The Review of Economic Studies Advance Access published January 3, 2013

Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach

MAREK KAPIČKA

University of California, Santa Barbara and CERGE-EI, Prague

First version received October 2008; final version accepted October 2012 (*Eds.*)

This paper studies efficient allocations in a dynamic private information economy with a continuum of idiosyncratic shocks that are persistent. I develop a first-order approach for this environment and show that the problem has a simple recursive structure that relies on only a small number of state variables, making the problem tractable. I find sufficient conditions that guarantee that the first-order approach is valid.

To illustrate the first-order approach I numerically compute the efficient allocations in a Mirrleesean economy with productivity shocks that follow a random walk and verify the validity of the first-order approach. I show that persistent shocks create a new trade-off where the social planner decreases the informational rent of the agent today at the cost of providing higher insurance in the future.

1. INTRODUCTION

This paper studies efficient allocations in a dynamic economy with a continuum of idiosyncratic private information shocks that follow a Markov process. It is well known that persistent shocks significantly increase the complexity of the problem. As shown by Fernandes and Phelan (2000), the efficient allocations have a recursive structure, but the dimensionality of the state space is proportional to the number of possible shock values. Quantitative analysis thus is feasible only for a small number of shocks. This paper shows that when the first-order approach is used, the state space can be reduced to a manageable dimension of two endogenous state variables even when the shocks take a continuum of values. This greatly increases the tractability of the problem and the ability to investigate the efficient allocations both qualitatively and quantitatively. While I use the first order approach to study the efficient allocations in a Mirrleesean private information economy, the methodology developed in this paper has much wider applicability. It can be used to answer other important questions where both private information or persistence of shocks are essential, like the analysis of firm financing constraints or the analysis of optimal health insurance.

When the shocks are i.i.d., the recursive formulation of the efficient contract is known to take a very simple form.[1] The agent's reporting history up to any period t can be summarized by a single statistic, called promised utility, which is the lifetime utility the agent is entitled to receive from period t onwards. The promised utility is sufficient to summarize agent's past because all agents have identical preferences over the allocations from the current period onwards, regardless of whether they have reported truthfully in previous periods or not. This is no longer true when shocks are persistent. The probability distribution, and hence preferences, over the allocations now depend on the previous shock. Thus, to ensure incentive compatibility, the continuation utility must

1. See Green (1987), Thomas and Worrall (1990) or Atkeson and Lucas (1992).

depend on both the current report and on the current shock. Since the social planner does not observe the shock, the continuation utility must be chosen for all possible shock values. With a continuum of shocks, the state variable must therefore be a *function* that specifies the agent's promised utility for each possible shock the agent might have had in the previous period.

The advantage of the first-order approach in static economies is that it restricts attention to local deviations from the optimum (Mirrlees (1971)). This paper develops a dynamic version of the first-order approach. I provide necessary conditions for incentive compatibility using the envelope condition and show that it only requires knowing how the continuation utility varies with the agent's previous period shock *on the margin*. The marginal continuation utility corresponds to the future informational rent of the agent.

Replacing the incentive compatibility constraint with the envelope condition leads to a relaxed problem. The relaxed problem is shown to have a simple recursive structure in the sense that the set of all admissible lifetime utility and lifetime marginal utility pairs, $\mathcal{V}^R$, can be factorized into a current allocation and a pair of continuation utilities and marginal continuation utilities drawn from the same set $\mathcal{V}^R$. As a consequence, the social planner's problem can be written recursively and its state space can be reduced from to two numbers, instead of a function. I provide conditions for the validity of the first-order approach and show that, if those conditions are satisfied, the recursive relaxed problem solves the original unrelaxed problem.

I apply the first-order approach in a simple Mirrleesean economy with endogenous consumption and labor supply and privately observed productivity shocks that follow a random walk and are Pareto distributed. I find that the intratemporal wedge tends to decrease with the marginal promised utility. This is consistent with the mechanics of the efficient contract: for lower marginal promised utilities the informational rent and labor supply decreases, and lower labor supply is associated with higher intratemporal wedges. Quantitatively, the initial intratemporal wedge is small, equal to only about 2.5%. The expected intertemporal wedge is hump shaped, and is between 1-2%. I numerically verify the validity of the relaxed problem.[2]

The examples highlight a new type of trade-off that the social planner faces when the shocks are persistent. The trade-off involves two opposing effects in which the marginal continuation utility affects the social planner's costs. On one hand, a higher marginal continuation utility decreases the informational rent of the agent and makes it easier to provide incentives in the current period. On the other hand, a higher marginal continuation utility distorts the allocation in the next period by forcing the social planner to choose an allocation that is less sensitive to the current shock. In other words, the social planner is forced to provide insurance that is, from the ex-post perspective, too high. In the optimum, the social planner optimally balances those two effects.

1.1. *Related Literature*

The first-order approach has been so far used mainly in static environments. Dynamics extensions include Courty and Li (2000), who study a two period price discrimination problem, Eso and Szentes (2007), who study an optimal auction design, and especially Pavan et al. (2009). Eso and Szentes (2007) characterize the efficient allocations in a

2. In the online Appendix B I also study an economy with linear-quadratic utility and uniformly distributed shocks. I provide a closed form solution for the relaxed problem. I also *analytically* verify that the relaxed problem is valid for all degrees of shock persistence.

two period model with private information being revealed gradually over time. They are able to prove the validity of the first order approach under a simple assumption that the transformed payoff function has a bounded derivative. Pavan et al. (2009) have recently extended the first-order approach to dynamic environments with a more general stochastic structure. There are two main differences between their paper and this one. First, Pavan et al. (2009) do not characterize the recursive structure of the efficient allocations, which is the main goal of this paper. Second, they show the validity of the first-order approach under slightly different conditions that rely on the Lebesque Dominated Convergence Theorem, while this paper uses the Monotone Convergence Theorem to prove the applicability of the first-order approach. While the difference seems subtle, using the Monotone Convergence Theorem allows one to apply the first-order approach to some major parametric classes of distributions, for instance the lognormal distribution.

The literature on recursive dynamic contracting with persistent private information has mainly focused on an environment with a finite number of shocks, starting with Fernandes and Phelan (2000). Doepke and Townsend (2006) extend the results for economies with both private information and hidden action. Battaglini and Coate (2008) study efficient income taxation in a Mirrleesean economy with hidden productivity, while Tchistyi (2006) studies the optimal security design with hidden cash flows. Both papers exploit a very special environment where Markov shocks can take only two values and the agents are risk neutral. There is also a growing literature that studies private information economies with persistent shocks in continuous time. Zhang (2009) characterizes the efficient allocations in a Mirrleesean economy with two shocks. His quantitative findings are similar to my findings, namely that both intertemporal and intratemporal wedges are larger than in an economy with i.i.d. shocks. Williams (2011) studies an economy with hidden income that follows a Brownian motion. As he points out, the trade-offs in continuous time models are very different from discrete time models. In particular, current consumption has measure zero in a continuous time setting, and so all incentives are provided through the variations in the continuation utility. Williams (2011) also assumes that the agent cannot overstate her true shock, and it is not obvious how important this restriction is for the validity of his first-order approach. Nevertheless, his recursive formulation is similar in a sense that one needs to introduce an additional costate variable, corresponding to the marginal continuation utility in this paper.

Farhi and Werning (2010) and Golosov et al. (2010) both use the first-order approach developed in this paper to study dynamic Mirrleesean economies with a finite horizon. Both papers consider utility functions with a constant Frisch elasticity of labor supply. Golosov et al. (2010) assume in addition that there are no income effects on labor supply. Both papers differ in their assumptions about the distribution of the shocks: in Farhi and Werning (2010) the shocks are lognormally distributed and follow a random walk, while in Golosov et al. (2010) the shock distribution is calibrated from the data. In contrast to both papers, the Mirrleesean economy studied in this paper assumes infinite horizon and Pareto distributed shocks.

Finally, the first-order approach is technically related to the dynamic Ramsey problem with commitment (Kydland and Prescott (1980)) or without commitment in a strategic framework (Phelan and Stacchetti (1999), see also Chang (1998)).[3] The

3. Similar techniques have also been used by Werning (2001) and Abraham and Pavoni (2008) to study dynamic hidden savings problem and by Jarque (2010) to analyze moral hazard environment with effort persistence.

dynamic Ramsey problem can be characterized recursively by introducing an additional costate variable that plays a role similar to the role played by the marginal promised utility here. In both cases, their role is to rule out marginal deviations from the agent's optimum (deviations in savings in the Ramsey problem and deviations in the report in the dynamic private information problem). Kydland and Prescott (1980) see the introduction of the additional costate variable as a restriction on the planner that overcomes the time consistency problem and ensures that the solution is optimal. The same logic applies in the dynamic private information problem with persistent shocks: the solution is not time consistent, and the planner would, if allowed, like to reoptimize at the beginning of each period (except for the initial one). Adding the marginal promised utility as a state variable to the planning problem enforces the efficient, but time inconsistent, solution.

2. THE MODEL

The economy is populated by a continuum of infinitely lived agents of measure one. Each period the agents care about consumption $x_t \in \mathbb{X}$. They receive an idiosyncratic shock $\theta_t \in \Theta \subseteq \mathbb{R}_{++}$, where Θ is an open interval, its infimum is $\underline{\theta}$ and its supremum is $\overline{\theta}$. Agents' preferences are represented by an expected utility function

$$\mathbb{E}_0 \sum_{t=1}^{\infty} \beta^t U(x_t, \theta_t) \qquad 0 < \beta < 1.$$

The following is assumed:

Assumption 1. *The consumption set $\mathbb{X}$ is convex and compact.*

Assumption 2. *The utility function $U : \mathbb{X} \times \Theta \to \mathbb{R}$ is continuous, strictly increasing and concave in x.*

Assumption 1 could be substantially relaxed, and is made here mainly to reduce the mathematical complexity of the problem. Both assumptions together imply that $|U|$ is bounded by $\overline{U}_0 + \theta \overline{U}_1$ for some finite constants $\overline{U}_0$ and $\overline{U}_1$. The utility function is allowed to be unbounded in the shock. This is important because a common specification is that the utility function is affine in θ, and many important parametric distributions have unbounded support.

The shock θ_t follows a first-order Markov process. Its transition cumulative distribution function is given by $F : \Theta^2 \to [0,1]$, such that $F(\cdot|\theta_-)$ is, for all $\theta_- \in \Theta$, a cumulative distribution function, and $F(\theta|\cdot)$ is, for all $\theta \in \Theta$, a Borel measurable function. Let, for any $1 \leq j < t$, $\theta_j^t = (\theta_j, \ldots, \theta_t) \in \Theta^{t-j+1}$ denote a partial history of shocks in periods j through t. Given the transition cumulative distribution function F and any $\theta_{j-1} \in \Theta$, let $\mu_j^t(\theta_j^t|\theta_{j-1}) : \Theta^{t-j+2} \to [0,1]$ be the cumulative distribution of the partial histories $\theta_j^t \in \Theta^{1+t-j}$.[4] The distribution of shocks is such that the present value of shocks is always finite:

Assumption 3. *The conditional expectation $\mathbb{E}_0(\theta'|\theta)$ is bounded by $B_0 + B_1\theta$ for some finite constants B_0 and $B_1 < \beta^{-1}$.*

4. For $j = 1$ a simplified notation $\mu^t = \mu_1^t$ will be used.

The social planner is committed to deliver a lifetime utility $w_1 \in \mathcal{W}(\theta_0)$ and his objective is to minimize the present value of costs. The social planner has access to credit markets and can freely borrow or save. The intertemporal price of consumption is exogenously given by $q < 1$. Each period, the costs of allocating $x \in \mathbb{X}$ to the agent are given by a function $R : \mathbb{X} \to \mathbb{R}$ that satisfies the following:

Assumption 4. *The cost function R is continuous, increasing and convex.*

The timing and information structure are as follows. At the beginning of each period, agents observe their current shock. After that, a report is sent to the social planner, and current consumption is determined. While consumption is observed by the social planner, the shock of the agent is her private information. The only exception is the initial shock, θ_0, which is observed by the social planner and is the same for everyone.

3. CONSTRAINED PARETO OPTIMA

At time zero, the social planner selects an *allocation* $X = \{X_t\}_{t=1}^{\infty}$, which is a sequence of measurable functions $X_t : \Theta^t \to \mathbb{X}$, $t \geq 1$. Let $\mathcal{X}$ be the set of all allocations. The agent's preferences over an allocation X are given by

$$W(X;\theta_0) = \sum_{t=1}^{\infty} \beta^{t-1} \int_{\Theta^t} U\left(X_t(\theta^t), \theta_t\right)\, d\mu^t(\theta^t|\theta_0).$$

At the beginning of each period, the agents report their current type to the social planner. The *reporting strategy* $\sigma = \{\sigma_t\}_{t=1}^{\infty}$ is a sequence of measurable functions $\sigma_t : \Theta^t \to \Theta$. The set of all reporting strategies is denoted by Σ, and the history of reports up to period t is denoted by $\hat{\theta}^t \in \Theta^t$. Since the shocks are private information of the agent, the allocation must be such that the agent always prefers to report her shock truthfully. If the agent chooses reporting strategy $\sigma \in \Sigma$ she receives $(X \circ \sigma)(\theta^t) = \{X_t(\sigma^t(\theta^t)\}_{t\geq 1}$. Thus, the allocation is *incentive compatible* if

$$W(X;\theta_0) \geq W(X \circ \sigma;\theta_0) \quad \forall \sigma \in \Sigma. \tag{3.1}$$

3.1. *Temporary Incentive Compatibility Constraint*

It will now be shown that the incentive compatibility constraint (3.1) is essentially equivalent to a temporary incentive compatibility constraint that explicitly only rules out one period deviations. Let $X_{t+1}^{\infty}(\theta^t) \in \mathcal{X}$ be the continuation of an allocation $X \in \mathcal{X}$ from period $t+1$ on, given a history θ^t. Define an allocation X to be *temporarily incentive compatible* after a history of shocks θ^t if

$$U(X_t(\theta^t),\theta_t) + \beta W(X_{t+1}^{\infty}(\theta^t);\theta_t) \geq U(X_t(\theta^{t-1},\hat{\theta}),\theta_t) + \beta W(X_{t+1}^{\infty}(\theta^{t-1},\hat{\theta});\theta_t). \tag{3.2}$$

for all reports $\hat{\theta} \in \Theta$. The relationship between incentive compatibility and temporary incentive compatibility is given in the next Lemma.[5]

Lemma 1.

5. The proof is similar to the proof of Theorem 2.1 of Fernandes and Phelan (2000) and is in the online Appendix C.

i. Suppose that an allocation X is incentive compatible. Then it is temporarily incentive compatible for all $t \geq 1$ and for almost all $\theta^t \in \Theta^t$.

ii. Suppose that an allocation X is temporarily incentive compatible for all $t \geq 1$, for all $\theta^{t-1} \in \Theta^{t-1}$ and for almost all $\theta_t \in \Theta$. Then it is incentive compatible.

Imposing the temporary incentive compatibility constraint for all past histories θ^{t-1} is thus a stronger requirement than imposing the incentive compatibility constraint. The asymmetry disappears if the shocks are discrete and each of them has strictly positive probability, see Theorem 2.1 of Fernandes and Phelan (2000). Since the differences between the incentive compatibility constraint and the temporary incentive compatibility constraint occur only on a set of measure zero, I will restrict attention to allocations that satisfy the temporary incentive compatibility constraint (3.2) after *all* histories.

3.2. *The Set of Implementable Utilities*

Allocations that satisfy the temporary incentive compatibility constraint (3.2) for all histories generate a set of lifetime utilities $\mathcal{W}(\theta_0)$:

$$\mathcal{W}(\theta_0) = \{W(X;\theta_0) \,|\, X \text{ satisfies (3.2)}\}.$$

Assumptions 1-3 imply that $|\mathcal{W}(\theta_0)|$ is bounded by $K_0 + K_1\theta_0$, where K_0 and K_1 are finite constants. Let $\Gamma^{IC} : \mathcal{W} \twoheadrightarrow \mathcal{X}$ be the set of all allocations that satisfy the temporary incentive compatibility constraint (3.2) and deliver a lifetime utility of w_1 to an agent with initial shock θ_0:

$$\Gamma^{IC}(w_1,\theta_0) = \{X \in \mathcal{X} \,|\, W(X;\theta_0) = w_1\}. \tag{3.3}$$

It is desirable that the set $\Gamma^{IC}(w_1,\theta_0)$ be convex. It is easy to show that $\Gamma^{IC}(w_1,\theta_0)$ is convex in X if the following assumption holds:

Assumption 5. *The utility function U is affine in x.*

Assumption 5 is relatively innocuous for two reasons. First, if the utility function is separable and affine in θ, that is, if there are functions U^1 and U^2 such that $U(x,\theta) = U^1(x^1) + \theta U^2(x^2)$ for $x = (x^1,x^2) \in \mathbb{X}$, then one can transform the utility and the costs by redefining the choice variable to be $\left(U^1(x^1), U^2(x^2)\right)$ directly. The utility is then affine.[6] Second, even if the utility is not separable and affine in θ one can convexify and linearize the problem by introducing lotteries over consumption.[7] In the remainder of the paper, Assumption 5 is assumed to hold, although to reduce notation, I will continue using the general expression for the utility function.

3.3. *Social Planner's Problem*

The present value of the costs that are implied by an allocation X is given by

$$D(X;\theta_0) = \sum_{t=1}^{\infty} q^{t-1} \int_{\Theta^t} R\left(X_t(\theta^t)\right)\, d\mu^t(\theta^t|\theta_0).$$

6. Note that a standard taste shock utility with $U(x,\theta) = \theta U^2(x)$, as well as a utility function with consumption and leisure with $x = (c,-l)$ and $U(x,\theta) = U^1(c) + \theta U^2(l)$ are both separable and affine in θ.

7. Online Appendix A formally shows how to transform the problem in both cases, and proves that the transformed problem satisfies Assumptions 2, 4 and 5 with the choice variable being $\left(U^1(x^1), U^2(x^2)\right)$ in the first case and a probability distribution over $\mathbb{X}$ in the second case.

The social planner's objective is to select an allocation that minimizes the present value of costs among all the incentive compatible allocations that give the agents a lifetime utility $w_1 \in \mathcal{W}(\theta_0)$. The cost function $P^{IC} : \mathcal{W} \to \mathbb{R}$ satisfies

$$P^{IC}(w_1, \theta_0) = \min_{X \in \mathcal{X}} \left\{ D(X; \theta_0) \,|\, X \in \Gamma^{IC}(w_1, \theta_0) \right\}. \tag{3.4}$$

An allocation X^* is efficient if it attains the minimum of (3.4). Assumption 1 implies that the correspondence Γ^{IC} is compact valued. It is also continuous in w_1. By Assumption (2) the objective function is continuous in X. Hence, by Berge's Theorem of Maximum,[8] a solution exists and P^{IC} is continuous in w_1.

4. FERNANDES-PHELAN RECURSIVE FORMULATION

To derive a recursive formulation of the social planner's problem, it is necessary to show that its state space has a recursive representation. Unfortunately, the set of lifetime utilities $\mathcal{W}(\theta_0)$ does not have a recursive structure. To see why, note that one needs to know the value of $W(X_{t+1}^{\infty}(\theta^{t-1}, \hat{\theta}); \theta_t)$ for each $\hat{\theta} \in \Theta$ and each $\theta \in \Theta$ to determine whether an allocation satisfies the temporary incentive compatibility constraint (3.2). Specifically, one needs to know not only the *equilibrium* lifetime utility $W(X_{t+1}^{\infty}(\theta^t); \theta_t)$, but also the *off-equilibrium* lifetime utility $W(X_{t+1}^{\infty}(\theta^{t-1}, \hat{\theta}); \theta_t)$ for $\hat{\theta} \neq \theta_t$. The set $\mathcal{W}(\theta_0)$ contains only the equilibrium lifetime utilities, and so is too small to have a recursive structure. To put it differently, knowing that $W(X_{t+1}^{\infty}(\theta^t); \theta_t) \in \mathcal{W}(\theta_t)$ is not enough to determine whether $W(X_t^{\infty}(\theta^{t-1}); \theta_{t-1}) \in \mathcal{W}(\theta_{t-1})$.

Fernandes and Phelan (2000) propose a solution by showing[9] that the set of all equilibrium and off-equilibrium lifetime utilities

$$\mathcal{V}^{IC} = \{W(X; \cdot) \,|\, X \text{ satisfies } (3.2)\}$$

possesses a recursive representation. That is, any function $\mathbf{w} : \Theta \to \mathcal{W}$ that belongs to$\mathcal{V}^{IC}$ can be, using the language of Abreu et al. (1990), factorized into the current allocation and a continuation equilibrium and off-equilibrium lifetime utility function chosen again from the set $\mathcal{V}^{IC}$. The social planner's problem can then be shown to have a recursive representation as well, and its state space includes $\mathcal{V}^{IC}$.

While the Fernandes-Phelan recursive formulation works well in theory, it is clear that having a function $\mathbf{w} : \Theta \to \mathcal{W}$ as one of the state variables makes the social planner's problem, as well as the computation of the set $\mathcal{V}^{IC}$, prohibitively complex. In addition, it is hard to see how to provide insights into how the efficient allocations work if the problem depends on the dynamics of a function.

To obtain a simpler recursive structure, one first needs to provide a simpler characterization of the temporary incentive compatibility constraint. The approach taken in this paper is to use the first-order approach. The first-order approach uses an envelope theorem to replace the temporary incentive compatibility constraint (3.2) by an envelope condition. It will be shown that the envelope condition involves only the equilibrium continuation utility $W(X_{t+1}^{\infty}(\theta^{t-1}, \theta_t); \theta_t)$ and its partial derivative with respect to the current shock θ_t.

8. See for example Aliprantis and Border (2005), Theorem 17.31.
9. Fernandes and Phelan (2000) show the result for the case of finite number of shock values. The generalization for a continuum of shocks is straightforward.

Naturally, the first-order approach relaxes the constraints on the social planner because the envelope condition is sufficient but not necessary. I therefore provide sufficient second-order conditions that can be checked to determine whether the first-order approach is valid.

5. FIRST-ORDER APPROACH

In order to apply the first-order approach, it is necessary to ensure that the function $W(X, \theta)$ is differentiable with respect to θ and that the derivative is well behaved. To this end, it will be assumed that the transition function is continuously differentiable:

Assumption 6. *$F(\theta'|\theta)$ is differentiable in θ.*

Denote the derivative of F by $F_2(\theta'|\theta)$. It will also be assumed that

Assumption 7. *$F_2(\theta'|\theta) \leq 0$ and $F(\theta'|\theta)$ is either concave or convex in θ on Θ.*

That is, F_2 is required to be nonpositive, which means that the shocks are positively autocorrelated. It is also required to be monotone, either decreasing or increasing, but is not allowed to be strictly increasing on some part of the domain, and strictly decreasing on the rest of the domain. Under these conditions, the following result holds:

Lemma 2. *The function $W(X;\theta)$ is differentiable in θ with its derivative*

$$W_\theta(X;\theta) = \int_\Theta \left[U\left(X_1(\varepsilon),\varepsilon\right) + \beta W(X_2^\infty(\varepsilon);\varepsilon)\right] dF_2(\varepsilon|\theta). \tag{5.5}$$

Assumption 6 is clearly required by Lemma 2. On the other hand, Assumption 7 can be avoided. It is imposed in order to apply the Monotone Convergence Theorem and could be replaced by a condition that $|F_2|$ has some integrable bound.[10] However, some prominent parametric classes of distributions, for instance the Lognormal distribution, do not satisfy the integrable bound assumption, but they satisfy Assumption 7 (see Example 1 below).

Note also that the assumption that F is differentiable in θ may sometimes be too restrictive. For instance, if F does not have a full support then, in most cases, F will fail to be differentiable everywhere on Θ.[11] However, Lemma 2 can be extended to allow for moving support, provided that the support is well behaved. In particular, if $F(\theta'|\theta)$ has a density $f(\theta'|\theta)$ and satisfies the assumptions of Lemma 2 only on the interior of $[a(\theta), b(\theta)]$ and both a and b are differentiable, then one can show that the derivative of

10. As in Pavan et al. (2009), second part of Assumption 5. One would then apply the Lebesque Dominated Convergence Theorem to prove the result. Note also that Assumptions 1 and 2 and 5 (first part) of Pavan et al. (2009) are satisfied as well, while Assumptions 3,4 and 6 are not needed for this result.

11. For example, suppose that $F(\theta'|\theta)$ is uniformly distributed on $[\theta-\frac{1}{2}, \theta+\frac{1}{2}]$. Then the derivative $F_2(\theta'|\theta)$ does not exist at $\theta-\frac{1}{2}$ and $\theta+\frac{1}{2}$.

W at the interior of the support is given by

$$\begin{aligned} W_\theta(X;\theta) &= b'(\theta)\left[U\left(X_1(b(\theta)), b(\theta)\right) + \beta W(X_2^\infty; b(\theta))\right] f(b(\theta))|\theta) \\ &\quad - a'(\theta)\left[U\left(X_1(a(\theta)), a(\theta)\right) + \beta W(X_2^\infty; a(\theta))\right] f(a(\theta))|\theta) \\ &\quad + \int_{a(\theta)}^{b(\theta)} \left[U\left(X_1(\varepsilon), \varepsilon\right) + \beta W(X_2^\infty(\varepsilon); \varepsilon)\right] f_2(\varepsilon|\theta)\, d\varepsilon. \end{aligned} \tag{5.6}$$

In the theoretical analysis that follows, this case will not be explicitly studied. This is done only for algebraic convenience, and the results can be extended to the more general case. An example with moving support will be studied in the quantitative exercise in Section 7.

To provide a sharper characterization of the function $W(X,\theta)$, the following restriction is imposed:

Assumption 8. *$\mathbb{E}(\theta'|\theta)$ is Lipschitz continuous in θ on $[\eta, \overline{\theta})$ for any $\eta \in \Theta$.*

The restrictions imposed by Assumption 8 are weaker than Lipschitz continuity in θ on Θ. This is an important generalization, because for many standard parametric distributions $\Theta = (0,\infty)$ and $\mathbb{E}(\theta'|\theta) = \theta^\rho$, which is not Lipschitz continuous on Θ, but satisfies Assumption 8. Next Lemma shows that this property is inherited by W:

Lemma 3. *$W(X,\theta)$ is Lipschitz continuous in θ on $[\eta, \overline{\theta})$ for any $\eta \in \Theta$.*

Lemma (3) implies that for any $\theta \in \Theta$, the marginal continuation utility $W_\theta(X,\theta)$ is bounded.

5.1. *Necessary and Sufficient Conditions*

Necessary conditions for an allocation to be temporarily incentive compatible are derived next using the envelope theorem. It will be possible to apply the envelope theorem because of Lemmas (2) and (3), as well as because of the following assumption:[12]

Assumption 9. *U is differentiable and Lipschitz continuous in θ.*

Denote the derivative of the utility function by U_θ. Let also $\underline{W}(X_t^\infty(\theta^{t-1})) = \lim_{\theta\to\underline{\theta}} U(X_t(\theta^{t-1},\theta),\theta) + \beta W(X_{t+1}^\infty(\theta^{t-1},\theta);\theta)$. Then the following result holds:

Theorem 4. *Suppose that an allocation X satisfies the temporary incentive compatibility constraint (3.2). Then for any $t \geq 1$ and any $\theta^t \in \Theta^t$,*

$$\begin{aligned} U(X_t(\theta^t),\theta_t) + \beta W(X_{t+1}^\infty(\theta^t);\theta_t) &= \int_{\underline{\theta}}^{\theta_t} \left[U_\theta\left(X_t(\theta^{t-1},\varepsilon),\varepsilon\right) + \beta W_\theta(X_{t+1}^\infty(\theta^{t-1},\varepsilon);\varepsilon)\right] d\varepsilon \\ &\quad + \underline{W}(X_t^\infty(\theta^{t-1})). \end{aligned} \tag{5.7}$$

The theorem is proven by verifying the conditions of Theorem 2 in Milgrom and Segal (2002). The envelope condition (5.7) shows how the lifetime utility must vary with the current period shock in order to be incentive compatible. The expression under the

12. Pavan et al. (2009) use a similar assumption, see their Assumption 4.

integral consists of two terms. The first term is the current informational rent: the extra current period utility an agent receives from a marginal increase in current skills. The second term corresponds to the future informational rent: extra lifetime utility from tomorrow on that an agent receives from a marginal shift in the future distribution of shocks. The second term is zero if shocks are i.i.d. because the current shock bears no information about future shocks.

Condition (5.7) is necessary but not sufficient for an allocation to be temporarily incentive compatible. A sufficient condition is given next:

Theorem 5. *Suppose that an allocation X satisfies (5.7) and, in addition,*

$$U_\theta\left(X_t(\theta^{t-1},\hat{\theta}_t),\theta_t\right)+\beta W_\theta\left(X_{t+1}^{\infty}(\theta^{t-1},\hat{\theta}_t);\theta_t\right) \tag{5.8}$$

is increasing in $\hat{\theta}_t$ for all $t\geq 1$, all $\theta^{t-1}\in\Theta^{t-1}$ and almost all $\theta_t\in\Theta$. Then X is temporarily incentive compatible.

If the shocks are i.i.d. and U satisfies the Spence-Mirrlees condition, the second term in the expression (5.8) drops out and Theorem 5 is equivalent to a simple requirement that $X_t(\theta^t)$ is increasing in θ_t, and one can show that this requirement is necessary as well. When the shocks are persistent, monotonicity of X_t is clearly neither required nor implied. Note that condition (5) is identical to condition ii. of Proposition 3 in Pavan et al. (2009), although the proof is somewhat different.[13]

5.2. *Some Parametric Examples*

Since the applicability of the first-order approach is essential for the recursive formulation developed in the next section, three major parametric distributions are now studied in detail to verify that they satisfy its assumptions. The first two examples are such that F has full support, while the last one involve a moving support.

Example 1 (Lognormal distribution). *Suppose that $\sigma>0$, $\rho\in[0,1]$ and $\Theta=(0,\infty)$. Define $F(\theta'|\theta)=\Phi(\frac{\ln\theta'-\rho\ln\theta}{\sigma^2}-\frac{1}{2})$ where Φ is cdf of a standard normal distribution. The distribution clearly satisfies Assumption 3. Assumption 6 holds as well since F is differentiable in its second argument on Θ with $F_2(\theta'|\theta)=-\frac{\rho}{\sigma^2}\frac{1}{\theta}\Phi'\leq 0$. The second derivative is $F_{22}(\theta'|\theta)=\frac{\rho}{\sigma^2}\frac{1}{\theta^2}(\Phi'-\frac{\rho}{\sigma^2}\Phi'')\geq 0$. Hence Assumption 7 is satisfied as well and Lemma (2) applies. Since $\mathbb{E}(\theta'|\theta)=\theta^\rho$, Assumption 8 holds, and Theorem 4 may be applied.*

Example 2 (Mixture distribution). *Suppose that $\Theta=(0,\overline{\theta})$, $K\geq 2$ and define $F(\theta'|\theta)=\sum_{k=1}^K\rho_k(\theta)F_k(\theta')$ where $F_k:\Theta\to[0,1]$ satisfy $F_k\geq F_{k+1}$ and $\rho_k(\theta):\Theta\to[0,1]$ are twice differentiable functions that are Lipschitz continuous on $[\eta,\infty)$ for any $\eta>0$ and satisfy $\sum_{k=1}^K\rho_k(\theta)=1$ for all $\theta\in\Theta$. The distribution satisfies Assumption 3. Assumption 6 holds as well since F is differentiable with $F_2(\theta'|\theta)=\sum_{k=1}^K\rho_k'(\theta)F_k(\theta')$. If, in addition, $\sum_{j=1}^k\rho_j''(\theta)\geq 0$ for all $k=1,\ldots,K$ then the second derivative $F_{22}(\theta'|\theta)=\sum_{k=1}^K\rho_k''(\theta)F_k(\theta')$ is nonnegative. Hence, Assumption*

13. Condition i. of their Proposition 3 follows from (5.7), while Condition iii. of their Proposition 3 follows from (1).

7 is satisfied and Lemma (2) applies. The lifetime utility and its derivative are given by $W(X,\theta) = \sum_{k=1}^{K} \rho_k(\theta) W_k(X)$ and $W_\theta(X,\theta) = \sum_{k=1}^{K} \rho_k'(\theta) W_k(X)$ where $W_k(X)$ is the expected utility under F_k. Since ρ is Lipschitz continuous on $[\eta,\infty)$, Assumption 8 holds, and Theorem 4 may be applied.

Example 3 (Pareto distribution). *Suppose that $\lambda > 1$, $\rho \in [0,1]$ and $\Theta = (\kappa^{1-\rho},\infty)$, where $\kappa = \frac{\lambda-1}{\lambda}$. Define $F(\theta'|\theta) = 1 - \left(\frac{\theta'}{\kappa\theta^\rho}\right)^{-\lambda}$ if $\theta' > \kappa\theta^\rho$ and zero otherwise. The distribution clearly satisfies Assumption 3. Assumption 6 holds as well since F is differentiable for $\theta' \geq \kappa\theta^\rho$ with $F_2(\theta'|\theta) = -\rho\lambda\frac{1}{\theta}\left(\frac{\theta'}{\kappa\theta^\rho}\right)^{-\lambda}$. The second derivative is $F_{22}(\theta'|\theta) = \rho\lambda(1-\rho\lambda)\frac{1}{\theta^2}\left(\frac{\theta'}{\kappa\theta^\rho}\right)^{-\lambda}$. The sign of F_{22} depends on whether $\rho\lambda$ is greater or smaller than one, but in both cases it does not change sign and F_2 satisfies Assumption 7. Hence W_θ is given by (5.6), which simplifies to $W_\theta(X,\theta) = \frac{\rho\lambda}{\theta}\left[W(X,\theta) - \underline{W}(X,\theta)\right]$ where $\underline{W}(X,\theta) = U\left(X_1(\kappa\theta^\rho),\kappa\theta^\rho\right) + \beta W(X_2^\infty(\kappa\theta^\rho);\kappa\theta^\rho)$. Since $\mathbb{E}(\theta'|\theta) = \theta^\rho$, Assumption 8 holds, and Theorem 4 may be applied.*

6. A RELAXED PROBLEM

Define a *relaxed* problem by replacing the temporary incentive compatibility constraint (3.2) by the envelope condition (5.7) for all histories. The critical simplification brought by the relaxed problem is that the envelope condition (5.7) only contains the lifetime utility of the truthteller and her marginal lifetime utility, rather than the continuation utility of all possible types. As will be shown, this implies that the set of pairs of the lifetime utility and the marginal lifetime utility that can be obtained by some allocation satisfying the envelope condition (5.7) has a recursive structure. The social planner's problem will then also have a recursive structure, and its state space will be given by the space of lifetime and marginal lifetime utilities.

6.1. *Characterizing Admissible Utilities*

An allocation $X \in \mathcal{X}$ is said to be *admissible* if it satisfies the envelope condition (5.7) for all histories. The set of the lifetime utility and the marginal lifetime utility pairs, $\mathcal{V}^R(\theta)$, obtained by some admissible allocation is given by:

$$\mathcal{V}^R(\theta) = \{W(X;\theta), W_\theta(X;\theta) \,|\, X \text{ is admissible}\},$$

where R stands for "Relaxed". The set $\mathcal{V}^R(\theta)$ has the following properties:

Lemma 6. *$\mathcal{V}^R(\theta)$ is nonempty, compact and convex for all $\theta \in \Theta$.*

A recursive representation of $\mathcal{V}^R$ is derived as follows. Define an *allocation rule* $z \equiv (x,v,h)$ to be a triplet of measurable functions $x : \Theta \to \mathbb{X}$, $v : \Theta \to \mathbb{R}$ and $h : \Theta \to \mathbb{R}$. The function x specifies the current consumption of an agent, v represents the continuation utility of the truthtelling agent, and h represents a marginal change in the continuation utility of the truthtelling agent. Take $\mathcal{V}$ to be any nonempty subset of $\mathbb{R}^2 \times \Theta$. The allocation rule is said to be *admissible* with respect to $\mathcal{V}$ if the continuation utility and the marginal continuation utility belong to the set $\mathcal{V}(\theta)$:

$$[v(\theta), h(\theta)] \in \mathcal{V}(\theta) \qquad \forall\theta \in \Theta, \tag{6.9}$$

and the envelope condition holds:

$$U\left(x(\theta),\theta\right)+\beta v(\theta)=\int_{\underline{\theta}}^{\theta}\left[U_{\theta}\left(x(\varepsilon),\varepsilon\right)+\beta h(\varepsilon)\right]d\varepsilon+\underline{w}, \tag{6.10}$$

where $\underline{w}=\lim_{\theta\to\underline{\theta}}U\left(x(\theta),\theta\right)+\beta v(\theta)$. The set of all allocation rules that are admissible w.r.t. $\mathcal{V}$ is denoted by $\mathcal{Z}(\mathcal{V})$.

An admissible allocation rule z generates a lifetime utility $w(z;\theta_-)$ and a marginal lifetime utility $g(z;\theta_-)$ given by

$$w(z;\theta_-)=\int_{\Theta}\left[U\left(x(\varepsilon),\varepsilon\right)+\beta v(\varepsilon)\right]dF(\varepsilon|\theta_-), \tag{6.11}$$

$$g(z;\theta_-)=\int_{\Theta}\left[U\left(x(\varepsilon),\varepsilon\right)+\beta v(\varepsilon)\right]dF_2(\varepsilon|\theta_-). \tag{6.12}$$

The set of all lifetime utility and marginal utility pairs that are generated by some allocation rule that is admissible with respect to $\mathcal{V}$ defines an operator $\mathcal{T}$:

$$\mathcal{T}\mathcal{V}(\theta_-)=\left\{(w(z;\theta_-),g(z;\theta_-))\,|\,z\text{ is admissible w.r.t. }\mathcal{V}\right\}$$

The set $\mathcal{V}^R$ is the fixed point of the operator $\mathcal{T}$:

Theorem 7.

$$\mathcal{T}\mathcal{V}^R=\mathcal{V}^R.$$

Hence, any pair of lifetime and marginal lifetime utilities that can be obtained by an admissible allocation can also be obtained by an allocation rule that is admissible with respect to $\mathcal{V}^R$ and vice versa.

6.2. *A Relaxed Social Planner's Problem*

It is useful to separate the social planner's problem into two stages. In the second stage (called the auxiliary planning problem in Fernandes and Phelan (2000)), the social planner minimizes the costs of delivering a given pair of a promised and marginal promised utility $(w,g)\in\mathcal{V}^R(\theta)$ to a $\theta-$ type agent. In the first stage, the social planner chooses the initial marginal promised utility g_1 that minimizes the costs of delivering the lifetime utility $w_1\in\mathcal{W}(\theta_0)$.

Consider first the second stage. An allocation $X\in\mathcal{X}$ is said to *support* (w,g,θ) if it is admissible and delivers a lifetime utility $w=W(X;\theta)$ and a marginal lifetime utility $g=W_\theta(X;\theta)$. Define $\Gamma^R:\mathcal{V}^R\twoheadrightarrow\mathcal{X}$ to be the set of all such allocations:

$$\Gamma^R(w,g,\theta)=\left\{X\in\mathcal{X}\,|\;X\text{ supports }(w,g,\theta)\right\}.$$

The social planner selects an allocation that minimizes the costs among all the allocations that support (w,g,θ). The minimized costs $P^*:\mathcal{V}^R\to\mathbb{R}$ are given by

$$P^*(w,g,\theta_-)=\min_{X\in\mathcal{X}}\left\{D(X;\theta)\,|\,X\in\Gamma^R(w,g,\theta_-)\right\}. \tag{6.13}$$

If $X^*:\mathcal{V}^R\to\mathcal{X}$ attains the minimum of the social planner's problem (6.13) then it is called *second stage efficient*.

6.2.1. Second Stage: A Recursive Representation. An allocation rule is said to *support* (w, g, θ) if it is admissible with respect to $\mathcal{V}^R$ and delivers a lifetime utility $w = w(z; \theta)$ and a marginal lifetime utility $g = g(z; \theta)$.[14] The set of all such allocation rules $\gamma^R : \mathcal{V}^R \twoheadrightarrow \mathcal{Z}(\mathcal{V}^R)$ is defined by

$$\gamma^R(w, g, \theta) = \{z \in \mathcal{Z} \mid \ z \text{ supports } (w, g, \theta)\} . \tag{6.14}$$

Consider a cost function $P : \mathcal{V}^R \to \mathbb{R}$ that solves the following Bellman equation:

$$P(w, g, \theta_-) = \min_{z \in \gamma^R(w,g,\theta_-)} \int_\Theta \left[R\left(x(\theta)\right) + qP\left(v(\theta), h(\theta), \theta\right)\right] dF(\theta|\theta_-), \tag{6.15}$$

The *efficient* allocation rule is given by functions $z^* : \mathcal{V}^R \times \Theta \to \mathbb{X} \times \mathcal{V}^R$, that attain the minimum of the right-hand side of (6.15). The solution of the Bellman equation (6.15) has the following properties:

Theorem 8.

(i) There exists a unique bounded and continuous function P that solves the Bellman equation (6.15). The efficient allocation rule z^* admits a measurable selection.
(ii) The function P is convex in (w, g). If R is strictly convex in x then P is strictly convex in (w, g) and the efficient allocation rule z^* is essentially unique and measurable.

The relationship between the solution of the social planner's problem (6.13) and the corresponding recursive formulation (6.15) is now studied. It is first shown that there is a one-on-one relationship between admissible allocations and sequences of admissible allocation rules. To state the result formally, define $V_t(\theta^{t-1}) = W(X_t^\infty(\theta^{t-1}); \theta_{t-1})$ and $H_t(\theta^{t-1}) = W_\theta(X_t^\infty(\theta^{t-1}); \theta_{t-1})$ to be the stochastic precesses for lifetime and marginal lifetime utilities in the relaxed problem. Then one obtains the following result:

Lemma 9. *$X \in \Gamma^R(w, g, \theta_0)$ if and only if, for all $\theta^t \in \Theta^t$ and all $t \geq 1$,*

$$[X_1(\cdot), V_2(\cdot), H_2(\cdot)] \in \gamma^R(w, g, \theta_0) \tag{6.16}$$

$$\left[X_t(\theta^{t-1}, \cdot), V_{t+1}(\theta^{t-1}, \cdot), H_{t+1}(\theta^{t-1}, \cdot)\right] \in \gamma^R\left(V_t(\theta^{t-1}), H_t(\theta^{t-1}), \theta_{t-1}\right) . \tag{6.17}$$

Now, take the efficient allocation rule z^* and any $(w, g, \theta_0) \in \mathcal{V}^R$ and define an allocation X as follows. Set $V_1(w, g, \theta_0) = w$ and $H_1(w, g, \theta_0) = g$. For $t \geq 2$ let $V_t(w, g, \theta^{t-1})$ and $H_t(w, g, \theta^{t-1})$ solve difference equations

$$V_{t+1}(w, g, \theta^t) = v^*\left(V_t(w, g, \theta^{t-1}), H_t(w, g, \theta^{t-1}), \theta_{t-1}, \theta_t\right)$$
$$H_{t+1}(w, g, \theta^t) = h^*\left(V_t(w, g, \theta^{t-1}), H_t(w, g, \theta^{t-1}), \theta_{t-1}, \theta_t\right)$$

Since v^* and h^* are measurable with respect to its arguments by Theorem 8, V_{t+1} and H_{t+1} are measurable with respect to θ^t for all $t \geq 1$. Then define X by

$$X_t(w, g, \theta^t) = x^*\left(V_t(w, g, \theta^{t-1}), H_t(w, g, \theta^{t-1}), \theta_{t-1}, \theta_t\right) .$$

Since x^* is measurable with respect to its first argument and V_t and H_t are measurable with respect to θ^{t-1}, the allocation X_t is measurable with respect to θ^t for all $t \geq 1$ as

14. Those constraints are called a promise keeping constraint and a marginal threat keeping constraint.

well. Hence $X(w,g,\theta_0)$ is an allocation and $X:\mathcal{V}^R \to\in \mathcal{X}$. The allocation X is called an allocation *generated by* the efficient allocation rule z^*. Using Lemma 9, one obtains:

Theorem 10.

(i) $P = P^*$.
(ii) If z^* *is efficient and* X^* *is generated by* z^**, then* X^* *is second stage efficient.*

6.2.2. First Stage. The first stage corresponds to period 1 when the social planner is only constrained by the promised utility of the truthtelling agent, $w_1 \in \mathcal{W}(\theta_0)$, and so chooses the cost minimizing promised utility function that delivers w_1 to the truthteller. That is, he is free to choose any marginal lifetime utility g as long as it delivers a lifetime utility w_1 to the truthteller and satisfies $(w_1,g)\in\mathcal{V}^R(\theta_0)$:

$$P^R(w_1,\theta_0) = \min_{g\in\mathbb{R}} \left\{P^*(w_1,g,\theta_0)\,|\,(w_1,g)\in\mathcal{V}^R(\theta_0)\right\}. \tag{6.18}$$

The initial marginal promised utility is *efficient* if it attains the minimum of the right-hand side of (6.18) and is denoted by $h_1^*:\mathcal{W}(\Theta)\times\Theta\to\mathbb{R}$. It follows from Theorem (8) and from compactness of $\mathcal{V}^R$ that such a function exists. Also, if X^* is generated by the efficient allocation rule z^*, then the allocation $X^*\left(w_1,h_1^*(\theta_0),\theta_0\right)$ solves the relaxed social planner's problem.

6.3. *Validity of the Relaxed Problem*

I will now examine conditions under which the solution to the relaxed problem coincides with the solution to the unrelaxed problem. One way is to check whether the allocation $X^*\left(w_1,h_1^*(\theta_0),\theta_0\right)$ generated by the efficient allocation rule z^* satisfies the temporary incentive compatibility constraint (3.2) or the sufficiency conditions (5.8). A more convenient approach is to find conditions on the allocation rule directly.

To do so, one needs to characterize the lifetime utility for all possible reporting strategies. If a θ- type agent reports $\hat\theta$, the continuation and marginal continuation utilities are given by

$$\hat v(w,g,\theta_-,\hat\theta;\theta) = \int_\Theta s^*\left(v^*(w,g,\theta_-,\hat\theta),h^*(w,g,\theta_-,\hat\theta),\hat\theta,\varepsilon\right)\,dF(\varepsilon|\theta) \tag{6.19}$$

$$\hat h(w,g,\theta_-,\hat\theta;\theta) = \int_\Theta s^*\left(v^*(w,g,\theta_-,\hat\theta),h^*(w,g,\theta_-,\hat\theta),\hat\theta,\varepsilon\right)\,dF_2(\varepsilon|\theta), \tag{6.20}$$

where

$$s^*(w,g,\theta_-,\theta) = U\left(x^*(w,g,\theta_-,\theta),\theta\right)+\beta v(w,g,\theta_-,\theta)$$

is the lifetime utility of an agent after the current shock has been revealed (*ex-post utility*). A relaxed problem is called *valid* if the efficient allocation rule is such that for all $(w,g,\theta_-)\in\mathcal{V}^R$ the agent prefers to report truthfully:

$$\theta\in\arg\max_{\hat\theta\in\Theta}\left\{U\left(x^*(w,g,\theta_-,\hat\theta),\theta\right)+\beta\hat v(w,g,\theta_-,\hat\theta;\theta)\right\}. \tag{6.21}$$

If the relaxed social planner's problem is valid, then the solution coincides with the solution to the unrelaxed social planner's problem (3.4):

Theorem 11. *If the relaxed social planner's problem is valid then* $P^R = P^{IC}$.

The verification procedure is similar to the one examined by Abraham and Pavoni (2008) for the hidden savings problem. However, condition (6.21) is simpler in one key aspect: one does not need to recompute the whole agent's problem to verify it.[15] Condition (6.21) is used by Farhi and Werning (2010) to verify the validity of their relaxed problem. Condition (6.21) can be checked directly, but one can also derive a sufficiency condition analogous to the monotonicity condition (5.8):

Lemma 12. *If*

$$D(w,g,\theta_-,\hat{\theta};\theta) \equiv U_\theta\left(x^*(w,g,\theta_-,\hat{\theta}),\theta\right) + \beta\hat{h}(w,g,\theta_-,\hat{\theta};\theta) \tag{6.22}$$

increases in $\hat{\theta}$ for almost all $\theta \in \Theta$ then the relaxed problem is valid.

The presence of the first term on the right-hand side of (6.22) is well understood from the static theory. The second term $\hat{h}$ is, however, quite complex. To get a better understanding, assume that the utility function is separable and affine in θ and let

$$m^*(w,g,\theta_-,\theta) = U_\theta\left(x^*(w,g,\theta_-,\theta)\right) + \beta h^*(w,g,\theta_-,\theta)$$

be the informational rent of a θ-type agent. Then one can write

$$D(w,g,\theta_-,\hat{\theta};\theta) = m^*(w,g,\theta_-,\hat{\theta}) + \beta M(w,g,\theta_-,\hat{\theta};\theta). \tag{6.23}$$

where, after integration by parts,

$$M(w,g,\theta_-,\hat{\theta};\theta) = \int_\Theta \left[F_2(\varepsilon|\hat{\theta}) - F_2(\varepsilon|\theta)\right] m^*\left[v^*(w,g,\theta_-,\hat{\theta}), h^*(w,g,\theta_-,\hat{\theta}),\hat{\theta},\varepsilon\right] d\varepsilon.$$

The report $\hat{\theta}$ thus affects the function D in two ways. First, the informational rent of the truthteller m^* depends on the current report $\hat{\theta}$. If the planner chooses a higher informational rent for agents with higher skills, the sufficiency condition is more likely to hold. Second, the misreporting agent experiences a different distribution of shocks. The significance of the distributional shift is captured in the function M.

Differentiating M with respect to $\hat{\theta}$ yields that for local deviations from the truthful report the function M is going to be increasing in $\hat{\theta}$ if $F_{22} \geq 0$. For global deviations from the optimum the relationship between $\hat{\theta}$ and M also depends on whether the realized informational rent *next period*, $m^*\left[v^*(w,g,\theta_-,\hat{\theta}), h^*(w,g,\theta_-,\hat{\theta}),\hat{\theta},\varepsilon\right]$, increases or decreases with the current report. Although it is not possible to determine the sign in general, the intuition is similar to the one for local deviations: If the distributional shift tends to increase the expected informational rent of the deviating agent, the monotonicity condition is more likely to hold. If $F_{22} = 0$ then $M = 0$ and we have:

Corollary 13. *Suppose that the utility is separable and affine in θ and that $F_{22} = 0$. Then the relaxed problem is valid if $m^*(w,g,\theta_-,\theta)$ increases in θ.*

The sufficiency condition is then essentially no different from the sufficiency condition in a static model: it requires the term under the integral on the right-hand side of the envelope condition to be increasing in the report, or, equivalently, the ex-post utility

15. Note also that the condition (6.21) may sometimes be unnecessarily strong. If the efficient contract always chooses the lifetime and marginal lifetime utilities from some subset $\hat{\mathcal{V}}^R$ of $\mathcal{V}^R$ and that $\left(w_1, h_1^*(w_1,\theta_0)\right) \in \hat{\mathcal{V}}^R(\theta_0)$, then it is enough to check whether (6.22) holds on $\hat{\mathcal{V}}^R$.

s^* to be convex in the report. Lemma 13 is applicable, for instance, in the case of the Mixture distribution as in Example 2 when ρ_k is linear in θ for all k.[16] Note also that Lemma 13 only applies when F has full support.

Typically, the sufficiency conditions will have to be checked numerically. One exception is an economy with linear quadratic utility and uniformly distributed shocks. In the online Appendix B I provide a closed form solution for its relaxed problem. I show that the function M is always decreasing in $\hat{\theta}$. Despite that, the relaxed problem is valid, because m^* is increasing in $\hat{\theta}$ sufficiently fast. The result holds for all coefficients of autocorrelation $\rho \in [0,1]$, with $\rho = 1$ being a marginal case where D is independent of $\hat{\theta}$. The

7. A MIRRLEESEAN ECONOMY WITH RANDOM WALK

Consider a simple economy where $c \in \mathbb{R}_+$ is consumption and $y \in \mathbb{R}_+$ is output. The utility function is given by

$$\hat{U}(c,y,\theta) = \ln c - \frac{\alpha}{1+\eta}\left(\frac{y}{\theta}\right)^{1+\eta}$$

where η is inverse of the compensated elasticity of labor supply. This is a typical Mirrlees-type economy with labor supply given by $l = \frac{y}{\theta}$ where the private information shock θ represents individual's skill. The cost function is given by $\hat{R} = c - y$. The skill shocks are Pareto distributed as in Example (3) and are assumed to follow a random walk with $\mathbb{E}(\theta'|\theta) = \theta$. The utility specification and the assumption of random walk are both standard ingredients in the empirical macro and public finance literature.[17] The assumption of Pareto distributed shocks is consistent with the distribution of earnings at the right tail.[18]

I transform the problem as follows. Let $x_1 = \ln c$, $x_2 = \alpha \frac{y^{1+\eta}}{1+\eta}$. Then, one can write $U(x,\theta) = x_1 - \theta^{-(1+\eta)} x_2$ $R(x) = e^{x_1} - \gamma(1+\eta) x_2^{\frac{1}{1+\eta}}$, where $\gamma = \left(\frac{1+\eta}{\alpha}\right)^{\frac{1}{1+\eta}} \frac{1}{1+\eta}$. The transformed functions satisfy Assumptions (2), (4), and (5).

7.1. *The Relaxed Social Planner's Problem*

The problem has several properties that simplify the problem. First, the set of admissible utilities $\mathcal{V}^R$ has a very simple characterization: $\mathcal{V}^R = \mathbb{R} \times \mathbb{R}_+ \times \Theta$. That is, any promised utility and nonnegative marginal threat utility is implementable. Second, as Example (3) shows, the marginal lifetime utility (6.12) is proportional to the spread between the expected lifetime utility and the utility of the agent with the lowest possible shock, $\underline{w}(z;\theta_-)$:

$$g(z;\theta_-) = \frac{\lambda}{\theta_-}[w(z;\theta_-) - \underline{w}(z;\theta_-)]. \tag{7.24}$$

Third, the value function satisfies the following homogeneity property:[19]

16. The example of the mixture distribution is also studied by Fukushima and Waki (2011). They show that one can reduce the state space to K state variables. The first-order approach further strengthens the results in that the state space can be reduced to 2 state variables for any K.

17. See e.g. Heathcote et al. (2009) or Heathcote et al. (2011).

18. See Diamond (1998) or Saez (2001).

19. Proof of the Lemma is in the online Appendix C.

Lemma 14. *The value function satisfies*

$$P(w, g, \theta_-) = \theta_- P(w - \frac{\ln \theta_-}{1-\beta}, g\theta_-, 1).$$

In what follows I will solve the normalized problem with $\theta_- = 1$. It is convenient to rescale the policy rules h and v by defining $r(\theta) = \theta h(\theta)$ and $q(\theta) = v(\theta) - \frac{\ln \theta}{1-\beta}$. The set of allocation rules $\gamma^R(w, g, 1)$ is characterized as follows. First, the lifetime utility the allocation rules deliver needs to be w, where

$$w = \int_\kappa^\infty \left[x_1(\theta) - \theta^{-(1+\eta)} x_2(\theta) + \beta q(\theta)\right] f(\theta)\, d\theta + \frac{\beta}{1-\beta}\left(\frac{1}{\phi} + \ln \kappa\right). \tag{7.25}$$

Second, the marginal lifetime utility has to be equal to g. By using equation (7.24), the marginal threat keeping constraint is

$$g = \lambda(w - \underline{w}). \tag{7.26}$$

Third, the continuation lifetime and marginal lifetime utility has to be admissible with respect to $\mathcal{V}^R$. This requires that the envelope condition holds,

$$x_1(\theta) - \theta^{-(1+\eta)} x_2(\theta) + \beta q(\theta) + \frac{\beta \ln \theta}{1-\beta} = \int_\kappa^\infty \left[(1+\eta)\varepsilon^{-(2+\eta)} x_2(\varepsilon) + \beta r(\varepsilon)\right] \frac{d\varepsilon}{\varepsilon} + \underline{w}, \tag{7.27}$$

and that

$$r(\theta) \geq 0 \quad \forall \theta \geq \kappa. \tag{7.28}$$

The function $P : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ then satisfies the following Bellman equation:

$$P(w, g) = \min_{x,q,r} \int_\kappa^\infty \left[e^{x_1(\theta)} - \gamma(1+\eta) x_2(\theta)^{\frac{1}{1+\eta}} + \beta\theta P\left(q(\theta), r(\theta)\right)\right] f(\theta|1)\, d\theta, \tag{7.29}$$

subject to $x_2 \geq 0$ and the constraints (7.25), (7.26), (7.27), and (7.28). The optimal policy rules are $x_1(w, g, \theta)$, $x_2(w, g, \theta)$, $q(w, g, \theta)$ and $r(w, g, \theta)$.

There are several notable properties of the dynamic program (7.29). The value function P reaches, for a given w, its minimum at some $g^*(w) \geq 0$. Moreover, if $g \leq g^*(w)$ one can show that $r(w, g, \theta) \leq g^*(w)$ for all $\theta \geq \kappa$. This result is important because it implies that the downward sloping part of the value $[0, g^*(w)]$ constitutes an ergodic set. Since in the first period the social planner chooses $g = g^*(w)$, the efficient allocation will always lie in this ergodic set.

The dynamics of the efficient allocation is driven by the fact that the marginal continuation utility affects the efficient contract in two distinct ways. First, in the next period, the marginal continuation utility determines the amount of dispersion in the ex-post utilities and therefore the amount of insurance the social planner must provide. Lower marginal continuation utilities are associated with less dispersions in ex-post utilities. Second, the marginal continuation utility affects the future informational rent. Higher marginal continuation utilities are associated with higher informational rents. Those two effects are obviously related, since a lower dispersion of ex-post utilities in the future makes the current shock less relevant (if there is no dispersion in the ex-post utilities in the next period then the current shock is irrelevant, and bears no future informational rent). Informational rents are not desirable because they limit the amount of insurance the social planner can provide. This explains why the efficient contract never chooses $r(w, g, \theta) > g^*(w)$. Such a choice would increase the current informational rent and, at the same time, make the dispersion in ex-post utilities inefficiently high. On the other hand, choosing some $r(w,g,\theta) \leq g^*(w)$ involves a trade-off: it makes it cheaper to provide insurance today, at the expense of being forced to provide more insurance tomorrow.

7.2. *Numerical Simulations*

The discount rate β is set to 0.96. The shape parameter of the Pareto distribution λ is chosen to be such that the variance of the log shock innovations in the productivity shock ζ is 0.007, based on the evidence in Heathcote et al. (2005).[20] This yields $\lambda = 11.952$. The utility function parameter η is set equal to 5 to match the Frisch elasticity of labor supply of 0.2. The parameter α is normalized to 1. The parametrization, as well as the functional forms, are similar to the ones used by Farhi and Werning (2010).[21]

The space of shocks is discretized with $N = 100$ logarithmically distributed grid points, with $\theta_1 = \kappa$ and θ_N being chosen in such away that $F(\theta_N) = 0.999$. On each interval $[\theta_n, \theta_{n+1})$, the optimal policy functions x_2, r and z are assumed constant, while x_1 is affine, with its slope chosen to satisfy the envelope condition on every interval. The same approximation is taken on the right tail $[\theta_N, \infty)$. The dynamic program (7.29) is solved by iterating on the value function until it converges. The convergence criterion ϵ is set equal to 1^{-3}. The value function is approximated on a 2-dimensional grid (w_i, g_j), $i = 1, \ldots, I$, $j = 1, \ldots, J$ with linear interpolation for values outside of the grid. I set $g_1 = 0$ and g_J large enough so that the minimum of the value function is below g_J for each w_i. I extrapolate the value function below w_1 and above w_I. I choose $I = 64$ and $J = 90$ for the computations.[22]

Figure 1 shows the value function $P(w,g)$, as well as the cost minimizing marginal promised utility $g^*(w)$. When $g < g^*(w)$, the costs are decreasing relatively steeply in g. This is not surprising, since low g is associated with low informational rent, high amount of insurance, and hence with low labor supply and low marginal continuation utility. Moreover, low marginal continuation utility implies that those features will persist in the future.

Figure 2 shows how the social planner solves the tradeoff involved in choosing the marginal continuation utility. It plots the expected marginal continuation utility, $\mathbb{E}r(w,g) = \int_\kappa^\infty r(w,g,\varepsilon) f(\varepsilon)\, d\varepsilon$, for a promised utility w_0 that is such that the social planner breaks even at $g^*(w_0)$ (i.e. satisfying $P[w_0, g^*(w_0)] = 0$). The figure is drawn for $g \in [0, g^*]$. The expected marginal continuation utility is clearly increasing in g. In addition, the figure shows that $\mathbb{E}r(w_0, g) < g$. The social planner thus trades off a cheaper insurance today (because the informational rent today decreases) for an ex-post inefficiently high insurance in the future.

20. The variance of the hourly wage rate increase by 21 log points between age 21 and age 55 with time effects. This increase implies that the variance of the annual innovations must be 0.21 / 30 = 0.007.

21. There are three significant differences: Farhi and Werning (2010) consider an economy with finite lifetimes, they assume that the shocks are lognormally distributed, and they choose a Frisch elasticity of labor equal to 0.5.

22. The most critical assumption from the perspective of numerical computations is that the shocks follow a random walk since it allows the problem to be normalized and reduces its dimensionality. The assumption of Pareto distributed innovations simplifies the problem as well in that the marginal threat keeping constraint has an explicit expression, see equation (7.24). On the other hand, the assumption of logarithmic utility could be relaxed quite easily.

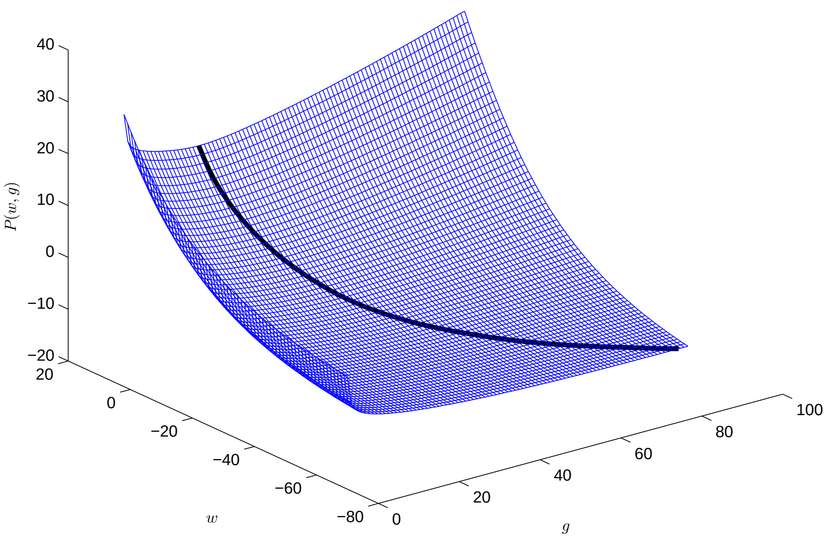

FIGURE 1
Value Function $P(w, g)$. Black line denotes $g^*(w)$

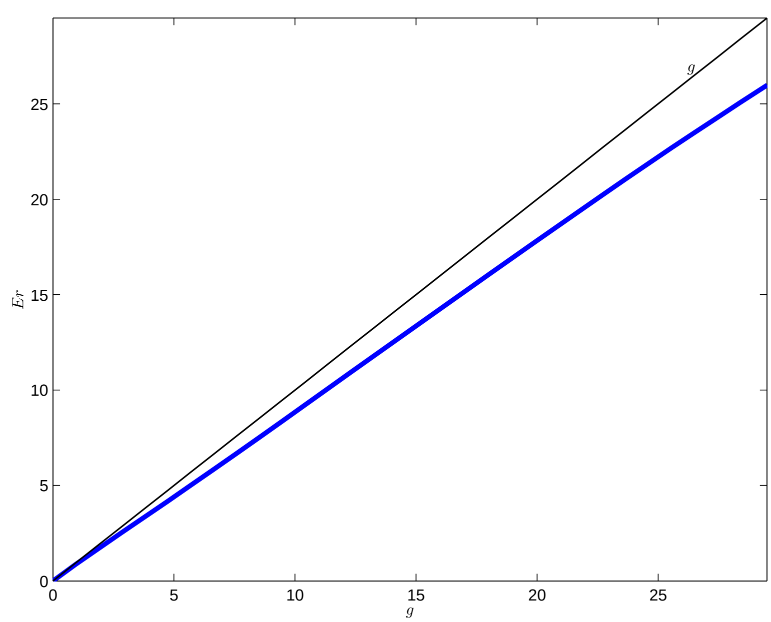

FIGURE 2
Expected Marginal Continuation Utility $\mathbb{E}r(w_0, g)$.

7.2.1. The wedges. A convenient way to summarize the extent of insurance is to look at intratemporal and intertemporal wedges. Because of the homogeneity of the

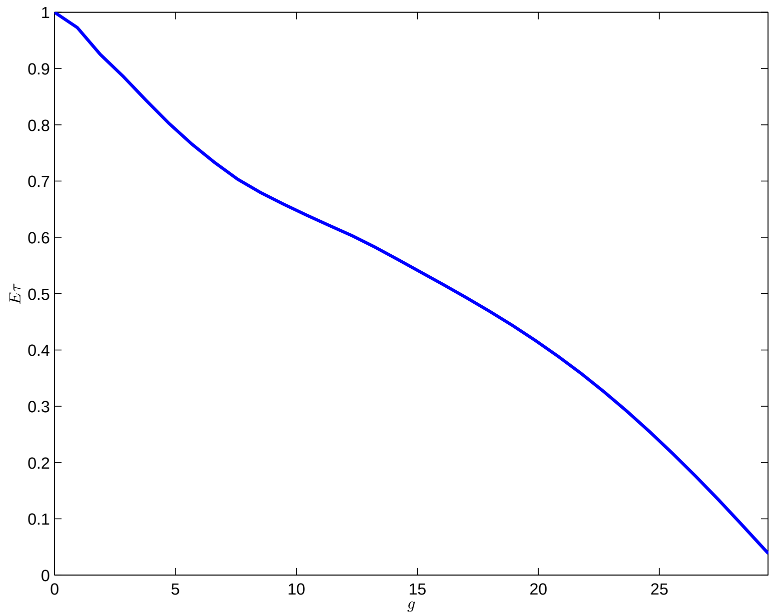

Figure 3
Expected Intratemporal Wedge $\mathbb{E}\tau(w_0, g)$.

optimal policy functions, the intratemporal wedge $\tau(w, g, \theta)$ and the intertemporal wedge $\delta(w, g, \theta)$ can be written as

$$\tau(w, g, \theta) = 1 - \frac{\theta}{\gamma} e^{x_1(w,g,\theta)} x_2(w, g, \theta)^{\frac{\eta}{1+\eta}}$$

$$\delta(w, g, \theta) = 1 - \theta e^{-x_1(w,g,\theta)} \left(\int_\kappa^\infty e^{-x_1[x(w,g,\theta), r(w,g,\theta), \varepsilon]} f(\varepsilon)\, d\varepsilon \right)^{-1}.$$

Figure 3 examines the intratemporal wedge. It shows the expected intratemporal wedge, $\mathbb{E}\tau(w, g) = \int_\kappa^\infty \tau(w, g, \varepsilon) f(\varepsilon)\, d\varepsilon$, for the break even promised utility w_0 as a function of the marginal promised utility. The expected intratemporal wedge decreases with the marginal promised utility. At the cost minimizing promised utility $g^*(w_0)$, the expected intratemporal wedge is only 5%. It decreases with the marginal promised utility and converges to 1 as the marginal promised utility g decreases toward zero. The simulations also reveal that the intratemporal wedge decreases with the continuation utility as well. The dynamics of the expected intratemporal wedge, as well as its magnitude at the beginning are similar to the one obtained by Farhi and Werning (2010)).

Ales and Maziero (2009) study an economy with finite lifetimes and i.i.d. shocks and find that the intratemporal wedge increases with age. While their model differs from ours, the argument is somewhat similar. In their model the intratemporal wedge increases because as people age, the social planner is less able to vary the continuation utility and must provide incentives through intratemporal distortions. In this model, lower marginal promised utilities are associated with lower dispersion of ex post lifetime utilities. Hence, the current shock becomes less important in determining the continuation lifetime utility and the social planner must rely more heavily on intratemporal distortions to provide

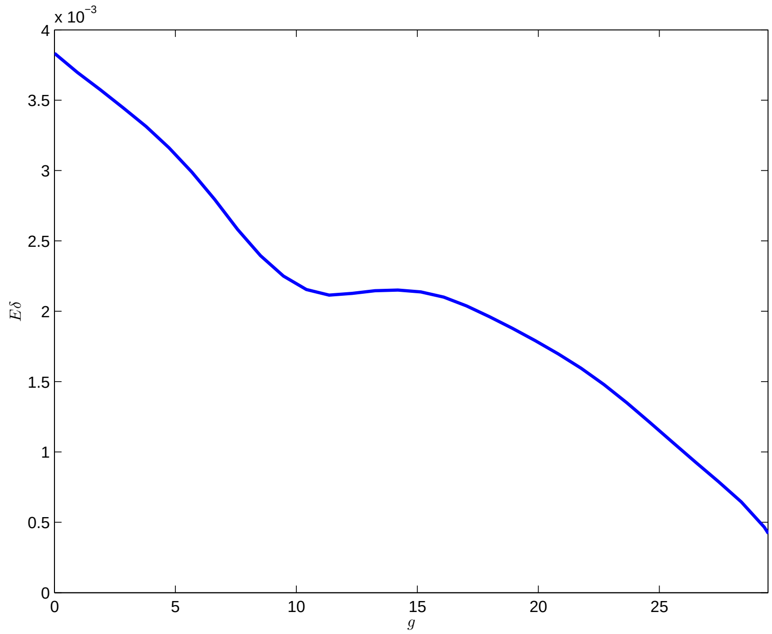

FIGURE 4
Expected Intertemporal Wedge $\mathbb{E}\delta(w_0, g)$

incentives.

The intertemporal wedge is plotted in figure 4, again for the break even promised utility w_0. As expected, it is always nonnegative. The intertemporal wedge is decreasing in the marginal promised utility. It also tends to increase in the promised utility. In terms of magnitude, it is between 0.05% and 0.4%, which is consistent with the findings in Farhi and Werning (2010)).

7.2.2. Validity of the Relaxed Problem. The solution to the dynamic program does not specify what happens if the agent reports $\hat{\theta} > \theta$ and receives shock $\theta' \in [\kappa\theta, \kappa\hat{\theta})$ next period. This information is, however, needed to compute the deviator's continuation utility (6.19), and one needs to extend the allocation for such values.

Because of the homogeneity of the optimal policy functions, one only needs to extend the functions $x_1(w, g, \theta)$, $x_2(w, g, \theta)$, $v(w, g, \theta)$, and $r(w, g, \theta)$ for the off-equilibrium values $\theta \in (0, \kappa)$ in a way that does not violate incentive compatibility. There is no unique way of doing so, and any extension that satisfies the envelope condition is a suitable candidate. I do so by setting, for all $\theta \in (0, \kappa)$, $x_2^*(w, g, \theta) = \bar{x}_2$ and $r^*(w, g, \theta) = \bar{r}$ for some values of $\bar{x}_2$ and $\bar{r}$, and then compute $s^*(w, g, \theta)$ from the envelope condition. The envelope condition then holds by construction.

I verify the validity of the relaxed problem by directly computing the right-hand side of equation (6.21) and check whether truthtelling is optimal. I compute the right-hand side of (6.21) for all gridpoints (w_i, g_j) for which the dynamic program was computed, for 10 equispaced shock values θ, ranging from κ to θ_N, and for 100 equispaced reports $\hat{\theta}$, ranging again from κ to θ_N. The value of reporting $\hat{\theta}$ for a θ-type agent at a gridpoint

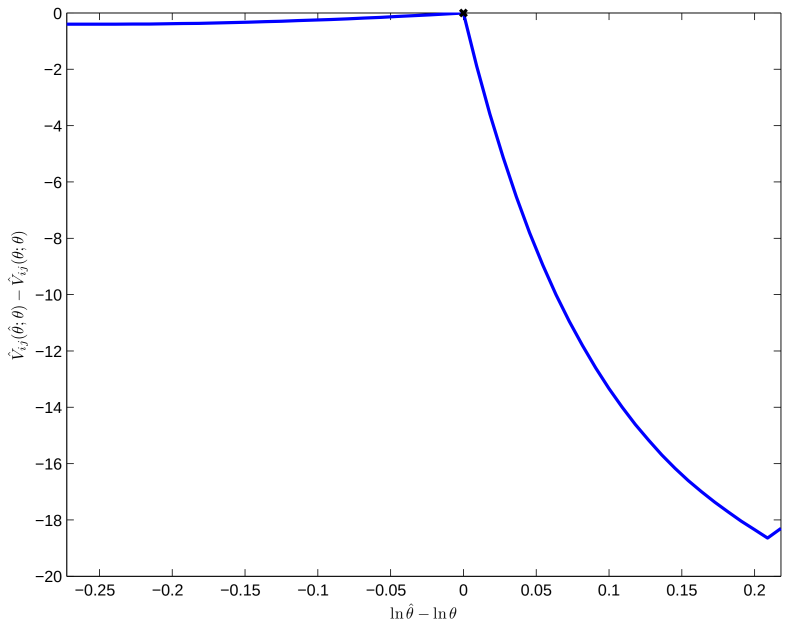

FIGURE 5
Gains from deviation for a given i,j and θ, a typical example

(w_i, g_j) is denoted by $\hat{V}_{ij}(\hat{\theta};\theta)$. Since the policy functions are either piecewise linear or step functions, their values can be computed for any $\hat{\theta}$ without a need for approximation.

The condition (6.21) is satisfied for all gridpoints in the state space. A typical pattern of the gains from deviation is shown in figure 5. The horizontal axis shows the difference between the report $\ln\hat{\theta}$ and the shock $\ln\theta$ for a given shock θ. The vertical axis shows the difference between the utility from reporting $\ln\hat{\theta}$, $\hat{V}_{ij}(\hat{\theta};\theta)$, and the utility from reporting truthfully, $\hat{V}_{ij}(\theta;\theta)$. Truthtelling is optimal, and the function is maximized at zero. In most cases, the most attractive non-local alternative to truthtelling is to report the lowest possible shock.[23] Nevertheless, the gain from reporting the lowest possible shock is negative for all gridpoints and all shocks.

8. CONCLUSIONS

The main contribution of this paper is methodological. The paper develops a method of solving dynamic private information models with a continuum of persistent shocks that is simple enough to be solved numerically. The first-order approach is used to simplify the structure of the incentive constraints. It is shown that the first-order approach not only simplifies the problem within each period (as is well known from static models), but has the added benefit that one only needs to keep track of the continuation utility of the truthtelling agent and her marginal continuation utility. In contrast, without the first-order approach one needs to keep track of the continuation utility of all possible deviating

23. Such a report would be attractive if the least able agent faces a large dispersion in ex-post utilities next period. Since the deviating agent has a more favorable distribution of shocks next period, he is more likely to obtain high ex-post utilities next period.

agents. Sufficiency conditions that guarantee that the first-order approach is valid are presented as well. I illustrate the first-order approach by quantitatively analyzing the efficient allocations in a Mirrleesean consumption-labor economy with Pareto distributed shocks that follow a random walk.

A major limitation of the first-order approach is that it is useful only as long as it is valid.[24] Since in most cases there are no conditions on the primitives that would guarantee validity, it is not possible to determine beforehand whether the first-order approach is valid or not. The application in this paper, as well as the applications in Farhi and Werning (2010) and Golosov et al. (2010), check the sufficiency conditions and all find that the first-order approach is valid. In previous versions of this paper I have computed the efficient allocations in taste shock economies with exponential utility and Pareto distributed shocks, and the first-order approach was valid as well. Perhaps the unanswered question now is not whether the first-order approach can be usefully applied at all, but where its limits are.

9. APPENDIX: PROOFS

Proof (**Lemma 2**). Fix X and let $S(\theta) = U\left(X_1(\theta), \theta\right) + \beta W(X_2^\infty(\theta); \theta)$. Then, $W(X;\theta)$ can be written as

$$W(X;\theta) = \int_\Theta S(\varepsilon)\, dF(\varepsilon|\theta). \tag{A-1}$$

Integrating the Lebesque-Stieltjes integral (A-1) by parts, one gets

$$\begin{aligned} W(X;\theta) - W(X;\hat\theta) &= \left[S(\varepsilon)\left(F(\varepsilon|\theta) - F(\varepsilon|\hat\theta)\right)\right]_{\underline{\theta}}^{\overline{\theta}} - \int_\Theta \left[F(\varepsilon|\theta) - F(\varepsilon|\hat\theta)\right] dS(\varepsilon) \\ &= -\int_\Theta \left[F(\varepsilon|\theta) - F(\varepsilon|\hat\theta)\right] dS(\varepsilon). \end{aligned} \tag{A-2}$$

If $\overline{\theta} < \infty$, then the second equality follows immediately. Suppose that $\overline{\theta} = \infty$. Since $|U|$ is bounded by $\overline{U}_0 + \overline{U}_1\theta$ and $|W|$ is bounded by $K_0 + K_1\theta$ by Assumptions 1, 2 and 3, $|S(\theta)| \le \hat B_0 + \hat B_1\theta$ for $\hat B_0 = \overline{U}_0 + \beta K_0$ and $\hat B_1 = \overline{U}_1 + \beta K_1$. Hence,

$$|S(\theta')\left(F(\theta'|\theta) - F(\theta'|\hat\theta)\right)| \le (\hat B_0 + \hat B_1\theta')|F(\theta'|\theta) - F(\theta'|\hat\theta)|.$$

The last term converges to zero as $\varepsilon \to \infty$. To see this, suppose by contradiction that $\theta'|F(\theta'|\theta) - F(\theta'|\hat\theta)|$ converges to some nonzero constant $\tilde B$. Then $|F(\theta'|\theta) - F(\theta'|\hat\theta)|$ converges to $\frac{\tilde B}{\theta'}$ and $F(\theta'|\theta)$ must converge to $1 - \frac{\tilde{\tilde B}}{\theta'}$ for some $\tilde{\tilde B}$. This implies that the conditional expectation $\mathbb{E}(\theta'|\theta)$ must be infinite, a contradiction with Assumption 3. One can then write

$$W_\theta(X;\theta) = \lim_{\eta\to 0}\frac{W(X;\theta+\eta) - W(X;\theta_0)}{\eta} = -\lim_{\eta\to 0}\int_\Theta \frac{F(\varepsilon|\theta+\eta) - F(\varepsilon|\theta)}{\eta}\, dS(\varepsilon)$$

where the second equality follows from (A-2). Note, that since Θ is open, $\theta_0 + \eta \in \Theta$ for η sufficiently small, and so W_θ is correctly defined.

By Assumption 7, $F(\theta'|\theta)$ is either convex or concave in θ. Suppose it is convex. Then, $F(\theta'|\alpha\hat\theta + (1-\alpha)\theta) \ge \alpha F(\theta'|\hat\theta) + (1-\alpha)F(\theta'|\theta)$. for any $\alpha \in [0,1]$. Let $\eta_2 \ge \eta_1 \ge 0$

24. Werning (2001) argues that an invalid first-order approach can still be useful as an upper bound on the Pareto frontier.

and set $\alpha = \frac{\eta_1}{\eta_2}$ and $\hat{\theta} = \theta + \eta_2$. Rearranging terms, one gets

$$-\frac{F(\theta'|\theta+\eta_1) - F(\theta'|\theta)}{\eta_1} \leq -\frac{F(\theta'|\theta+\eta_2) - F(\theta'|\theta)}{\eta_2}.$$

Hence, the sequence $-\frac{F(\theta'|\theta+\eta)-F(\theta'|\theta)}{\eta}$ is monotonically increasing in η. Since $F_2 \leq 0$ by Assumption 7, this sequence is nonnegative. Assumption 6 ensures that the limit of this sequence exists and is equal to $-F_2(\theta'|\theta)$. The Monotone Convergence Theorem can be applied and

$$W_\theta(X;\theta) = -\int_\Theta \lim_{\eta\to 0} \frac{F(\varepsilon|\theta+\eta) - F(\varepsilon|\theta)}{\eta}\, dS(\varepsilon) = -\int_\Theta F_2(\varepsilon|\theta) dS(\varepsilon).$$

where the second equality follows from the definition of F_2. Similar results obtain if $F(\theta'|\theta)$ is concave in θ.

Suppose now that θ_n converges to θ. By Assumption 7, $F_2(\theta'|\theta)$ is monotone in θ, and so applying the Monotone Convergence Theorem once more yields

$$\begin{aligned}\lim_{n\to\infty} W_\theta(X;\theta_n) &= -\lim_{n\to\infty}\int_\Theta F_2(\varepsilon|\theta_n) dS(\varepsilon) = -\int_\Theta \lim_{n\to\infty} F_2(\varepsilon|\theta_n) dS(\varepsilon) = -\int_\Theta F_2(\varepsilon|\theta) dS(\varepsilon) \\ &= W_\theta(X;\theta).\end{aligned}$$

Hence, W_θ is continuous in θ. Finally, integrating by parts again, one gets

$$W_\theta(X;\theta) = [S(\varepsilon)F_2(\varepsilon|\theta)]_{\underline{\theta}}^{\overline{\theta}} + \int_\Theta S(\varepsilon)\, dF_2(\varepsilon|\theta) = \int_\Theta S(\varepsilon)\, dF_2(\varepsilon|\theta),$$

where the first term on the right-hand side vanishes by the same arguments that were used to show (A-2).

Proof (**Lemma 3**). Take any $\eta \in \Theta$. By Assumption 8, $\mathbb{E}(\theta'|\theta)$ is Lipschitz continuous in θ on $[\eta, \overline{\theta})$. Let L be the Lipschitz constant. Since $|U|$ is bounded by $\overline{U}_0 + \overline{U}_1\theta$ and $|W|$ is bounded by $K_0 + K_1\theta$ by Assumptions 1, 2 and 3, one has for all $\theta, \hat{\theta} \in [\theta_0, \overline{\theta}]$

$$\begin{aligned}|W(X;\theta) - W(X;\hat{\theta})| &\leq \int_\Theta |U(X_1(\varepsilon), \varepsilon) + \beta W(X_2^\infty;\varepsilon)| \left|dF(\varepsilon|\theta) - dF(\varepsilon|\hat{\theta})\right| \\ &\leq \int_\Theta \left(\overline{U}_0 + \overline{U}_1\varepsilon + \beta K_0 + \beta K_1\varepsilon\right) \left|dF(\varepsilon|\theta) - dF(\varepsilon|\hat{\theta})\right| \\ &= (\overline{U}_1 + \beta K_1) \int_\Theta \varepsilon \left|dF(\varepsilon|\theta) - dF(\varepsilon|\hat{\theta})\right| \\ &\leq (\overline{U}_1 + \beta K_1) L |\theta - \hat{\theta}|.\end{aligned}$$

Hence, $W(X;\theta)$ is Lipschitz continuous on $[\eta, \overline{\theta})$.

Proof (**Theorem 4**). The proof verifies the conditions of Theorem 2 of Milgrom and Segal (2002). Take any $\eta \in \Theta$ and $\theta_t \geq \eta$. By Lemma (2), $W(X;\theta)$ is differentiable in θ, and by Lemma (3), it is Lipschitz continuous on $[\eta, \overline{\theta})$. By Assumption 9, $U(x,\theta)$ is differentiable and Lipschitz continuous in θ. Therefore,

$$\hat{S}_t(\theta^{t-1}, \hat{\theta}_t, \theta_t) = U\left(X_t(\theta^{t-1}, \hat{\theta}_t), \theta_t\right) + \beta W\left(X_{t+1}^\infty(\theta^{t-1}, \hat{\theta}_t); \theta_t\right). \tag{A-3}$$

is differentiable in θ_t on $(\eta, \overline{\theta})$ and Lipschitz continuous in θ_t on $[\eta, \overline{\theta})$. Hence, it is absolutely continuous and has a bounded derivative w.r.t. θ_t on $(\eta, \overline{\theta})$. Since (3.2) holds,

$\hat{S}_t(\theta^{t-1}, \hat{\theta}_t, \theta_t)$ is maximized at $\hat{\theta}_t = \theta_t$ and by Theorem 2 of Milgrom and Segal (2002), it can be represented by an integral of its derivative. Thus,

$$\hat{S}_t(\theta^{t-1}, \theta_t, \theta_t) = \int_{\eta}^{\theta_t} \frac{\partial}{\partial \theta_t} \hat{S}_t(\theta^{t-1}, \varepsilon, \varepsilon)\, d\varepsilon + \hat{S}_t(\theta^{t-1}, \eta, \eta). \tag{A-4}$$

The equation (5.7) obtains by taking the limit as $\eta \to \underline{\theta}$ and noting that

$$\frac{\partial}{\partial \theta_t} \hat{S}_t(\theta^{t-1}, \hat{\theta}_t, \theta_t) = U_\theta(X_t(\theta^{t-1}, \hat{\theta}_t)) + \beta W_\theta(X^{\infty}_{t+1}(\theta^{t-1}, \hat{\theta}_t); \theta_t).$$

Proof (**Theorem 5**). Fix $t \geq 1$ and $\theta^{t-1} \in \Theta^{t-1}$. Define $S(\theta^{t-1}, \theta_t) = \hat{S}(\theta^{t-1}, \theta_t; \theta_t)$ where $\hat{S}$ is defined in (A-3). An allocation is temporarily incentive compatible if

$$S(\theta^{t-1}, \theta_t) - \hat{S}(\theta^{t-1}, \hat{\theta}_t; \theta_t) \geq 0. \tag{A-5}$$

By the envelope condition (5.7), $S(\theta^{t-1}, \theta)$ is differentiable for almost all $\theta \in \Theta$, with its derivative $\frac{\partial}{\partial \theta} S(\theta^{t-1}, \theta) = U_\theta\left(X_t(\theta^{t-1}, \theta), \theta\right) + \beta W_\theta\left(X^{\infty}_{t+1}(\theta^{t-1}, \theta); \theta\right)$. Hence, one can write

$$\begin{aligned}
S(\theta^{t-1}, \theta_t) - S(\theta^{t-1}, \hat{\theta}_t) &= \int_{\hat{\theta}}^{\theta} \frac{\partial}{\partial \theta} S(\theta^{t-1}, \varepsilon)\, d\varepsilon \\
&= \int_{\hat{\theta}}^{\theta} \left[U_\theta\left(X_t(\theta^{t-1}, \varepsilon), \varepsilon\right) + \beta W_\theta\left(X^{\infty}_{t+1}(\theta^{t-1}, \varepsilon); \varepsilon\right)\right] d\varepsilon \\
&\geq \int_{\hat{\theta}}^{\theta} \left[U_\theta\left(X_t(\theta^{t-1}, \hat{\theta}_t), \varepsilon\right) + \beta W_\theta\left(X^{\infty}_{t+1}(\theta^{t-1}, \hat{\theta}_t); \varepsilon\right)\right] d\varepsilon \\
&= \hat{S}(\theta^{t-1}, \hat{\theta}_t; \theta_t) - S(\theta^{t-1}, \hat{\theta}_t),
\end{aligned}$$

where the inequality follows from the fact that the expression (5.8) is increasing in $\hat{\theta}_t$, and the last equality follows from the fact that $\hat{S}(\theta^{t-1}, \hat{\theta}_t, \theta)$ is differentiable in θ. Canceling terms, one gets (A-5).

Proof (**Lemma 6**). Fix $\theta \in \Theta$. The set $\mathcal{V}^R(\theta)$ is clearly nonempty because any allocation that is independent of the report is incentive compatible. $\mathcal{V}^R(\theta)$ is bounded because $W(X;\theta)$ is bounded by $K_0 + K_1\theta$ by Assumptions 1-3 and boundedness of $W_\theta(X;\theta)$ follows from Lemma (3).

To prove that $\mathcal{V}^R(\theta)$ is closed, let $\{w_n, g_n\}_{n=1}^{\infty}$ be a Cauchy sequence in $\mathcal{V}^R(\theta)$. Denote its limit by (w, g). Let also $\{X_n\}_{n=1}^{\infty}$ be a sequence of allocations such that $w_n = W(X_n;\theta)$ and $g_n = W_\theta(X_n;\theta)$. Since $\mathcal{X}$ is compact, X_n contains a convergent subsequence $\{X_{n_k}\}_{k=1}^{\infty}$. Let X_∞ be its limit. Because $\mathbb{X}$ is compact by Assumption 1 and a limit of a sequence of a measurable functions is measurable, $X_\infty \in \mathcal{X}$. One has

$$w = \lim_{k\to\infty} w_{n_k} = \lim_{k\to\infty} W(X_{n_k};\theta) = W(\lim_{k\to\infty} X_{n_k};\theta) = W(X_\infty;\theta),$$

where the second inequality follows from the fact that X_n supports (w_n, g_n), and the third one follows from affinity of W in X. Similarly,

$$g = \lim_{k\to\infty} g_{n_k} = \lim_{k\to\infty} W_\theta(X_{n_k};\theta) = W_\theta(\lim_{k\to\infty} X_{n_k};\theta) = W_\theta(X_\infty;\theta).$$

Finally, since X_n satisfies the envelope condition (5.7) and U,W,U_θ and W_θ are all continuous in X, X_∞ satisfies (5.7) as well. Thus, X_∞ supports (w, g), which implies that $(w,g) \in \mathcal{V}^R(\theta)$, and so $\mathcal{V}^R(\theta)$ is closed. Since it is both bounded and closed, $\mathcal{V}^R(\theta)$ is compact. Convexity of $\mathcal{V}^R(\theta)$ follows from the fact that the utility is affine in x, and so both $W(X;\theta)$ and $W_\theta(X;\theta)$ are affine in X.

Proof (**Theorem 7**). Let $(w,g) \in \mathcal{V}^R(\theta_-)$. Then, there exists some admissible allocation $X \in \mathcal{X}$ such that $W(X;\theta) = w$ and $W_\theta(X;\theta) = g$. Define the allocation rule $z = (x,v,h)$ by $x(\theta) = X_1(\theta)$, $v(\theta) = W(X_1^\infty(\theta);\theta)$ and $h(\theta) = W_\theta(X_1^\infty(\theta);\theta)$. Since X is measurable and the utility is affine in x, z is measurable as well, and so $z \in \mathcal{Z}$. Since $X_1^\infty(\theta)$ is an allocation and satisfies the envelope condition (5.7) after all histories, one has $[v(\theta),h(\theta)] \in \mathcal{V}^R(\theta)$ for all $\theta \in \Theta$. The allocation rule z satisfies (6.10) since

$$
\begin{aligned}
U\left(x(\theta),\theta\right) + \beta v(\theta) &= U\left(X_1(\theta)),\theta\right) + \beta W(X_1^\infty;\theta) \\
&= \int_{\underline{\theta}}^{\theta} \left[U_\theta\left(X_1(\varepsilon),\varepsilon\right) + \beta W_\theta(X_1^\infty(\varepsilon);\varepsilon)\right] d\varepsilon + \underline{W}(X_1^\infty) \\
&= \int_{\underline{\theta}}^{\theta} \left[U_\theta\left(x(\varepsilon),\varepsilon\right) + \beta h(\varepsilon)\right] d\varepsilon + \underline{w}.
\end{aligned}
$$

Hence, z is admissible w.r.t. $\mathcal{V}^R$. One can show that $w(z;\theta_-) = w$ and $g(z;\theta_-) = g$. Hence, $(w,g) \in \mathcal{T}\mathcal{V}^R(\theta_-)$ and so $\mathcal{V}^R(\theta_-) \subseteq (\mathcal{T}\mathcal{V}^R)(\theta_-)$. Since θ_- is arbitrary, $\mathcal{V}^R \subseteq \mathcal{T}\mathcal{V}^R$.

To show the reverse implication, suppose that $(w,g) \in \mathcal{T}\mathcal{V}^R(\theta_-)$. Then, there exists some allocation rule z such that (6.10) holds and $[v(\theta),h(\theta)] \in \mathcal{V}^R(\theta)$. Define an allocation X as follows. Let $X_1(\theta) = x(\theta)$. Since $[v(\theta),h(\theta)] \in \mathcal{V}^R(\theta)$, there is for each $\theta \in \Theta$ some allocation $\tilde{X}(\theta)$ such that $W(\tilde{X}(\theta);\theta) = v(\theta)$ and $W_\theta(\tilde{X}(\theta);\theta) = v(\theta)$. Define $X_1^\infty(\theta) = \tilde{X}(\theta)$. The allocation X constructed in this way is measurable and so $X \in \mathcal{X}$. It is also easy to see that X satisfies (5.7), and so is admissible, and that $W(X;\theta_-) = w$ and $W_\theta(X;\theta_-) = g$. Thus, $(w,g) \in \mathcal{V}^R(\theta_-)$ and so $\mathcal{T}\mathcal{V}^R(\theta_-) \subseteq \mathcal{V}^R(\theta_-)$. Since θ_- is arbitrary, $\mathcal{T}\mathcal{V}^R \subseteq \mathcal{V}^R$.

Proof (**Theorem 8**). Let $\mathcal{C}(\mathcal{V}^R)$ be a space of bounded and continuous functions $P : \mathcal{V}^R \to \mathbb{R}$, endowed with a sup norm. Let also T be the operator defined by the right-hand side of the Bellman equation (6.15). The set $\mathcal{C}(\mathcal{V}^R)$ is a complete metric space.

The correspondence γ^R is clearly nonempty and continuous in (w,g). Lemma (2) implies that both $w(z;\theta)$ and $g(z;\theta)$ are continuous in θ. Hence, $\gamma^R(w,g,\theta)$ is continuous in θ as well. Since $\mathbb{X}$ is compact by Assumption 1 and $\mathcal{V}^R$ is compact by Lemma (6), γ^R is compact valued.

Suppose now that $P \in \mathcal{C}(\mathcal{V}^R)$. Define a function $\hat{P} : \mathcal{V}^R \times \mathcal{Z}(\mathcal{V}^R) \to \mathbb{R}$ by

$$
\hat{P}(w,g,\theta_-,z) = \int_\Theta \left[R\left(x(\theta)\right) + qP\left(v(\theta),h(\theta),\theta\right)\right] dF(\theta|\theta_-).
$$

Since R is continuous by Assumption 4 and F is continuous in θ_- by Assumption 6, $\hat{P}$ is continuous in (w,g,θ_-,z). Berge's Theorem (Aliprantis and Border (2005), Theorem 17.31) then implies that T maps $\mathcal{C}(\mathcal{V}^R)$ on itself. Since $q < 1$, T is a contraction, and it follows from the contraction mapping theorem (Stokey et al. (1989) Theorem 3.2) that there is a unique fixed point of the operator T within $\mathcal{C}(\mathcal{V}^R)$.

Since both $\mathbb{X}$ and $\mathcal{V}^R$ are compact, $\mathcal{Z}(\mathcal{V}^R)$ is compact as well. Since Γ^R is a closed subset of $\mathcal{V}^R \times \mathcal{Z}(\mathcal{V}^R)$, by the measurable selection theorem (e.g. Proposition 7.33 of Bertsekas and Shreve (1978)), there exists a measurable selection from the efficient allocation rule.

The utility is affine in x, and so γ^R is convex in z. Since γ^R is affine in (w,g) for each $\theta \in \Theta$, it is convex for each $\theta \in \Theta$. The objective function R is convex in x. Convexity of P then follows from standard arguments. If R is strictly convex then P is strictly convex by Corollary 1 of Stokey et al. (1989). Since the objective function is strictly convex, the efficient allocation rule is essentially unique, and so must be measurable.

Proof (**Lemma 9**). *Proof of (if).* Suppose that $X \in \Gamma^R(w,g,\theta_0)$. Define $V_2(\theta_1) = W(X_1^\infty(\theta_1);\theta_1)$ and $H_2(\theta_1) = W_\theta(X_1^\infty(\theta_1);\theta_1)$. Since X is measurable and W is affine, $Z_1(\cdot) \equiv [X_1(\cdot), V_2(\cdot), H_2(\cdot)]$ is measurable as well. Hence, Z_1 is an allocation rule. Since $X \in \Gamma^R(w,g,\theta_0)$, it satisfies

$$w = \int_\Theta [U(X_1(\theta_1),\theta_1) + \beta W(X_1^\infty;\theta_1)]dF(\theta_1|\theta_0) = \int_\Theta [U(X_1(\theta_1),\theta_1) + \beta V_2(\theta_1)]dF(\theta_1|\theta_0)$$
$$\equiv w(Z_1;\theta_0),$$

where the second equality uses the definition of V_2. Similarly, one can show that $g = g(Z_1;\theta_0)$. Since $X_1^\infty(\theta)$ is an allocation and supports $[V_2(\theta_1), H_2(\theta_1)]$ at θ_1 by construction, $[V_2(\theta_1), H_2(\theta_1)] \in \mathcal{V}^R(\theta_1)$. Since X satisfies the envelope condition (5.7) at $t=1$, Z_1 satisfies the envelope condition (6.10). Hence, Z_1 is admissible w.r.t. $\mathcal{V}^R$ and (6.16) holds. Property (6.17) can be shown in a similar fashion.

Proof of (only if). If (6.16) and (6.17) hold, $V_t(\theta^{t-1})$ satisfies for all $t \geq 1$ and all $\theta^{t-1} \in \Theta^{t-1}$

$$V_t(\theta^{t-1}) = \int_\Theta [U\left(X_t(\theta^{t-1},\theta_t),\theta_t\right) + \beta V_{t+1}(\theta^{t-1},\theta_t)]\, dF(\theta_t|\theta_{t-1}), \tag{A-6}$$

where $V_0(\varnothing) \equiv w$. Recursively eliminating V_{t+s} on the right-hand side of (A-6) $T-$ times, taking the limit as $T \to \infty$ and noting that since $\left(V_{t+T}(\theta^{t-1},\theta^T), H_{t+T}(\theta^{t-1},\theta^T)\right) \in \mathcal{V}^R(\theta_T)$ and $\mathcal{V}^R(\theta_T)$ is compact, one gets

$$\begin{aligned} V_t(\theta^{t-1}) &= \sum_{s=1}^\infty \beta^{s-1} \int_{\Theta^s} U\left(X_{t+s-1}(\theta^{t-1},\theta^s),\theta_s\right)\, dF(\theta_{t+s-1}|\theta_{t+s-2})\dots dF(\theta_t|\theta_{t-1}) \\ &= \sum_{s=1}^\infty \beta^{s-1} \int_{\Theta^s} U\left(X_{t+s-1}(\theta^{t-1},\theta^s),\theta_s\right)\, d\mu^{t+s}(\theta^{t-1},\theta^s|\theta_0) \\ &\equiv W(X_t^\infty(\theta^{t-1});\theta_{t-1}) \end{aligned} \tag{A-7}$$

where the second equality follows from the definition of μ^t. Similarly,

$$H_t(\theta^{t-1}) = W_\theta(X_t^\infty(\theta^{t-1});\theta_{t-1}), \tag{A-8}$$

where $H_0(\varnothing) \equiv g$. Evaluating (A-7) and (A-8) at $t=0$ yields $w = W(X;\theta_0)$ and $g = W_\theta(X;\theta_0)$. Since (6.16) holds, $[X_1(\cdot), V_2(\cdot), H_2(\cdot)]$ satisfies the envelope condition (6.10). Because $V_2(\theta_1) = W(X_2^\infty(\theta_1);\theta_1)$ and $H_2(\theta_1) = W_\theta(X_2^\infty(\theta_1);\theta_1)$ by equations (A-7) and (A-8), X satisfies the envelope condition (5.7) at $t=1$. Analogous arguments show that X satisfies the envelope condition (5.7) after any history. Hence $X \in \Gamma^R(w,g,\theta_0)$.

Proof (**Theorem 10**). *Proof of i.* Take any $[X_1(\cdot), V_2(\cdot), H_2(\cdot)] \in \gamma^R(w,g,\theta_0)$. The value function P satisfies

$$P(w,g,\theta_0) = \min_{z\in\gamma^R(w,g,\theta_0)} \int_\Theta \left[R\left(x(\theta_1)\right) + qP\left(v(\theta_1), h(\theta_1), \theta_1\right)\right] dF(\theta_1|\theta_0)$$
$$\leq \int_\Theta \left[R\left(X_1(\theta_1)\right) + qP\left(V_2(\theta_1), H_2(\theta_1), \theta_1\right)\right] dF(\theta_1|\theta_0). \quad \text{(A-9)}$$

By the same argument, P satisfies for any $\left[X_t(\theta^{t-1},\cdot), V_{t+1}(\theta^{t-1},\cdot), H_{t+1}(\theta^{t-1},\cdot)\right] \in \gamma^R\left(V_t(\theta^{t-1}), H_t(\theta^{t-1}), \theta_{t-1}\right)$

$$P\left(V_t(\theta^{t-1}), H_t(\theta^{t-1}), \theta_{t-1}\right) \leq \int_\Theta \left[R\left(X_t(\theta^t)\right) + qP\left(V_{t+1}(\theta^t), H_2(\theta^t), \theta_t\right)\right] dF(\theta_t|\theta_{t-1}). \quad \text{(A-10)}$$

Eliminating the value function from the right-hand side of (A-9) $T-$times using (A-10), one gets

$$P(w,g,\theta_0) \leq \sum_{t=1}^{T} q^{t-1} \int_{\Theta^t} R\left(X_t(\theta^t)\right)\, dF(\theta_t|\theta_{t-1})\ldots dF(\theta_1|\theta_0)$$
$$+\, q^T \int_{\Theta^T} P\left(V_{T+1}(\theta^T), H_{T+1}(\theta^T), \theta_T\right)\, dF(\theta_T|\theta_{T-1})\ldots dF(\theta_1|\theta_0). \quad \text{(A-11)}$$

Taking the limit for $T\to\infty$ and noting that P is bounded by Lemma 8, one obtains

$$P(w,g,\theta_0) \leq \sum_{t=1}^{\infty} q^{t-1} \int_{\Theta^t} R\left(X_t(\theta^t)\right)\, dF(\theta_t|\theta_{t-1})\ldots dF(\theta_1|\theta_0)$$
$$= \sum_{t=1}^{\infty} q^{t-1} \int_{\Theta^t} R\left(X_t(\theta^t)\right)\, d\mu^t(\theta^t|\theta_0) \quad \text{(A-12)}$$
$$\equiv D(X;\theta_0), \quad \text{(A-13)}$$

where the second equality follows from the definition of μ^t. By Lemma 9, any $X \in \Gamma^R(w,g,\theta_0)$ can be constructed in this way. Hence $P \leq P^*$.

To show the opposite inequality, let z^* be the efficient allocation rule, and X be the allocation that is generated by z^*. Then, the argument above can be repeated with equality at every step. Since X may not be second stage efficient, $P^* \leq P$. Combining, one gets $P = P^*$.

Proof of ii. By previous arguments, if X is generated by the efficient allocation rule z^* then $P^*(w,g,\theta_0) = P(w,g,\theta_0) = D(X;\theta_0)$. Hence, X solves the social planner's problem (6.13) and so is second stage efficient.

Proof (**Theorem 11**). Let X be an allocation generated by the efficient allocation rule z^*. To reduce notation, suppress the dependence of X on w and g. Then, for any $t \geq 1$ and any $\theta^{t-1} \in \Theta^{t-1}$,

$$U\left(X_t(\theta^{t-1}, \hat{\theta}_t), \theta_t\right) = U\left(x^*\left(V_t(\theta^{t-1}), H_t(\theta^{t-1}), \hat{\theta}_t\right), \theta_t\right).$$

Lemma (9) implies that $W\left(X_t^\infty(\theta^{t-1}), \theta_{t-1}\right) = V_t(\theta^{t-1})$. The continuation utility can then be shown to satisfy

$$W\left(X_{t+1}^\infty(\theta^{t-1}, \hat{\theta}_t); \theta_t\right) = \hat{v}\left(V_t(\theta^{t-1}), H_t(\theta^{t-1}), \theta_{t-1}, \hat{\theta}_t; \theta_t\right).$$

Combining, one obtains

$$U\left(X_t(\theta^{t-1},\hat\theta_t),\theta_t\right)+\beta W\left(X_{t+1}^\infty(\theta^{t-1},\hat\theta_t);\theta_t\right)$$
$$=U\left(x^*\left(V_t(\theta^{t-1}),H_t(\theta^{t-1}),\theta_{t-1},\hat\theta_t\right),\theta_t\right)+\beta\hat v\left(V_t(\theta^{t-1}),H_t(\theta^{t-1}),\theta_{t-1},\hat\theta_t;\theta_t\right).$$

If (6.21) holds for any $(w,g,\theta_-)\in\mathcal{V}^R(\theta_-)$, then

$$\theta\in\arg\max_{\hat\theta\in\Theta}U\left(X_t(\theta^{t-1},\hat\theta_t),\theta_t\right)+\beta W\left(X_{t+1}^\infty(\theta^{t-1},\hat\theta_t);\theta_t\right).$$

The allocation X therefore satisfies the temporary incentive compatibility constraint (3.2) and solves the social planner's problem (3.4). Hence $P^{IC}=P^R$.

Proof (**Lemma 12**). Let $\hat s(w,g,\theta_-,\hat\theta;\theta)=U\left(x^*(w,g,\theta_-,\hat\theta),\theta\right)+\beta\hat v(w,g,\theta_-,\hat\theta;\theta)$. The relaxed problem id valid if

$$s^*(w,g,\theta_-,\theta)-\hat s(w,g,\theta_-,\hat\theta;\theta)\geq 0. \tag{A-14}$$

By the envelope condition (6.10), s^* is differentiable for almost all $\theta\in\Theta$, with its derivative $\frac{\partial}{\partial\theta}s^*(w,g,\theta_-,\theta)=U_\theta\left(x^*(w,g,\theta_-,\theta)\right)+\beta h^*(w,g,\theta_-,\theta)$. Hence, one can write

$$\begin{aligned}s^*(w,g,\theta_-,\theta)-s^*(w,g,\theta_-,\hat\theta)&=\int_{\hat\theta}^{\theta}\frac{\partial}{\partial\theta}s^*(w,g,\theta_-,\varepsilon)\,d\varepsilon\\&=\int_{\hat\theta}^{\theta}\left[U_\theta\left(x^*(w,g,\theta_-,\varepsilon),\varepsilon\right)+\beta h^*(w,g,\theta_-,\varepsilon)\right]d\varepsilon\\&\geq\int_{\hat\theta}^{\theta}\left[U_\theta\left(x^*(w,g,\theta_-,\hat\theta),\varepsilon\right)+\beta\hat h(w,g,\theta_-,\hat\theta;\varepsilon)\right]d\varepsilon\\&=\hat s(w,g,\theta_-,\hat\theta;\theta)-s^*(w,g,\theta_-,\hat\theta),\end{aligned}$$

where the inequality follows from the fact that the expression (6.22) is increasing in $\hat\theta$, and the last equality follows from the fact that $\hat s$ is differentiable in θ. Canceling terms, one gets (A-14).

Acknowledgements. I would like to thank the editor, Kjetil Storesletten, four anonymous referees for helpful suggestions. Audiences at Stanford, MIT, UCLA, Penn State, University of Iowa, University of Tokyo, UCSB, CERGE-EI, Mannheim, University of British Columbia and many conferences offered helpful advice. All errors are mine.

REFERENCES

Abraham, A. and N. Pavoni (2008). Efficient allocations with moral hazard and hidden borrowing and lending: A recursive formulation. *Review of Economic Dynamics 11*, 781–803.

Abreu, D., D. Pearce, and E. Stacchetti (1990). Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica 58*(5), 1041–1063.

Ales, L. and P. Maziero (2009). Accounting for private information. Working paper, University of Minnesota.

Aliprantis, C. D. and K. C. Border (2005). *Infinite Dimensional Analysis: A Hitchhiker's Guide* (3 ed.). Berlin, Germany: Springer-Verlag.

Atkeson, A. and R. E. Lucas, Jr. (1992). On efficient distribution with private information. *The Review of Economic Studies 59*, 427–453.

Battaglini, M. and S. Coate (2008). Pareto efficient income taxation with stochastic abilities. *Journal of Public Economics 92*(3-4), 844–868.

Bertsekas, D. P. and S. E. Shreve (1978). *Stochastic Optimal Control: The Discrete Time Case*. New York, N. Y.: Academic Press.

Chang, R. (1998). Credible monetary policy in an infinite horizon model: Recursive approaches. *Journal of Economic Theory 81*(2), 431–461.

Courty, P. and H. Li (2000). Sequential screening. *The Review of Economic Studies 67*(4), 697–717.

Diamond, P. A. (1998). Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates. *American Economic Review 88*, 83–95.

Doepke, M. and R. M. Townsend (2006). Dynamic mechanism design with hidden income and hidden actions. *Journal of Economic Theory 126*, 236–286.

Eso, P. and B. Szentes (2007). Optimal information disclosure in auctions and the handicap auction. *The Review of Economic Studies 74*, 705–731.

Farhi, E. and I. Werning (2010). Insurance and taxation over the life-cycle. Working paper, MIT.

Fernandes, A. and C. Phelan (2000). A recursive formulation for repeated agency with history dependence. *Journal of Economic Theory 91*(2), 223–247.

Fukushima, K. and Y. Waki (2011). Computing dynamic optimal mechanisms when hidden types are markov. Working paper, University of Minnesota.

Golosov, M., A. Tsyvinski, and M. Troshkin (2010). Optimal dynamic taxes. Working paper, Yale University.

Green, E. J. (1987). Lending and the smoothing of uninsurable income. In E. Prescott and N. Wallace (Eds.), *Contractual Arrangements for Intertemporal Trade*. Minneapolis: University of Minnesota Press.

Heathcote, J., K. Storesletten, and G. Violante (2009). Consumption and labor supply with partial insurance: An analytical framework. Working paper, Federal Reserve Bank of Minneapolis.

Heathcote, J., K. Storesletten, and G. Violante (2011). From wages to welfare: Decomposing gains and losses from rising inequality. Working paper, Federal Reserve Bank of Minneapolis.

Heathcote, J., K. Storesletten, and G. L. Violante (2005). Two views of inequality over the life cycle. *Journal of the European Economic Association 3*(2/3), pp. 765–775.

Jarque, A. (2010). Repeated moral hazard with effort persistence. *Journal of Economic Theory 145*, 2412–2423.

Kydland, F. E. and E. C. Prescott (1980). Dynamic optimal taxation, rational expectations and optimal control. *Journal of Economic Dynamics and Control 2*(0), 79 − 91.

Milgrom, P. and I. Segal (2002). Envelope theorems for arbitrary choice sets. *Econometrica 70*, 583–601.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies 38*, 175–208.

Pavan, A., I. Segal, and Y. Toikka (2009). Dynamic mechanism design: Revenue equivalence, profit maximization, and information disclosure. Working paper, Stanford University.

Phelan, C. and E. Stacchetti (1999). Sequential equilibria in a ramsey tax model. *Econometrica 69*(6), 1491–1518.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *The Review of Economic Studies 68*, 205–229.

Stokey, N. L., R. E. Lucas, Jr., and E. C. Prescott (1989). *Recursive Methods in Economic Dynamics*. Cambridge: Harvard University Press.

Tchistyi, A. (2006). Security design with correlated hidden cash flows: The optimality of performance pricing. Working paper, New York University.

Thomas, J. and T. Worrall (1990). Income fluctuations and asymmetric information: An example of the repeated principal agent problem. *Journal of Economic Theory 51*, 367–390.

Werning, I. (2001). Moral hazard with unobserved endowments: A recursive approach. Working paper, University of Chicago.

Williams, N. (2011). Persistent private information. *Econometrica 79*(4), 1233–1274.

Zhang, Y. (2009). Dynamic contracting, persistent shocks and optimal taxation. *Journal of Economic Theory 144*, 635–675.