y : \Theta \rightarrow \mathcal{R}_+, w' : \Theta^2 \rightarrow \mathcal{R}$$

such that

$$w(\hat{\theta}_-) = \int_\theta [U(c(\theta), \frac{y(\theta)}{\theta}) + \beta w'(\theta, \theta)]\pi(\theta|\hat{\theta}_-)d\theta$$

$$U(c(\theta), 1 - \frac{y(\theta)}{\theta}) + \beta w'(\theta, \theta) \geq U(c(\hat{\theta}), \frac{y(\hat{\theta})}{\theta}) + \beta w'(\hat{\theta}, \theta) \qquad \forall \hat{\theta} \text{ s.t. } y(\hat{\theta}) \leq \theta, \text{ all } \theta$$

$$w'(\theta, .) \in B_t^T\}$$

It is easy to see that $w'_T(w(.), \hat{\theta}_{t-1}, \theta_t, .) \in B_{T-1}^T$ for all $w(.)$ For instance, $c_T(w(.), \theta_{T-1}, .)$, $y_T(w(.), \theta_{T-1}, .)$.and $w'_{T+1}(w(.), \theta_{T-1}, \theta_T, .)$ satisfies the conditions of the equation..An induction argument implies that $w'_{t+1}(w(.), \theta_{t-1}, \theta_t, .) \in B_t^T(\theta_t)$,all $w(.) \in L^\Theta$, all θ_t, all $t = 1...T$. If we show that $\lim_{T \rightarrow \infty} B_t^T = B^*$ for all t then the proof is complete. To show this, note that by construction we have $B_t^T = B_1^{T-t+1}$. Thus, it is enough to show that $\lim_{T \rightarrow \infty} B_1^T = B$. This follows simply from the fact that B^* is a fixed point of (5). This concludes this part of the proof.

Proof of part iii). This is a trivial implication of parts i) and ii). ■ ■

Proof of theorem (1). If $\mathbf{c}^*, \mathbf{y}^*, \mathbf{w}'^*$ solves the component planner's problem, then by Lemma (8) there is an allocation $\mathbf{C}^*, \mathbf{Y}^*$ satisfying $\hat{V}_1(w_1, \theta_0) = \Omega(w_1, \theta_0)$. It is easy to show

that the resource constraint (6) implies that the resource constraint (A2) in the cost minimization sequence problem holds. It follows from Lemma (7) that $w_1 = u(\mathbf{C}^*, \mathbf{Y}^*, \theta_0)$. The proof of the corollary follows. $\blacksquare$

Lemma 9 *The continuation utility function $w'(\hat{\theta}, \theta)$ is twice differentiable with respect to θ. Moreover,*

$$\left| \frac{\partial}{\partial \theta} w'(\hat{\theta}, \theta) \right| \leq \bar{\kappa}(\theta)$$

for some integrable function $\bar{\kappa}(\theta)$.

Proof. We start with the observation that, from the definition of the set B^*, this function has to satisfy, for any $\hat{\theta}$,

$$w'(\hat{\theta}, \theta) = \int_\theta [U(\tilde{c}(\theta'), \frac{\tilde{y}(\theta')}{\theta'}) + \beta \tilde{w}'(\theta', \theta')] \pi(\theta'|\theta) d\theta$$

for some functions $\tilde{c}$, $\tilde{y}$ and $\tilde{w}'$. (The argument of the proof goes through for any such functions, so we do not need to worry about their properties). Notice that the left hand side of the equation contains θ only as an argument in $\pi(\theta', \theta)$. We know that this function is twice continuously differentiable. Thus, $w'(\hat{\theta}, \theta)$ must also be twice continuously differentiable in θ, otherwise the promise keeping constraint can never be satisfied. Hence if $w'(\theta, .)$ is not twice differentiable then $w'(\theta, .) \notin B^*$. Thus. $w'(\hat{\theta}, \theta)$ will always be twice differentiable with respect to θ.

We have assumed that there exists an integrable function $b : \Theta \to \mathcal{R}$ such that $\left| \frac{\pi_2(\theta'|\theta)}{\pi(\theta'|\theta)} \right| \leq \kappa(\theta)$ for all θ' and almost all θ. Thus,

$$\begin{aligned}
\frac{\partial}{\partial \theta} w'(\hat{\theta}, \theta) &= \int_\theta [U(\tilde{c}(\theta'), \frac{\tilde{y}(\theta')}{\theta'}) + \beta \tilde{w}'(\theta', \theta')] \pi_2(\theta'|\theta) d\theta \\
&= \int_\theta [U(\tilde{c}(\theta'), \frac{\tilde{y}(\theta')}{\theta'}) + \beta \tilde{w}'(\theta', \theta')] \frac{\pi_2(\theta'|\theta)}{\pi(\theta'|\theta)} \pi(\theta'|\theta) d\theta \\
&\leq
\end{aligned}$$

By properties of $\frac{\pi_2(\theta'|\theta)}{\pi(\theta'|\theta)}$ we have

$$\left| \frac{\partial}{\partial \theta} w'(\hat{\theta}, \theta) \right| \leq \kappa(\theta) \left| \int_\theta [U(\tilde{c}(\theta'), \frac{\tilde{y}(\theta')}{\theta'}) + \beta \tilde{w}'(\theta', \theta')] \pi(\theta'|\theta) d\theta \right|$$

The second term is bounded by some constant $\tilde{\kappa}$ because the utility is bounded. Hence

$$\left| \frac{\partial}{\partial \theta} w'(\hat{\theta}, \theta) \right| \leq \bar{\kappa}(\theta)$$

where $\bar{\kappa}(\theta) = \kappa(\theta)\tilde{\kappa}$. ∎

Proof of Theorem (2). I first prove the envelope theorem first and then sufficiency and necessity of the condition i).

Let $X(\hat{\theta}, \theta) = U(c(\hat{\theta}), 1 - \frac{y(\hat{\theta})}{\theta}) + \beta w'(\hat{\theta}, \theta))$. Let $X(\theta, \theta) = X^*(\theta)$. We have $\frac{\partial}{\partial \theta} X(\hat{\theta}, \theta) = U_l \frac{y(\hat{\theta})}{\theta^2} + \frac{\partial}{\partial \theta} w'(\hat{\theta}, \theta)$. From Lemma 9, the derivative is correctly defined. The envelope theorem requires that, in addition, the term $X_1(\theta, \hat{\theta})$ satisfies Lipschitz condition $\sup_{\hat{\theta}} \left| U_l \frac{y(\hat{\theta})}{\theta^2} + \frac{\partial}{\partial \theta} w'(\hat{\theta}, \theta) \right| \leq \tilde{\kappa}(\theta)$ for some integrable $\tilde{\kappa}(\theta)$. By Lemma 9 again, the second term satisfies this property. The first one satisfies the boundedness condition as well (SHOW- STANDARD ARGUMENTS).

$$
\begin{aligned}
\left| X^*(\theta) - X^*(\theta') \right| &= \left| \sup_{\hat{\theta} \in \Theta} X(\hat{\theta}, \theta) - \sup_{\hat{\theta} \in \Theta} X(\hat{\theta}, \theta') \right| \\
&\leq \sup_{\hat{\theta} \in \Theta} \left| X(\hat{\theta}, \theta) - X(\hat{\theta}, \theta') \right| \\
&= \sup_{\hat{\theta} \in \Theta} \left| \int_{\theta}^{\theta'} \frac{\partial}{\partial \theta} X(\hat{\theta}, \varsigma) d\varsigma \right| \leq \int_{\theta}^{\theta'} \sup_{\hat{\theta} \in \Theta} \left| \frac{\partial}{\partial \theta} X(\hat{\theta}, \varsigma) \right| d\varsigma \\
&\leq \int_{\theta}^{\theta'} \tilde{\kappa}(\varsigma) d\varsigma
\end{aligned}
$$

Hence $X^*(\theta)$ is absolutely continuous and differentiable almost everywhere, with derivative $X_1(\theta, \hat{\theta})$. Then we have

$$
X(\theta, \theta) = X(0, 0) + \int_0^{\theta} \frac{\partial}{\partial \theta} X(\varsigma, \varsigma) d\varsigma
$$

Using $U_0 = X(0, 0)$ and the definition of X we conclude the proof.

I show second order conditions assuming differentiability of the policy functions. A standard argument shows that second order condition can be equivalently written as

$$
\frac{\partial^2}{\partial \theta \partial \hat{\theta}} X(\theta, \theta) \geq 0
$$

i.e.

$$
(\frac{\partial}{\partial \theta} U_c) \frac{dc}{d\hat{\theta}} + (\frac{\partial}{\partial \theta} U_l \frac{1}{\theta}) \frac{dy}{d\hat{\theta}} + \beta \frac{\partial^2}{\partial \theta \partial \hat{\theta}} w'(\theta, \theta) \geq 0
$$

by using the first order condition

$$
U_c \frac{dc}{d\hat{\theta}} + U_l \frac{1}{\theta} \frac{dy}{d\hat{\theta}} + \beta \frac{\partial}{\partial \hat{\theta}} w'(\theta, \theta) = 0
$$

to eliminate $\frac{dc}{d\theta}$ and rearranging we get

$$\left(\frac{\partial}{\partial\theta}\frac{U_l}{U_c}\frac{1}{\theta}\right)\frac{dy}{d\hat{\theta}} + \beta\left(\frac{\partial}{\partial\theta}\frac{\frac{\partial}{\partial\hat{\theta}}w'}{U_c}\right) \geq 0$$

The first term on the left hand side is positive by assumptions about utility function. However, the properties of the second term are not known.

We can, however, see that if we define $M(\hat{\theta}, \theta)$ as follows:

$$M(\hat{\theta}, \theta) = \frac{dy}{d\hat{\theta}}\frac{U_l}{U_c}\frac{1}{\theta} + \beta\frac{\frac{\partial}{\partial\hat{\theta}}w'}{U_c}$$

then $M(\hat{\theta}, \theta)$ must be increasing in θ. for $\hat{\theta} = \theta$.

Sufficiency of this condition can be shown as follows. Suppose it holds, but there is some $\hat{\theta}$ such that $X(\hat{\theta}, \theta) - X(\theta, \theta) > 0$. Therefore

$$\int_{\theta}^{\hat{\theta}} \frac{\partial}{\partial\hat{\theta}} X(\varsigma, \theta) d\varsigma > 0$$

which can be written as

$$\int_{\theta}^{\hat{\theta}} U_c\left[\frac{dc}{d\hat{\theta}}(\varsigma) + \frac{U_l}{U_c}\frac{1}{\theta}\frac{dy(\varsigma)}{d\hat{\theta}} + \beta\frac{\frac{\partial}{\partial\hat{\theta}}w'(\varsigma, \theta)}{U_c}\right] d\varsigma > 0$$

By using the term $M(\hat{\theta}, \theta)$ we have

$$\int_{\theta}^{\hat{\theta}} U_c\left[\frac{dc}{d\hat{\theta}}(\varsigma) + M(\varsigma, \theta)\right] d\varsigma > 0$$

and, since the term $M(\varsigma, \theta)$ is increasing in its second argument, the equation implies that

$$\int_{\theta}^{\hat{\theta}} U_c\left[\frac{dc}{d\hat{\theta}}(\varsigma) + M(\varsigma, \varsigma)\right] d\varsigma > 0.$$

But this contradicts the first order condition. ∎

Proof of Lemma (4). Denote the Lagrange multiplier on the resource constraint, threat keeping constraint and incentive compatibility constraint as Λ_w, Λ_g and $M(\theta)$ respectively. The first order conditions in U give us

$$fe^U = \theta(f\Lambda_w + f_\theta\Lambda_g + M) - \int_{\theta}^{\infty} M(\varepsilon)d\varepsilon$$

$$0 = \int_{0}^{\infty} M(\varepsilon)d\varepsilon$$

The first order condition in $W'(\theta)$ is

$$fV_w(W', G', \theta) = f\Lambda_w + f_\theta\Lambda_g + M$$

and we also have a limiting condition $\lim_{\theta\to 0} fV_w(W', G', \theta) = \int_0^\infty M(\varepsilon)d\varepsilon$. The first order condition in $G'(\theta)$ is

$$fV_g(W', G', \theta) = -\int_\theta^\infty M(\varepsilon)d\varepsilon$$

Finally, the envelope conditions for the problem are

$$V_w(w, g, \theta_-) = \Lambda_w$$

$$V_g(w, g, \theta_-) = \Lambda_g$$

Now make a guess that the policy functions satisfy

$$U(w, g, \theta_-, \theta) = a_u w + u(\tilde{g}, \theta_-, \theta)$$

$$W'(w, g, \theta_-, \theta) = a_w w + w'(\tilde{g}, \theta_-, \theta)$$

$$G'(w, g, \theta_-, \theta) = a_g w + g'(\tilde{g}, \theta_-, \theta)$$

$$\Lambda_w(w, g, \theta_-) = \lambda_w(\tilde{g}, \theta_-)e^{a_{\lambda_w} w}$$

$$\Lambda_g(w, g, \theta_-) = \lambda_g(\tilde{g}, \theta_-)e^{a_{\lambda_g} w}$$

$$M(w, g, \theta_-, \theta) = \mu(\tilde{g}, \theta_-, \theta)e^{a_\mu w}$$

and that the value function satisfies

$$V(w, g, \theta_-) = v(\tilde{g}, \theta_-)e^{aw}$$

We will verify this guess and determine the coefficients a of the policy functions and the value function and the value of rescaled promised marginal utility $\tilde{g}$. If these policy functions are valid, the coefficients must be such that all the terms involving w cancel out. This is because the equations must hold for all w.

Let $\gamma(\theta_-) = E(\theta|\theta_-)$. From the promise keeping constraint we have that the coefficients must satisfy $1 = \gamma(\theta_-)a_u + \beta a_w$. The threat keeping constraint implies that $g = a_u \gamma_{\theta_-}(\theta_-)w$ since $\int a_w f_{\theta_-} = 0$. The incentive compatibility constraint holds for $a_g = 0$ and for any a_u and a_w since the terms involving them cancel out. First order condition in U implies that we must have $a_u = a_{\lambda_w} = a_{\lambda_g} = a_\mu$. First order condition in W' will be identically satisfied if $a_w a = a_u$ while

first order condition in g' holds if $a_w a = a_\mu$. The envelope conditions hold if $a_{\lambda_w} = a$ and $a_{\lambda_g} = a$. Finally, we can verify the guess by checking that all the terms involving w also cancel in the Bellman equation itself. This restricts the coefficients to be such that $a = a_u = a.a_w$.

All these equalities are mutually consistent only if $a = a_u = a_{\lambda_w} = a_{\lambda_g} = a_\mu = \frac{1-\beta}{\gamma(\theta_-)}$ and that $a_w = 1$. The envelope condition also implies that $\frac{1-\beta}{\gamma(\theta_-)} v(\tilde{g}, \theta_-) = \lambda_w(\tilde{g}, \theta_-)$. The threat keeping constraint can now be written as

$$g - a_u \gamma_{\theta_-}(\theta_-) w = \int [\theta u(\theta) + \beta w'(\theta)] f_{\theta_-}(\theta|\theta_-) d\theta$$

and we therefore set $\hat{g} = g - (1-\beta) \frac{\frac{d}{d\theta_-}\gamma_{\theta_-}}{\gamma_{\theta_-}} w$. This completes the proof. ∎

Proof of Lemma (5). We need to verify that these guesses satisfy the Bellman equation and the constraints. For the sake of completeness, we also verify that the first order conditions satisfy the guess and show how to normalize the Lagrange Multipliers.

Note first that since θ is from lognormal distribution, its density satisfies $\pi(\theta|\theta_-) = \theta_-^{-\rho} \pi(\frac{\theta}{\theta_-^\rho}|1)$. In addition, its derivative π_{θ_-} satisfies $\pi_{\theta_-}(\theta|\theta_-) = \theta_-^{-(\rho+1)} \pi_{\theta_-}(\frac{\theta}{\theta_-^\rho}|1)$.

The promise keeping constraint is

$$
\begin{aligned}
0 &= \int [\theta u(g, \theta, \theta_-) + \beta w'(g, \theta, \theta_-)] \pi(\theta|\theta_-) d\theta \\
&= \theta_-^\rho \int [\frac{\theta}{\theta_-^\rho} u(\theta_-^{1-\rho} g, \frac{\theta}{\theta_-^\rho}, 1) + \beta w'(\theta_-^{1-\rho} g, \frac{\theta}{\theta_-^\rho}, 1)] \pi(\frac{\theta}{\theta_-^\rho}|1) \frac{d\theta}{\theta_-^\rho} \\
&= \theta_-^\rho \int [\varepsilon u(\theta_-^{1-\rho} g, \varepsilon, 1) + \beta w'(\theta_-^{1-\rho} g, \varepsilon, 1)] \pi(\varepsilon|1) d\varepsilon \\
&= 0
\end{aligned}
$$

where we have used substitution $\varepsilon = \frac{\theta}{\theta_-^\rho}$.

The threat keeping constraint is

$$
\begin{aligned}
g &= \int [\theta u(g, \theta, \theta_-) + \beta w'(g, \theta, \theta_-)] \pi_{\theta_-}(\theta|\theta_-) d\theta \\
&= \theta_-^\rho \int [\frac{\theta}{\theta_-^\rho} u(\theta_-^{1-\rho} g, \frac{\theta}{\theta_-^\rho}, 1) + \beta w'(\theta_-^{1-\rho} g, \frac{\theta}{\theta_-^\rho}, 1)] \pi_{\theta_-}(\frac{\theta}{\theta_-^\rho}|1) \frac{d\theta}{\theta_-^{\rho+1}} \\
&= \theta_-^{\rho-1} \int [\varepsilon u(\theta_-^{1-\rho} g, \varepsilon, 1) + \beta w'(\theta_-^{1-\rho} g, \varepsilon, 1)] \pi_{\theta_-}(\varepsilon|1) d\varepsilon \\
&= \theta_-^{\rho-1} \theta_-^{1-\rho} g \\
&= g
\end{aligned}
$$

so that threat keeping constraint is also satisfied.

For the incentive compatibility constraint we use the fact that $\varphi(\theta) = (1-\beta)\frac{\rho}{\theta}$:

$$\frac{\theta}{\theta_-^\rho} u(\theta_-^{1-\rho}g, \frac{\theta}{\theta_-^\rho}, 1) + \beta w'(\theta_-^{1-\rho}g, \frac{\theta}{\theta_-^\rho}, 1) = \theta_-^{-\rho}[\theta u(g, \theta, \theta_-) + \beta w'(g, \theta, \theta_-)]$$

$$= \theta_-^{-\rho}\{\int_0^\theta [u(g, \varepsilon, \theta_-) + \beta z(g, \varepsilon, \theta_-) + (1-\beta)\frac{\beta\rho}{\varepsilon}w'(g, \varepsilon, \theta_-)]d\varepsilon + \beta w_0'(g, \theta_-)\}$$

$$= \int_0^\theta [u(\theta_-^{1-\rho}g, \frac{\varepsilon}{\theta_-^\rho}, 1) + \beta z(\theta_-^{1-\rho}g, \frac{\varepsilon}{\theta_-^\rho}, 1) + (1-\beta)\frac{\beta\rho}{\varepsilon}\theta_-^\rho w'(\theta_-^{1-\rho}g, \frac{\varepsilon}{\theta_-^\rho}, 1)]\frac{d\varepsilon}{\theta_-^{-\rho}} + \beta w_0'(\theta_-^{1-\rho}g, \frac{\theta_-}{\theta_-^\rho})$$

$$= \int_0^{\frac{\theta}{\theta_-^\rho}} [u(\theta_-^{1-\rho}g, \tilde\varepsilon, 1) + \beta z(\theta_-^{1-\rho}g, \tilde\varepsilon, 1) + (1-\beta)\frac{\beta\rho}{\tilde\varepsilon}w'(\theta_-^{1-\rho}g, \tilde\varepsilon, 1)]d\tilde\varepsilon + \beta w_0'(\theta_-^{1-\rho}g, \frac{\theta_-}{\theta_-^\rho})$$

where the substitution $\tilde\varepsilon = \frac{\varepsilon}{\theta_-^{-\rho}}$ was used. Thus, the incentive compatibility constraint is satisfied as well.

The Bellman equation is

$$v(g, \theta_-) = \int [e^{u(g,\theta,\theta_-)} + \beta e^{\frac{1-\beta}{\gamma\theta_-}w'(g,\theta,\theta_-)} v(z(g,\theta,\theta_-), \theta)]\pi(\theta|\theta_-)d\theta$$

$$= \int [e^{u(\theta_-^{1-\rho}g, \frac{\theta}{\theta_-^\rho},1)} + \beta e^{\frac{1-\beta}{\theta_-^\rho\gamma_1}\theta_-^\rho w'(\theta_-^{1-\rho}g, \frac{\theta}{\theta_-^\rho},1)} v(z(\theta_-^{1-\rho}g, \frac{\varepsilon}{\theta_-^\rho}, 1), \theta)]\pi(\frac{\theta}{\theta_-^\rho}|1)\frac{d\theta}{\theta_-^\rho}$$

$$= \int [e^{u(\theta_-^{1-\rho}g,\varepsilon,1)} + \beta e^{\frac{1-\beta}{\gamma_1}w'(\theta_-^{1-\rho}g,\varepsilon,1)} v(z(\theta_-^{1-\rho}g, \varepsilon, 1), \theta)]\pi(\varepsilon|1)d\varepsilon$$

$$= v(\theta_-^{1-\rho}g, 1)$$

We guess that the Lagrange Multipliers satisfy $\lambda_w(g, \theta_-) = \lambda_w(\theta_-^{1-\rho}g, 1)\theta_-^{-\rho}$, $\lambda_g(g, \theta_-) = \lambda_g(\theta_-^{1-\rho}g, 1)\theta_-^{1-\rho}$ and $\mu(g, \theta, \theta_-) = \mu(\theta_-^{1-\rho}g, \frac{\theta}{\theta_-^\rho}, 1)\theta_-^{-2\rho}$. The first order condition for u is

$$\pi(\frac{\theta}{\theta_-^\rho}|1)e^{u(\theta_-^{1-\rho}g, \frac{\theta}{\theta_-^\rho},1)} = \theta_-^\rho \pi(\theta|\theta_-)e^{u(g,\theta,\theta_-)}$$

$$= \theta_-^\rho \{\theta[\lambda_w(g, \theta_-)\pi(\theta|\theta_-) + \lambda_g(g, \theta_-)\pi_{\theta_-}(\theta|\theta_-) + \mu(g, \theta, \theta_-)] - \int_\theta^\infty \mu(g, \varepsilon, \theta_-)d\varepsilon\}$$

$$= \frac{\theta}{\theta_-^\rho}[\lambda_g(\theta_-^{1-\rho}g, 1)\pi(\frac{\theta}{\theta_-^\rho}|1) + \lambda_g(\theta_-^{1-\rho}g, 1)\pi_{\theta_-}(\frac{\theta}{\theta_-^\rho}|1) + \mu(\theta_-^{1-\rho}g, \frac{\theta}{\theta_-^\rho}, 1)] - \frac{1}{\theta_-^\rho}\int_\theta^\infty \mu(\theta_-^{1-\rho}g, \frac{\varepsilon}{\theta_-^\rho}, 1)d\varepsilon$$

$$= \frac{\theta}{\theta_-^\rho}[\lambda_g(\theta_-^{1-\rho}g, 1)\pi(\frac{\theta}{\theta_-^\rho}|1) + \lambda_g(\theta_-^{1-\rho}g, 1)\pi_{\theta_-}(\frac{\theta}{\theta_-^\rho}|1) + \mu(\theta_-^{1-\rho}g, \frac{\theta}{\theta_-^\rho}, 1)] - \int_{\frac{\theta}{\theta_-^\rho}}^\infty \mu(\theta_-^{1-\rho}g, \hat\varepsilon, 1)d\hat\varepsilon$$

Similar calculations go for other first order conditions. ∎