Misery, Persistence and Growth *

Alexander W. Bloedel[†] R. Vijay Krishna[‡]

30th April 2017

Abstract

We study dynamic insurance contracts when the agent has Markovian and persistent private information. The optimal contract leads to the same long-run immiseration result from classic iid models. While this suggests that the distributional implications of private information are robust to details of the *information structure*, we also show that they are sensitive to the underlying *technology*. When aggregate resources grow over time in an Atkeson and Lucas (1992) economy, *relative* but not *absolute* immiseration occurs: growth saves all agents from misery, but incentive provision requires inequality to increase without bound. Methodologically, we show how to analyze the Markovian problem recursively and identify the appropriate martingale to characterize contractual dynamics. These techniques allow us to highlight allocative distortions unique to the Markovian setting and provide a unified perspective on the role of persistence and monetary transfers in dynamic contracting problems.

JEL Numbers: D31; D80; D82; D86

Keywords: Relative and absolute immiseration; dynamic contracting; recursive contracts; principal-agent problem; persistent private information; growth

1. Introduction

Problems of insurance and distribution are inextricably linked in modern economies. Financial markets, tax systems and social insurance programs all serve, at least in part, to facilitate risk-sharing and thereby protect against the many ups and downs of economic life, such as spells of unemployment, unexpected changes in health and productivity, acute liquidity needs,

(*) We thank Gabriel Carroll, V.V. Chari, Sebastian Di Tella, Ed Green, Pablo Kurlat, Paul Milgrom, Shunya Noda, Ilya Segal, Andy Skrzypacz, and seminar participants at Stanford for useful comments and conversations.

(†) Stanford University. Email: abloedel@stanford.edu

(‡) Florida State University. Email: rvk3570@gmail.com

and other sources of income fluctuations. These institutions are also at the center of ongoing debates in academic and political spheres over the sources and consequences of growing income and wealth inequality in developed economies, and the United States in particular. For instance, according to a recent Pew Research Center[1] survey, Europeans and Americans focus on inequality as the world's greatest danger. Lucas (1992) notes that "…the idea that a society's income distribution arises, in large part, from the way it deals with individual risks is a very old and fundamental one, one that is at least implicit in all modern studies of distribution." While it is therefore no surprise that economists have devoted considerable attention to the question optimal insurance mechanisms, the main distributional finding of these studies is striking. In canonical settings, the optimal insurance mechanism results in *immiseration*, wherein the insured agent becomes completely impoverished in the long run.

Most theoretical studies of dynamic risk sharing focus on the case of iid shocks. However, the data suggest that the relevant risks — such as shocks to earnings, employment, and health — are not only inherently *dynamic*, but also *highly persistent*; see, for instance, Storesletten, Telmer and Yaron (2004a), Storesletten, Telmer and Yaron (2004b), and Meghir and Pistaferri (2004).[2] We, therefore, study a model dynamic risk sharing with endowment shocks that are arbitrarily (but imperfectly) persistent. We make three contributions: two economic and one methodological. Our economic results revisit the classic immiseration result from the individual- and aggregate-level perspectives. Our methodological contribution is a tractable recursive formulation of the dynamic contracting problem with Markovian private information.

Our first contribution is in formulating a tractable recursive model for the analysis of dynamic risk sharing with Markovian shocks and analytically characterizing the optimal contract. We study a Markovian version of the model in Thomas and Worrall (1990), wherein a risk-neutral principal designs an optimal insurance contract for a single risk-averse agent who faces unobservable, persistent endowment shocks. Our Theorem 4 shows that immiseration occurs for arbitrary (but imperfect) degrees of serial correlation in the shock process. Importantly, our results suggest that the long-run distributional implications of private information are robust to details of the *information structure*. Nonetheless, as shown in Proposition 5.1, persistence does introduce a qualitatively different type of short-run allocative distortion than the distortion present in classic iid models.

Our second contribution is to show that the long-run implications of private information are *not*, however, robust to variations of the underlying *technological assumptions* of the classic model. In particular, we formulate a "general equilibrium" version of the principal-agent problem as in the seminal work of Atkeson and Lucas (1992), but allow aggregate resources to *grow* over time. When resources grow at a constant, positive rate, we show in Proposition 7.1

(1) See `http://www.pewglobal.org/2014/10/16/middle-easterners-see-religious-and-ethnic-hat-red-as-top-global-threat/`.

(2) Storesletten, Telmer and Yaron (2004a) find that labor earnings approximately follow a random walk. Storesletten, Telmer and Yaron (2004b) and Meghir and Pistaferri (2004) emphasize time-varying risk of labor income, which is inconsistent with the iid model.

that *relative*, but not *absolute*, immiseration occurs in this class of economies. To understand what we mean by relative and absolute immiseration, note that, in large-economy settings, the classic immiseration result can be stated either in terms of *levels* or *distributions*. In the level interpretation, incentive provision requires that (almost) everyone in society becomes impoverished — or, as we say, *absolutely immiserated*. In the distributional interpretation, immiseration says that the income and wealth distributions become infinitely unequal — a phenomenon we refer to as *relative immiseration*. Thus, our result demonstrates that inequality, but not poverty, is the primary long-run prediction of private information frictions in macroeconomic settings. This result also suggests that the tradeoff between insurance and incentives operates differently when the rate of aggregate growth is endogenous. For example, even a Rawlsian planner, who is concerned with the worst-off agent in the population, may optimally choose an allocation that results in relative, but not absolute, immiseration when the cost of redistribution is a reduction of the growth rate.

Finally, our methodological contribution is to provide a set of tools for analyzing dynamic contracting models with persistent private information and discrete types, which are applicable well beyond our model of risk sharing. We develop recursive methods to analyze these problems in discrete time and identify the appropriate martingale for our Markovian environment — which has the virtue of being familiar from the iid benchmark — to characterize the dynamics of the optimal contract. These martingale techniques are especially useful in settings where explicit solutions are not available, and allow us to prove relatively more analytical results than in much of the extant literature. The recursive approach also clarifies connections to the literatures on dynamic mechanism design and static multidimensional screening, as we discuss in Section 6.

The rest of the paper is organized as follows. After discussing connections to the literature, we lay out the basic model in Section 2 and describe the recursive contracting problem with Markovian shocks in Section 3. Section 4 contains an analysis of the optimal contract's long-run dynamics and presents our main immiseration result, while short-run dynamics are characterized in Section 5 (and in more detail in Appendix H). Sections 6 and 7 conclude with a discussion of the Markovian model and the growth model, respectively. All proofs are contained in the appendices.

Related Literature

Our paper contributes to three literatures: dynamic contracting for insurance, dynamic agency with persistent private information, and dynamic mechanism design.

Dynamic contracting for insurance: Green (1987) and Thomas and Worrall (1990) develop the recursive approach to dynamic screening using ex ante promised utility as a state variable when the agent's private information iid over time. Thomas and Worrall (1990) identify the relevant martingale — namely, the derivative of the principal's value function — to characterize the optimal contract's long-run dynamics, and we build on their important contributions here. Both papers show that in dynamic insurance problems with iid shocks and full commitment on both sides, the optimal contract leads to immiseration.

In a classic paper, Atkeson and Lucas (1992) study a large-economy "general equilibrium" version of this contracting problem and show that the cross-sectional distribution of promised utility fans out over time, so that inequality increases without bound. Their model features a fixed, aggregate resource constraint, which implies that consumption decreases to its lower bound for almost every agent in the economy. More recently, Farhi and Werning (2007) show that if the social planner places geometrically-declining Pareto weights directly on "future generations" (and hence is more patient that the agents in the model), then immiseration is overturned and the optimal allocation exhibits a kind of social mobility.[3]

Our main innovation relative to this literature is to introduce persistent private information into the canonical Thomas and Worrall (1990) model and characterize the optimal contract in this more general setting.[4] A secondary contribution is to clarify the role of the technological assumptions underlying the immiseration result of Atkeson and Lucas (1992), just as Farhi and Werning (2007) clarify the role of the relative patience of the planner and the agents in the economy.

Dynamic agency with persistent private information: From a methodological standpoint, our paper is closely related to the growing literature on dynamic principal-agent problems with persistent private information. The three closest papers are Fernandes and Phelan (2000), Doepke and Townsend (2006), and Zhang (2009).[5]

Fernandes and Phelan (2000) emphasize the recursive approach to these problems and use as state variables promised utility and *threat-point* utility, both of which are *ex ante* quantities. Threat-point utility is the continuation utility promised to the agent if he lies about his current type, and as such never arises on-path. In contrast, our *interim* state variables (promised utilities contingent on type) are both on-path quantities and, as we discuss, lend themselves to a somewhat more natural interpretation.[6] Doepke and Townsend (2006) develop

(3) Similarly, Atkeson and Lucas (1995) directly impose a lower bound on the agent's promised utility at each date, and Phelan (2006) studies a closely related moral hazard problem with equal Pareto weight placed on each generation; both papers obtain similar distributional results.

(4) See Kocherlakota (2010, Chapter 6) and Golosov, Tsyvinski and Werquin (2016) for discussions of the technical challenges of models with persistent private information. Golosov and Tsyvinski (2006) study a taxation model with a special kind of persistent private information, namely, the agent's type process has an absorbing state. In independent and contemporaneous work, Broer, Kapička and Klein (2017) take a similar approach (with a slight difference in timing) to a related insurance model with limited enforcement and binary types, though they focus on numerical experiments.

(5) See also Halac and Yared (2014), which studies a dynamic delegation problem with Markovian private information and a time-inconsistent agent, using the recursive formulation of Fernandes and Phelan (2000).

(6) Of course, the formulations are equivalent in the sense that there is a one-to-one mapping between them.

similar techniques, with an eye toward more general hidden action problems and computational efficiency. Our primary contribution relative to these papers is a fairly complete analytical characterization of the recursive domain and optimal contract in our canonical insurance setting, while they restrict attention to numerical experiments in particular parametric examples. From a technical standpoint, the fact that our problem is unbounded requires different and occasionally subtle mathematical arguments.

Zhang (2009) develops a continuous-time version of the Fernandes-Phelan model with two shock types. Some of our results about short-run dynamics have analogues in Zhang's work, but the contracting environments are different. He studies a dynamic taxation model in which the inverse Euler equation holds, while it does not hold in our model.[7] Also, Zhang (2009) restricts attention to the non-generic case of Markov processes over two types with symmetric transition probabilities, whereas our results hold for *all* persistent Markov processes with finite state spaces. This illustrates an advantage of our approach, whereby we are able to derive substantive and more general economic results in a simpler and more familiar discrete time setting.

Also related are Kapička (2013) and Williams (2011) which, in discrete and continuous time, respectively, develop the first-order approach for analyzing Markovian contracting problems with a continuum of types. Farhi and Werning (2013) apply similar ideas to a Mirleesian taxation problem. Fu and Krishna (2014) use the same techniques as in this paper to study a firm financing problem with Markovian shocks. Their model features a risk-neutral agent and monetary transfers subject to a limited liability constraint, in contrast to our focus on insurance.

Dynamic mechanism design: This literature primarily focuses on settings with quasi-linear utility and asymmetric information at the time of contracting. Our model differs in both respects, though the recursive characterization is independent of these features. Pavan, Segal and Toikka (2014) study mechanism design in general dynamic environments; their approach to incentive compatibility in Markovian environments is the continuous-type analogue of ours. Closer to our paper, Battaglini (2005) analyzes a monopolistic screening problem with two types and Markovian shocks.[8] Neither of these papers emphasize the recursive approach.

More recently, the mechanism design literature has begun to move beyond the quasi-linear case (see Pavan (2016) for a recent survey). Two such papers have especially close links to

(7) The inverse Euler equation is a special case of our martingale condition (see Theorem 3), and would emerge under appropriate separability assumptions. Our setting does not satisfy the requisite separability conditions because the principal only has a single instrument, namely, consumption utiles. More precisely, in many moral hazard and public finance settings, the agent's effort or production cost is additively separable from his consumption utility. The goal is then to implement a given *allocation rule* for effort/production. The inverse Euler equation emerges as an optimal cost-minimization condition for the principal's *payment rule* denominated in consumption utiles.

(8) When there are more than two types, Battaglini and Lamba (2014) show that non-local incentive constraints may be binding, making the optimal mechanism much harder to characterize for reasons similar to those familiar from static multidimensional screening problems, which are surveyed in Rochet and Stole (2003).

the present work; both focus on the case in which the agent's private information follows a two-state Markov chain.

Farinha Luz (2015) studies dynamic insurance contracts but, unlike us, focuses on screening of *ex ante* private information and assumes that the agent's realized income in each period is observable and can be contracted upon. These differences lead to an optimal contract with markedly different dynamics and, in fact, features a pattern of vanishing distortions reminiscent of Battaglini (2005). From a technical perspective, he studies the sequence formulation of the principal's problem and does not use recursive methods.

Guo and Hörner (2014) study single-agent dynamic mechanism design without money and, independently, take a recursive approach that uses contingent utilities. Their setting is similar to ours in that the principal controls *only* the agent's consumption, so that all incentives must be provided through allocative distortions. However, their economic results and technical arguments are quite different, as they study the repeated allocation of an indivisible good to a risk-neutral agent. Consistent with Phelan (1998), their optimal contract leads to both immiseration and bliss as possible long-run outcomes. More importantly, they do not uncover the appropriate generalization of the Thomas and Worrall (1990) martingale for the Markovian environment, which is at the heart of our analysis.

2. Environment

A single risk-averse agent (he) faces an uncertain endowment stream, and a risk-neutral principal (she) is prepared to provide insurance. Time is discrete and infinite, indexed by $t \in \{0, 1, 2, \dots\}$, and both parties discount the future at the rate $\alpha \in (0, 1)$.

The agent receives an endowment of $\omega_s \in \mathbb{R}$ in each period, where $s \in S := \{1, 2\}$ and $\omega_1 < \omega_2$. The endowment shocks follow an exogenous first-order Markov process with transition probabilities

$$\mathbb{P}(s^{(t+1)} = j \mid s^{(t)} = i) = f_{ij}$$

We assume that $f_{ij} \in (0, 1)$ for all $i, j \in S$ so that the transition matrix is connected and, in particular, there are no absorbing states. We also assume that the Markov process over endowments is *persistent*, ie, positively correlated, which in our setting amounts to requiring $f_{11} \geq f_{21}$. Indeed, we set $\beta := f_{11} - f_{21} = f_{22} - f_{12} \geq 0$, so that β measures the degree of persistence, where $\beta = 0$ corresponds to the iid case.

The principal does not observe these endowment shocks and must rely on the agent's reports. By the Revelation Principle, it is without loss to restrict attention to contracts that elicit truthful reports. We will assume throughout that there is *no hidden borrowing*, so that the agent cannot borrow the consumption good in order to pretend his endowment is greater than it actually is.[9] For example, before receiving any transfers from the principal, the agent

(9) This assumption appears in the literature in various forms — see Williams (2011) for one example. Thomas

might be required to deposit some fraction of his endowment in an account that the principal can monitor. Similarly, we assume that the consumption good depreciates completely after each period so that there are *no hidden savings*.[10] Reporting strategies for the agent are said to be *admissible* if they respect these constraints.

The agent has CARA utility, so consuming y units of the consumption good delivers utility $u(y) = -e^{-y}$. Suppose that $c \in \mathbb{R}$ represents the transfer to the agent net of his endowment, so that the agent's utility in state s is $-e^{-\omega_s - c} = -e^{-\omega_s} e^{-c}$. We may simply define $\theta_s := e^{-\omega_s}$ and rewrite this utility as $\theta u(c)$. Hence, this is equivalent to a model in which the agent experiences *taste* (marginal utility) *shocks* θ_s, where $\theta_1 > \theta_2$. We will refer to $\theta_s \in \Theta := \{\theta_1, \theta_2\}$ and $s \in S$ interchangeably as the agent's *type*.

The restriction to binary types and CARA utility is for expositional simplicity, allowing us to present a fairly complete analytical characterization. As we discuss in Section 6, and show formally in Appendix I, our main results hold more generally.

3. Contracts

The most direct way to formulate the mechanism design problem is in terms of (direct revelation) *sequential contracts*, which specify a process $\tilde{x} = (x^{(t)})_{t \geq 0}$ of flow utility allocations[11] $x^{(t)}$ to the agent that are contingent on the public history (of *reported* types) at each date. These flow allocations are net of the agent's income shock, so that the agent's total flow utility is $\theta^{(t)} x^{(t)}$. The flow cost to the principal in each period is $C(x^{(t)}) := u^{-1}(x^{(t)})$. Given an initial vector of expected lifetime promised utilities $\mathbf{v} := (v_1, v_2) \in \mathbb{R}^2_{--}$, where v_i denotes the lifetime utility promise to an agent whose initial shock $s^{(0)} = i$, a sequential contract $\tilde{x}$ satisfies *promise keeping* constraint if the lifetime utility provided by $\tilde{x}$ is exactly equal to v_i whenever $s^{(0)} = i$. The contract is *incentive compatible* if the truth-telling strategy is optimal for the agent among all admissible reporting strategies. The details of this sequence problem, [**SP**], are standard and laid out in Appendix A. As we discuss below, up to purely technical issues, the sequence problem [**SP**] is equivalent to the recursive problem studied in this paper.

3.1. Recursive Contracts

For purposes of analysis, it is much more convenient to view contract and report choices as Markov Decision Processes (MDPs) with simple state spaces. In iid settings, Green (1987)

and Worrall (1990) show that this assumption is without loss of generality when shocks are iid. Feldman and Slemrod (2007) show that, in U.S. data, under-reporting is the empirically relevant concern.

(10) Allowing for hidden savings would introduce another hidden state variable, in addition to the agent's previous endowment shocks. Problems with hidden savings are complicated even without persistent types; see Williams (2015) and Di Tella and Sannikov (2015) for two of the few examples where analytical characterizations are available.

(11) Thus, there is a corresponding consumption process $\tilde{c} = (c^{(t)})$ such that $c^{(t)} = u^{-1}(x^{(t)})$ for all $t \geq 0$.

and Thomas and Worrall (1990) show that the appropriate state variable for the MDP is the agent's promised utility — ie, the lifetime expected utility starting from the given period that he would obtain if he were truthful in all future periods.

In the Markovian setting, promised utility is not a sufficient state variable because the agent's *true* endowment shock determines both his current marginal utility of consumption *and his preferences over continuation contracts*, so that both aspects of preferences are private information. The principal therefore needs more instruments to screen the agent through continuation contracts. Our formulation of the principal's problem uses the pair $(\mathbf{v}, s)$ of *contingent* promised utilities and yesterday's (reported) shock as state variables. Here, $\mathbf{v} := (v_1, v_2)$ and v_i denotes the utility promised to the agent conditional on reporting shock θ_i today, and s is the previous period's *reported* shock. In this notation, the promise keeping and incentive compatibility constraints are

$$\text{[}\mathbf{PK}_1\text{]} \qquad v_1 = \theta_1 x_1 + \alpha \mathbb{E}^{\mathbf{f}_1} [\mathbf{w}_1]$$

$$\text{[}\mathbf{PK}_2\text{]} \qquad v_2 = \theta_2 x_2 + \alpha \mathbb{E}^{\mathbf{f}_2} [\mathbf{w}_2]$$

$$\text{[}\mathbf{IC}\text{]} \qquad \theta_2 x_2 + \alpha \mathbb{E}^{\mathbf{f}_2} [\mathbf{w}_2] \geq \theta_2 x_1 + \alpha \mathbb{E}^{\mathbf{f}_2} [\mathbf{w}_1]$$

where $\mathbf{f}_i = (f_{i1}, f_{i2})$ and $\mathbf{w}_i = (w_{i1}, w_{i2}) \in \mathbb{R}^2_{--}$ for $i = 1, 2$, and $\mathbb{E}^{\mathbf{f}_i}$ is the expectation induced by the transition probabilities $\mathbf{f}_i$. The tuple $(\mathbf{w}_i, i)$ denotes the (contractual) state that an agent will transition to from $(\mathbf{v}, s)$ if he reports the current shock to be θ_i. Importantly, notice in [IC] that, even if the agent lies today, his expectation over tomorrow's endowment shock is still governed by his true current type.[12] This set of constraints is independent of the previous report s. In this way, the principal can incentivize truthful revelation in the current period regardless of the agent's previous history of actual and reported shocks. Thus, our formulation solves the issue of the agent's private preferences over continuation contracts in our setting.[13]

We need to specify which $\mathbf{w}_i$'s are feasible for the principal to offer to the agent. A subset of $\mathbb{R}^2_-$ is a *domain* for the principal's problem if, for every $\mathbf{v}$ in that set, there is a tuple $(x_i, \mathbf{w}_i)_{i=1,2}$ satisfying [$\mathbf{PK}_1$], [$\mathbf{PK}_2$], and [IC] such that $\mathbf{w}_1$ and $\mathbf{w}_2$ also lie in that set. The principal's value function is defined on the largest such domain, which is characterized in the following theorem.

Theorem 1. *The set* $V := \left\{\mathbf{v} \in \mathbb{R}^2 : v_1 < v_2 < 0\right\}$ *is the largest domain.*

The proof of Theorem 1 is in Appendix B. The domain V is illustrated in Figure 1. Notice that it must be a convex cone because CARA utility makes the constraint set linear in the x's and $\mathbf{w}$'s, which are unbounded below. Interestingly, the domain V is independent of the

(12) Recall that, because the agent cannot borrow the consumption good, the agent with a low endowment can never pretend to have a high endowment. Thus, there is only one incentive constraint that the principal need consider.

(13) This is analogous to the approach of Pavan, Segal and Toikka (2014) in Markovian environments, though they consider a continuum of shock types.

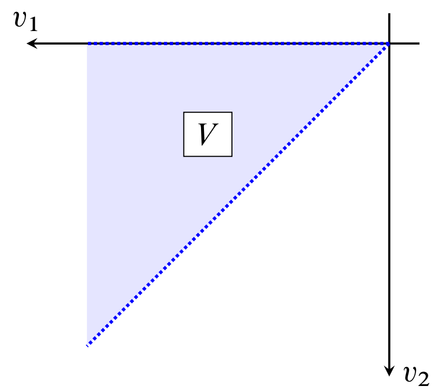

Figure 1: *The recursive domain V*

transition probabilities $\mathbf{f}_i$ and the shock sizes θ_i. Recalling that $\beta \geq 0$ measures the persistence of shocks, note that in the presence of the promise keeping constraints [**PK**$_1$] and [**PK**$_2$], the incentive compatibility constraint [**IC**] is equivalent to

$$[\mathbf{IC}^*] \qquad v_2 - v_1 \geq \underbrace{(\theta_2 - \theta_1)x_1}_{\text{iid info rent}} + \underbrace{\alpha\beta(w_{12} - w_{11})}_{\text{Markov info rent}}$$

One intuition for why Theorem 1 must be true is that both the iid and Markov information rents must be positive, given the agent's informational advantage over the principal.[14]

A *recursive contract* (or simply *contract*) is a map $\xi : V \times S \to \mathbb{R}_{--} \times V$, written as $\xi(\mathbf{v}, s) = (\xi^i(\mathbf{v}, s), \xi^c(\mathbf{v}, s))$, where $\xi^i(\mathbf{v}, s) = x_s(\mathbf{v}) \in \mathbb{R}_{--}$ provides *instantaneous* consumption utiles, and $\xi^c(\mathbf{v}, s) = \mathbf{w}_s(\mathbf{v}) \in V$ provides *contingent continuation* utiles. We say that ξ is *incentive compatible at* $\mathbf{v} \in V$ if $\xi(\mathbf{v}, \cdot) \in \Gamma(\mathbf{v})$, where

$$[\mathbf{3.1}] \qquad \Gamma(\mathbf{v}) := \left\{(x_s, \mathbf{w}_s) \in (\mathbb{R}_{--} \times V)^S : (x_s, \mathbf{w}_s)_{s=1,2} \text{ satisfies } [\mathbf{PK}_1], [\mathbf{PK}_2], [\mathbf{IC}]\right\}$$

is the principal's *feasible set at* $\mathbf{v} \in V$.[15] Naturally, ξ is *incentive compatible* if it is incentive compatible at all $\mathbf{v} \in V$. Let $\Xi(\mathbf{v})$ denote the set of incentive compatible recursive contracts that are initialized at $\mathbf{v} \in V$. Note that every $\mathbf{v} \in V$ and $\xi \in \Xi(\mathbf{v})$ together induce stochastic processes $\tilde{x}_\xi(\mathbf{v}) := \left(x_\xi^{(t)}\right)_{t=0}^\infty$, which we call the *induced allocation*, and $(\mathbf{v}^{(t)})_{t=1}^\infty$, which we call the *induced promises*. (The transition probabilities of these processes are determined by the agent's reporting strategy σ.)

(14) Suppose to the contrary that $v_2 \leq v_1$ at some history. Because the iid information rent is strictly positive and $\alpha\beta \in (0, 1)$, the spread in continuation promises $w_{12} - w_{11}$ must be *strictly more negative* than $v_2 - v_1 \leq 0$. Iterating on [**IC***] in this manner, it is easy to see that the spread in continuation utilities must diverge after any such history. A more formal argument can be found in Appendix B. In particular, we demonstrate in Lemma B.1 that $(0, v_2) \in \mathbb{R}^2_-$ cannot be implemented for any $v_2 < 0$.

(15) Recall that the recursive constraints are independent of the previous report $s \in S$, so that the constraint correspondence $\Gamma(\cdot)$ depends only on $\mathbf{v} \in V$.

3.2. Principal's Recursive Problem

The principal's *recursive problem* is to choose the recursive contract that minimizes the lifetime expected cost of the induced allocation, subject to the recursive constraints at each step:

$$\text{[RP]} \qquad P(\mathbf{v}, s) := \inf_{\xi \in \Xi(\mathbf{v})} \mathbb{E}\left[\sum_{t=0}^{\infty} \alpha^t C\left(x_\xi^{(t)}\right)\right]$$

Note that the expectation is taken with respect to the true probability measure over paths of income shocks, which is the measure over reported paths induced by the agent selecting the truthful reporting strategy. A recursive contract ξ^* is *recursively optimal* if it attains the infimum in [**RP**].[16]

Let $\mathscr{H} := S^\infty$ denote the space of all infinite sequences of income shocks, or paths, with generic element $h \in \mathscr{H}$. We say that ξ satisfies *agent transversality at* $\mathbf{v} \in V$ if, starting from $\mathbf{v}$, the induced discounted promises vanish asymptotically, uniformly across paths, ie, if

$$\text{[TVC]} \qquad \lim_{t\to\infty} \inf_{h \in \mathscr{H}} \alpha^t \mathbf{v}^{(t)}(h) = 0$$

where $(\mathbf{v}^{(t)}(h))_{t=0}^{\infty}$ denotes the (deterministic) sequence of contingent promises along the path $h \in \mathscr{H}$. Any incentive compatible recursive contract ξ that satisfies [**TVC**] is said to be [**TVC**]-*implementable*. This is a continuity condition for the agent's reporting problem, and guarantees that the one-shot deviation principle holds so that truthtelling is a globally optimal strategy for the agent – ie, the agent cannot benefit from strategies involving *infinitely many* misreports, which are not ruled out by the recursive version of incentive compatibility alone.[17] Consequently, if a recursively optimal contract satisfies [**TVC**], we say that it is *globally optimal*.[18]

3.3. Full-Information Benchmark

Before characterizing the optimal contract with private information, it is useful to understand the full-information optimal contract as a benchmark. The optimal contract in this setting is summarized in the following proposition.

(16) Recursive contracts, as we have defined them, are *stationary* in the sense that they do not depend explicitly on the public history or on time. But, as the Bellman equation in Theorem 2 suggests, even if we were to allow for such non-stationarity, it would be without loss to search for *recursively optimal* contracts that are stationary.

(17) It also guarantees that the induced allocation $\tilde{x}_\xi$ actually delivers the lifetime utility promises $\mathbf{v} = (v_1, v_2)$. With unbounded agent utility, this is not guaranteed by the recursive promise-keeping constraints alone.

(18) We show in Appendix G that the space of [**TVC**]-implementable recursive contracts is non-empty by explicitly constructing an incentive compatible recursive contract that satisfies [**TVC**].

Proposition 3.1. There is a unique globally optimal contract under full information. It is stationary and fully insures the agent conditional on his initial type: for each initial state $(\mathbf{v}, s) \in \mathbb{R}^2_{--} \times S$ and current shock $i \in S$,

$$x_i^*(\mathbf{v}, s) = \frac{(1-\alpha)v_s}{\theta_i} \tag{3.2}$$

$$\mathbf{w}_i^*(\mathbf{v}, s) = (v_s, v_s) \tag{3.3}$$

Moreover, the full-information value function $Q^* : \mathbb{R}^2_{--} \times S \to \mathbb{R}$ is strictly convex and strictly increasing in the direction $(1, 1)$.

The proof of Proposition 3.1 is in Appendix C. When the principal is able to observe the agent's endowment shocks in each period, there is no agency friction and it is optimal to provide full insurance. Notice from [3.2] that the optimal contract perfectly smooths the agent's consumption profile, ie, conditional on the initial type, the agent's flow utility $\theta^{(t)}x^{(t)}$ is constant for the rest of time. Graphically, starting from any initial $\mathbf{v} \in \mathbb{R}^2_{--}$, the optimal full-information contract jumps to the lower boundary of V (the diagonal in Figure 1) and then self-generates at a single point. Because we have shown in Theorem 1 that the domain V is open and does not contain the diagonal, it follows immediately that the full-information optimal contract is not incentive compatible. That is, the full-information contract is not implementable when the agent privately observes his endowment shocks.

3.4. Recursive Representation

The main result of this section shows that the principal's value function P satisfies a simple Bellman equation, characterizes its properties, and gives conditions under which the contract generated from the policy functions in fact solves the sequence problem [SP].

We denote the derivative of P at $(\mathbf{v}, s)$ by $\mathsf{D}P(\mathbf{v}, s) = (P_1(\mathbf{v}, s), P_2(\mathbf{v}, s))$, where $P_i(\mathbf{v}, s)$ is the partial derivative with respect to the component v_i. Similarly, the directional derivative in direction $(1, 1)$ is defined by $\mathsf{D}_{(1,1)}P(\mathbf{v}, s) := \lim_{\varepsilon \downarrow 0} [P(\mathbf{v} + (\varepsilon, \varepsilon), s) - P(\mathbf{v}, s)]/\varepsilon$.

Theorem 2. *The principal's value function* $P : V \times S \to \mathbb{R}$ *satisfies the functional equation*

$$P(\mathbf{v}, s) = \min_{(x_i, \mathbf{w}_i) \in \Gamma(\mathbf{v})} \sum_{i=1,2} f_{si}\big[C(x_i) + \alpha P(\mathbf{w}_i, i)\big] \tag{FE}$$

and, for each $s \in S$, $P(\cdot, s)$ *is convex and continuously differentiable. Moreover:*

(a) P *is the pointwise smallest solution to* [FE] *that lies pointwise above* Q^**, the full-information value function.*

(b) P *is strictly increasing in the first component of* $\mathbf{v}$ *and non-monotone in the second component. For any sequence* $\mathbf{v}_n \to \mathbf{v} \in \partial V$*, the boundary of* V*, we have* $P(\mathbf{v}_n, s) \to +\infty$.

(c) *The directional derivative* $\mathsf{D}_{(1,1)}P(\mathbf{v}, s)$ *is non-negative.*

(d) *For any* $a > 0$, P *satisfies the homogeneity property*

$$P(a\mathbf{v}, s) = P(\mathbf{v}, s) - \frac{\log(a)}{1-\alpha}$$

(e) *The partial derivatives are homogenous of degree* -1, *ie, for any* $a > 0$ *and* $i = 1, 2$

$$P_i(a\mathbf{v}, s) = \frac{1}{a} \cdot P_i(\mathbf{v}, s)$$

(f) *The optimal policy exists and is a recursively optimal contract* ξ^*, *which is non-random, independent of the previous report* $s \in S$, *and homogenous of degree* 1 *in* $\mathbf{v}$. *Moreover, the variables* x_2 *and* $\mathbf{w}_2$ *depend on* $\mathbf{v}$ *only through its second component.*

(g) *If* ξ^* *is globally optimal (ie,* ξ^* *satisfies* [**TVC**]*), then the induced allocation* $\tilde{x}_{\xi^*}$ *solves the sequence problem* [**SP**].

The proof of Theorem 2 is in Appendix D. That P satisfies [**FE**] is fairly standard. However, because V is not compact and the return function $C(\cdot)$ is unbounded above and below, we cannot rely on the usual contraction arguments to establish that P is the unique solution to [**FE**]. In any event, part (a) of Theorem 2 shows that it is the smallest solution in an appropriate sense.

The behavior of P near the boundaries of V, as described in part (b), follows from the construction of the domain V. Near the upper boundary $\{\mathbf{v} \in \mathbb{R}_- : v_2 = 0\}$, delivering the contingent promise v_2 becomes very expensive (even in the absence of private information). Near the lower boundary $\{\mathbf{v} \in \mathbb{R}_{--} : v_1 = v_2\}$, it becomes arbitrarily expensive to satisfy [**IC***].[19] The first-best value function Q^* remains finite on this lower boundary because it is derived from the full-information problem.

The homogeneity properties described in parts (d) and (e) follow from the assumption of CARA utility. Their only role is to make the characterization of the recursively optimal contract's short-run dynamics in Section 5 as transparent as possible.

Two properties deserve special comment. First, to understand the non-monotonicity of $P(\cdot, s)$ described in part (b), note that an increase in v_1 tightens both [$\mathbf{PK}_1$] and [**IC***] while leaving the constraint [$\mathbf{PK}_2$] unaffected. This implies that $P(\cdot,\cdot)$ unambiguously increases. On the other hand, an increase in v_2 tightens [$\mathbf{PK}_2$] but simultaneously adds slack to [**IC***]. Depending on the current state $(\mathbf{v}, s)$, loosening the incentive constraint can outweigh the extra cost of promising the agent additional utility, leading to non-monotonicity in v_2.

Second, the independence of the recursively optimal contractual variables on the agent's previous report, as described in part (f), is best understood by observing that the principal's optimization problem in [**FE**] is actually equivalent to two de-coupled problems, each of which conditions on the agent's *current* report.[20] Namely, because we use interim promised utilities,

(19) To see this, notice that both terms on the right hand side of [**IC***] are positive. Thus, if $v_2 - v_1$ is arbitrarily small, it must be that x_1 is also arbitrarily small, which is extremely costly.

(20) See Appendix D.4 for more details.

it follows naturally that the principal's optimization can be carried out at the interim stage. In other words, the interim promised utility vector $\mathbf{v} \in V$ is a "sufficient statistic" for $(\mathbf{v}, s)$ for the purpose of computing the recursively optimal contract. (Of course, the agent's previous report $s \in S$ is important for determining the principal's expected payoffs at each history.) Note that this separation is not possible in the Fernandes and Phelan (2000) approach, which uses ex ante promised utilities as state variables.

Finally, part (g) of Theorem 2 provides the connection between solutions to the principal's recursive problem [**RP**] and the full sequence problem [**SP**]. Because the agent's utility function is unbounded below, the one-shot deviation principle does not necessarily apply to contracts that are not [**TVC**]-implementable. Thus, there is a purely technical gap between [**RP**] (to which Theorem 2 applies) and the full sequence problem [**SP**].[21] While we cannot directly prove existence of a globally optimal contract, we conjecture that the recursively optimal contract described in Theorem 2 is in fact globally optimal, as this is known to be true when shocks are iid or when there is an absorbing state. Thus, we make the following standing assumption:

Assumption 1. There exists a globally optimal contract for [**RP**].

Note that neither Theorem 1 nor Theorem 2 rely on this assumption.[22] But to provide a more complete characterization of the optimal recursive contract, we will need to use some additional properties of the value function that are implied by Assumption 1.

Proposition 3.2. Under Assumption 1, the following hold:

(a) The recursively optimal contract ξ^* from part (f) of Theorem 2 is globally optimal, and is the *unique* optimal contract (in both the recursive and sequential sense). Moreover, it is continuous in $(\mathbf{v}, s)$.
(b) The value functions and (unique) solutions for the recursive problem [**RP**] and the sequence problem [**SP**] are equal.
(c) For each $s \in S$, $P(\cdot, s)$ is *strictly* convex.
(d) For each $s \in S$, the directional derivative $\mathsf{D}_{(1,1)} P(\cdot, s)$ is *strictly* positive.
(e) [**IC**] is always *active* under ξ^*.[23]

(21) It is worth emphasizing that this problem is endemic to all contracting environments that feature an agent with a utility function that is unbounded below. The typical approach for addressing this discrepancy is to solve for the policy functions corresponding to the recursively optimal contract and then verify [**TVC**]-implementability ex post. For examples of this verification in the iid case, see Lemmas 2 and 7 in Green (1987) for CARA utility or Lemma 6.1 in Atkeson and Lucas (1992) for CARA and CRRA utility; both papers essentially solve for the value function in closed form to carry out this step. Thomas and Worrall (1990), who also focus on the recursive problem, do not address this discrepancy. With Markovian shocks and a multidimensional state variable, closed form solutions are not available even with CARA utility, rendering this verification approach untenable.

(22) Also note that the *proof* of Theorem 1, contained in Appendices B and G, establishes that there exists a [**TVC**]-implementable contract for each $\mathbf{v} \in V$.

(23) A constraint is *active* if it holds with equality and if ignoring the constraint renders feasible policies that cost strictly less for the principal.

The proof of Proposition 3.2 is in Appendix D. Part (b) shows that, under Assumption 1, there is no gap between [**RP**] and the sequence problem [**SP**]. In particular, Assumption 1 guarantees that *some* solution of [**RP**] satisfies the constraints of [**SP**] (as delineated formally in Appendix A). Part (a) shows, moreover, that there is a *unique* solution to [**RP**], so that we may speak of *the* optimal contract.

Part (e) of Proposition 3.2 is particularly notable when contrasted with the findings of Fu and Krishna (2014). They study a model of firm financing in which the agent faces a limited liability constraint and show that, even though the analogue of [**IC**] always binds in the iid version of their model, it necessarily holds as a strict inequality in some parts of the state space when types are sufficiently persistent.

To ease the exposition, Assumption 1 is assumed to hold throughout Sections 4 and H. Thus, all results in these sections should be read as contingent on Assumption 1.

4. The Optimal Contract — Long-Run Properties

We now describe the long-run properties of the optimal contract, which is unique by Proposition 3.2. A detailed analysis of the optimal contract's short-run properties is deferred to Section 5. Section 4.1 identifies the appropriate martingale, and Section 4.2 contains our main result on immiseration with persistent shocks.

4.1. The Differential Martingale

In the iid setting, Thomas and Worrall (1990) show that the derivative of their value function (which takes a scalar promised utility as an argument) is a martingale. The appropriate martingale in our Markovian setting and formulation is a somewhat subtle, but natural, generalization.

Theorem 3. *The optimal contract induces a stochastic process* $\mathsf{D}_{(1,1)} P(\mathbf{v}^{(t)}, s^{(t)})$*, which is a strictly positive martingale.*

The proof of Theorem 3 is in Appendix E. We will refer to the process $\big(\mathsf{D}_{(1,1)}\, P(\mathbf{v}^{(t)}, s^{(t)})\big)$ as the *differential martingale*. To see the intuition for Theorem 3, fix a vector $\mathbf{v} \in V$, and consider the marginal cost to the principal of increasing this promise to $\mathbf{v}' := (v_1 + \varepsilon, v_2 + \varepsilon)$ for some small $\varepsilon > 0$. One way to deliver the additional utility in an incentive-compatible way is to increase the continuation promises $\mathbf{w}_i$ to $\mathbf{w}_i' := (w_{i1} + \varepsilon/\alpha, w_{i2} + \varepsilon/\alpha)$. To a first-order approximation, the cost of this perturbation is $f_{s1}\mathsf{D}_{(1,1)}P(\mathbf{w}_1, 1) + f_{s2}\mathsf{D}_{(1,1)}P(\mathbf{w}_2, 2)$. An envelope theorem argument implies that this perturbation is locally optimal, so that this marginal cost equals $\mathsf{D}_{(1,1)}P(\mathbf{v}, s)$, giving us precisely the martingale property in Theorem 3.

To understand why the directional derivative must be taken in the direction $(1, 1)$, notice that this is the *unique* direction of change for the $\mathbf{w}_i$ that increases (or decreases) the agent's *ex* *ante* continuation utility *while leaving his information rent unchanged*.[24] In other words, it is the unique direction of change that increases (or decreases) the agent's expected continuation promise equally under *both* transition probabilities $\mathbf{f}_1$ and $\mathbf{f}_2$. Suppose, to the contrary, that in the previous paragraph we instead were to increase the continuation promise vectors to some $\hat{\mathbf{w}}_i \neq \mathbf{w}'_i$ such that $\mathbb{E}^{\mathbf{f}_i}\left[\hat{\mathbf{w}}_i\right] - \mathbb{E}^{\mathbf{f}_i}\left[\mathbf{w}_i\right] = \varepsilon$. This clearly delivers the extra $\varepsilon > 0$ of utility given each reported type, but may not be incentive compatible, as it follows from $\mathbf{f}_1 \neq \mathbf{f}_2$ and $\hat{\mathbf{w}}_1 \neq \mathbf{w}'_1$ that $\mathbb{E}^{\mathbf{f}_2}\left[\hat{\mathbf{w}}_1\right] - \mathbb{E}^{\mathbf{f}_2}\left[\mathbf{w}_1\right] \neq \varepsilon$. Indeed, if [IC] held as an equality under the original contract (as it must at the optimum, by part (e) of Proposition 3.2), then it would clearly be violated under this new one.

4.2. Convergence to Misery

With the appropriate martingale identified, we now describe the long-run properties of the optimal contract, our main economic result.

Theorem 4. *Under the optimal contract, with probability one: (i) $P_i(\mathbf{v}^{(t)}, s^{(t)}) \to 0$ for $i = 1, 2$ so that $\mathsf{D}_{(1,1)}P(\mathbf{v}^{(t)}, s^{(t)}) \to 0$, (ii) $v_i^{(t)} \to -\infty$ for $i = 1, 2$, (iii) $x_i^{(t)} \to -\infty$ for $i = 1, 2$, and (iv) $v_2^{(t)} - v_1^{(t)} \to \infty$.*

The proof of Theorem 4 is in Appendix E. The theorem shows that immiseration occurs regardless of the degree of persistence of shocks.[25] The key to this result is the notion of *martingale splitting*. In particular, we say that the differential martingale *splits at state* $(\mathbf{v}, s)$ if

$$\mathsf{D}_{(1,1)}P(\mathbf{w}_i, i) < \mathsf{D}_{(1,1)}P(\mathbf{v}, s) < \mathsf{D}_{(1,1)}P(\mathbf{w}_{3-i}, 3-i)$$

for some $i \in \{1, 2\}$. We show in Proposition 5.1 that, on the path induced by the optimal contract, the differential martingale must split at states of the form $(\mathbf{v}^{(t)}, s^{(t)} = 2)$, that is, after high endowment (θ_2) shocks. Thus, for any path along which there are infinitely-many high-endowment shocks (the set of such paths has full measure), the differential martingale cannot converge to a point in the interior of its range. But the martingale convergence theorem implies that it must settle down along almost every path, so it must converge to zero almost surely. This turns out to imply that the instantaneous allocations $x^{(t)}$ and contingent promised utilities $v_i^{(t)}$ all converge to $-\infty$, almost surely. Moreover, the difference $v_2^{(t)} - v_1^{(t)} \to \infty$ meaning that, in terms of levels of continuation utilities, the optimal contract is as far from full insurance (which, recall from Proposition 3.1, requires $v_1 = v_2$) as possible in the long run.

(24) It may be useful to draw a connection to the first-order approach in models with continuous types. In such a model, if we use ex ante promised utility and ex ante marginal promised utility as state variables, as in Williams (2011), the appropriate analogue to the differential martingale is the partial derivative of the principal's value function with respect to promised utility. Holding marginal promised utility constant means, via an envelope theorem, that the agent's information rents are unchanged.

(25) Recall that we assumed in Section 2 that the type process is ergodic. If there were absorbing states, the optimal contract would "freeze" in finite time and immiseration could not possibly occur.

It is worth emphasizing that persistence makes the arguments underlying the convergence result more subtle than in the iid case. Intuitively, with contingent utilities, the *iso-differential sets*

$$I_M := \{\mathbf{v} \in V : \mathsf{D}_{(1,1)} P(\mathbf{v}, s) = M\}$$

for $M > 0$ are *curves* in V. With the scalar state variable in Thomas and Worrall (1990), these sets are singletons and there is a one-to-one correspondence between ex ante promised utilities and martinagle levels.[26] With persistent shocks, it is possible that the contingent utility vector $\mathbf{v}^{(t)}$ drifts along some iso-differential set after consecutive low endowment (θ_1) shocks. This difference requires more elaborate mathematical arguments, and is also closely related to a new *persistence wedge* (see display [**PW**] below) in the optimal contract distinct to the case with persistent shocks. We elaborate on these points in Section 5 below.

It is worth briefly relating Theorem 4 to two recent results in the literature, namely, Theorem 3 of Zhang (2009) and the implications of Theorem 2 in Guo and Hörner (2014). Both of these results, like ours, concern the long-run behavior of screening contracts when types are persistent and follow a two-state Markov process. (As discussed in Section 6, Theorem 4 and the idea of its proof generalize beyond binary types and CARA utility.) Zhang, who studies a distinct but related taxation model in continuous time, only proves that immiseration occurs when the transition probabilities are symmetric across the two states (ie, when $f_{11} = f_{22}$), while we do so for all persistent transition matrices. Guo and Hörner find that both immiseration and convergence to bliss are (the only) possible long-run outcomes, which is a product of their restriction to bounded consumption levels and a risk-neutral agent. Importantly, Guo and Hörner do not uncover the differential martingale $\big(\mathsf{D}_{(1,1)} P(\mathbf{v}^{(t)}, s^{(t)})\big)$, which is central to our characterization of the long-run properties of the optimal contract.

5. Short-Run Dynamics — Primal Variables

In this section, we characterize the essential short-run features of the optimal contract and distill the role of persistent private information for allocative distortions. Additional details, including the full set of the principal's optimality conditions, can be found in Appendix H, where we characterize the short-run dynamics in terms of the principal's *dual variables*.[27]

As a first step, notice that for any level of persistence, the presence of private information implies that the principal cannot perfectly smooth the agent's consumption. To see this explicitly,

(26) In their setting, the state variable $v \in \mathbb{R}_{--}$ is ex ante promised utility and the value function $P : \mathbb{R}_{--} \to \mathbb{R}$ is strictly convex. It follows that their martingale $P'(\cdot)$ is strictly increasing.

(27) While the assumption of CARA utility (or some other form of homogeneity) is essential for the present exercise, the dual characterization in Appendix H holds more generally.

we may rearrange the first-order conditions to get

[**5.1**] $$\mathsf{D}_{(1,1)}P(\mathbf{w}_1, 1) = \frac{C'(x_1)}{\theta_1} - \frac{\mu(\mathbf{v}, s)}{f_{s1}} \cdot \frac{(\theta_1 - \theta_2)}{\theta_1}$$

[**5.2**] $$\mathsf{D}_{(1,1)}P(\mathbf{w}_2, 2) = \frac{C'(x_2)}{\theta_2}$$

where $\mu(\mathbf{v}, s) \geq 0$ is the multiplier on the incentive constraint, [**IC**]. Compared to the analogous conditions from the full-information problem (see Appendix C for details),

[**5.3**] $$\mathsf{D}_{(1,1)}\, Q^*(\mathbf{w}_i^*, i) = \frac{C'(x_1^*)}{\theta_1} = \frac{C'(x_2^*)}{\theta_2} \qquad \text{for } i \in S$$

we see that there are two types of allocative distortions. On the *intratemporal* margin, the principal distorts the allocation for low-endowment (θ_1) types downward to prevent under-reporting by the high-endowment types, as suggested by the second term on the right-hand side of [**5.1**]. Because the principal's value function differs from the first-best value function (ie, $P \neq Q^*$), the optimal contract typically also distorts allocations for both types on the *intertemporal* margin, meaning that the flow utilities are not constant over time as in the full-information optimum described in Proposition 3.1. Moreover, by taking the expectation of [**5.1**] and [**5.2**] under the transition probability $\mathbf{f}_s$ and invoking the martingale property, we may rewrite the differential martingale as

[**5.4**] $$\mathsf{D}_{(1,1)}P(\mathbf{v}, s) = \underbrace{\mathbb{E}^{\mathbf{f}_s}\left[\frac{C'(x)}{\theta}\right]}_{\text{insurance}} - \underbrace{\left(\frac{\theta_1 - \theta_2}{\theta_1}\right) \cdot \mu}_{\text{incentives}}$$

while the analogous condition in the full-information problem, derived from [**5.3**], is

[**5.5**] $$\mathsf{D}_{(1,1)}\, Q^*(\mathbf{v}, s) = \mathbb{E}^{\mathbf{f}_s}\left[\frac{C'(x^*)}{\theta}\right]$$

Thus, the contractual dynamics are driven by the optimal tradeoff between insurance, which is best achieved by perfectly-smoothed consumption streams, and incentive provision, which requires dispersion of the allocation to separate low- and high-endowment types. The following proposition states this more precisely.

Proposition 5.1. Under the optimal mechanism:

(a) There exists rays $\tilde{E}_1$ and $\tilde{E}_2$ in V such that when $\beta > 0$, $\tilde{E}_1$ lies strictly above $\tilde{E}_2$, and when $\beta = 0$, the rays coincide.

(b) When $\beta > 0$, the optimal policy induces $\mathbf{w}_2^{(t)} \in \tilde{E}_2$ and $\mathbf{w}_1^{(t)}$ that lies strictly below $\tilde{E}_1$. When $\beta = 0$, promised utility always lies on $\tilde{E}_1 = \tilde{E}_2$.

(c) For any $\beta \in [0, 1)$, after consecutive θ_2 shocks, $x_2^{(t)}$ and $\mathbf{w}_2^{(t)}$ monotonically increase, and there is *under-insurance*

$$C'(x_2)/\theta_2 > C'(x_1)/\theta_1$$

(d) If $\beta = 0$, there is *insurance*

$$x_1 > x_2$$

and *co-insurance*

$$\mathbb{E}^{\mathbf{p}}[\mathbf{w}_1] < \mathbb{E}^{\mathbf{p}}[\mathbf{v}] < \mathbb{E}^{\mathbf{p}}[\mathbf{w}_2]$$

after both types of shocks.

(e) For any degree of persistence $\beta \in [0, 1)$, we have *martingale splitting* after high endowment shocks: starting from any $\mathbf{v} \in \tilde{E}_2$,

$$\mathsf{D}_{(1,1)}P(\mathbf{w}_1, 1) < \mathsf{D}_{(1,1)}P(\mathbf{v}, 2) < \mathsf{D}_{(1,1)}P(\mathbf{w}_2, 2)$$

If $\beta = 0$, we also have martingale splitting after low endowment shocks.

The proof of Proposition 5.1 is in Appendix H. The implied dynamics of contingent promised utility are depicted in Figure 2, where the labeled points illustrate the dynamics after θ_2 shocks and the dark shaded region represents the support of $\mathbf{v}$ after θ_1 shocks. The content of Proposition 5.1 can be divided into two pieces: (i) "robust" properties that hold regardless of the degree of persistence, and (ii) properties that are new to the Markovian case.

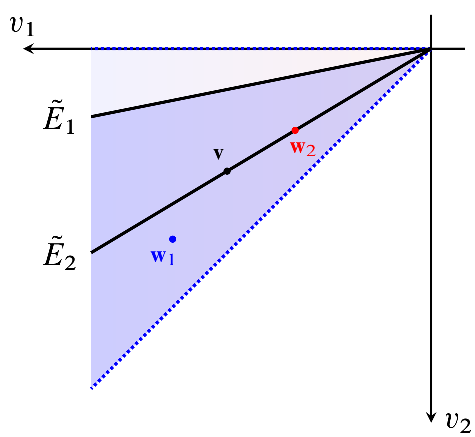

Figure 2: *Dynamics of* $\mathbf{v}^{(t)}$ *under the optimal mechanism*

Robust properties: Part (d) corresponds exactly to Lemma 5 of Thomas and Worrall (1990), which characterizes the optimal short-run dynamics in the iid case. Parts (b) and (c) offer a slightly looser, but qualitatively similar, characterization for the case of persistent shocks. They formalize the intuition, sketched above, that the consumption stream under the optimal contract needs to be excessively variable (relative to the full-information case) in order to screen the agent's private information. The martingale splitting described in part (e) is the driving force behind Theorem 4 and is a direct generalization of the analysis from the iid case. It states that the principal's marginal continuation cost decreases after two consecutive good (high endowment) shocks and increases after a good shock is immediately followed by a bad (low endowment) shock; it is a direct consequence of the variability of the optimal allocation.

One measure of the impact of private information is $\mathsf{D}_{(1,1)}\, Q^* - \mathsf{D}_{(1,1)}\, P$, which measures difference in marginal cost for the principal of raising contingent utiles by the same amount. From equations [5.4] and [5.5], it follows that

$$\underbrace{\mathsf{D}_{(1,1)}\, Q^*(\mathbf{v}, s) - \mathsf{D}_{(1,1)}\, P(\mathbf{v}, s)}_{\Delta\text{MC}} = \underbrace{\mathbb{E}^{\mathbf{f}_s}\left[\frac{C'(x^*) - C'(x)}{\theta}\right]}_{\text{allocative distortion}} + \underbrace{\frac{(\theta_1 - \theta_2)}{\theta_1}\mu(\mathbf{v}, s)}_{\text{shadow cost of private info}}$$

and from Theorem 4 (and its proof in Appendix E), we see that, under the optimal contract, all three terms vanish asymptotically. Thus, in the long-run, private information no longer matters, but this is only because the agent's utility becomes increasingly small, and so the cost to the principal also diminishes.

New properties: Parts (a) and (b) of Proposition 5.1, on the other hand, point to properties that are unique to the Markovian case. Essentially, they say that persistence introduces additional distortions relative to the iid case because it gives the agent an additional informational advantage over the principal, namely, the Markovian information rent term in [IC*]. More precisely:

- For any fixed *expected* continuation promise $w_i := \mathbb{E}^{\mathbf{f}_i}[\mathbf{w}_i]$, the unique cost-minimizing vector of contingent promises that delivers w_i lies on the ray $\tilde{E}_i$.[28] Thus, we refer to the rays $\tilde{E}_1$ and $\tilde{E}_2$ as *efficiency lines*. When shocks are iid ($\beta = 0$), the efficiency lines coincide, and optimal contract travels along this ray.
- When shocks are persistent ($\beta > 0$), the efficiency lines do not coincide. Moreover, after reporting a low endowment shock, the agent's interim continuation promise w_{11} is *higher*, and w_{12} is *lower*, than the vector lying in $\tilde{E}_1$ that delivers the same ex ante continuation promise. Thus, the dispersion of contingent promises $w_{12} - w_{11} > 0$ is *decreased* by moving off the efficiency line. This decreased dispersion is intuitive because, as shown in [IC*], with persistent private information the principal needs to manage the agent's Markovian information rent.
- As we discuss in Appendix H, the cost of this novel distortion can be quantified by using the first-order conditions to derive the *persistence wedge*:

[PW] $$\Upsilon(\mathbf{v}, \beta) := \frac{P_1(\mathbf{w}_1, 1)}{f_{11}} - \frac{P_2(\mathbf{w}_1, 1)}{f_{12}} = \frac{\mu(\mathbf{v}, s)}{f_{s1}} \cdot \frac{\beta}{f_{11} f_{12}} > 0$$

As $\Upsilon(\mathbf{v}, s)$ increases, the continuation vector $\mathbf{w}_1$ moves farther from the efficiency line $\tilde{E}_1$. Its magnitude is determined by direct and indirect effects. The direct effect corresponds

(28) Formally, the ray $\tilde{E}_i$ consists of those vectors $\mathbf{w}_i$ that, for some $w_i \in \mathbb{R}_{--}$, solve the problem $\min_{\mathbf{w}_i \in V} P(\mathbf{w}_i, i)$ subject to $w_i = \mathbb{E}^{\mathbf{f}_i}[\mathbf{w}_i]$. See Appendix H for details.

to the term $\beta/f_{11}f_{12}$, which is strictly increasing in β (for fixed f_{11}) and equal to zero if and only if $\beta = 0$. The indirect effect corresponds to the magnitude of the weighted [**IC**] multiplier $\mu(\mathbf{v}, s)/f_{s1}$.[29] While it is natural to expect that this increases in β as well, clean comparative statics are unavailable.

There is one final difference between the pathwise properties of the optimal contract in the case of iid and persistent shocks. Proposition 6 in Thomas and Worrall (1990) shows that, when the agent has CARA utility, the evolution of promised utility on the optimal path depends only on the number of high and low endowment shocks realized at any given date. In particular, promised utility evolves independently of the order of the shocks — a kind of *path independence* property. In the Markov case, this cannot be true because the endowment process itself depends on the most recently realized shock.

6. Discussion

To conclude the analysis of the Markovian model, it is useful to discuss extensions of our model to more general environments and highlight some connections to the broader dynamic contracting literature.

6.1. Assumptions and Extensions

For expositional simplicity, the analysis in Sections 3 – 5 relied on three simplifying assumptions laid out in Section 2: (i) CARA utility, (ii) binary endowment types, and (iii) *no hidden borrowing*, whereby the agent cannot over-report his endowment. Importantly, we show in Appendix I that our main results — namely, Theorems 1, 2, 3, and 4 — extend naturally to more general environments in which the agent's utility function has decreasing absolute risk aversion (DARA) and the endowment follows *any* fully connected finite-state Markov chain. These extensions maintain the *no hidden borrowing* assumption. Here, we provide a qualitative discussion.

General utility functions: The CARA assumption is used only in three parts of the analysis, namely, to (i) determine the shape of the set V in Theorem 1, (ii) facilitate the computation of some bounds in the proof of Theorem 2, and (iii) make the analysis of short-run dynamics in Section 5 as transparent as possible.[30] The domain V is still convex (and the value function P finite, convex, and smooth) under DARA utility, though harder to characterize analytically and requiring certain arguments to be adapted on a case-by-case basis. Appendix H shows that we are able to characterize the short-run dynamics in terms of the principal's *dual variables*

(29) As we show in the appendix, this quantity does not depend on the state s.

(30) The interested reader is referred to, respectively, (i) Appendices B and G, (ii) Lemmas C.1 and C.3 in Appendix C and Lemma D.5 in Appendix D, and (iii) Appendix H.3.

quite generally, though, in line with the dynamic insurance literature, we require homogeneity of some form to state these results in terms of the *primal* contractual variables, namely, the agent's promised utility. Thus, the CARA assumption is made merely for convenience and, beyond minimal technical qualifications, plays no role in the proofs of our main results.

On the other hand, one might conjecture that the special homogeneity properties implied by CARA utility — see, eg, parts (d) and (e) of Theorem 2 — would allow us to write the principal's problem using a one-dimensional state variable and obtain "closed form" solutions as in Green (1987) or Phelan (1998). This is not true with persistent shocks.[31] In contrast, it is straightforward to show in the iid case that the state space collapses down to a line.

More than two types: From a methodological perspective, the recursive representation of the principal's problem (see Section 3 and, in particular, Theorem 2) and the differential martingale (see Theorem 3) are valid for general finite type spaces and for general single-agent contracting problems that extend well beyond our canonical insurance setting. From an economic perspective, immiseration (see Theorem 4) continues to hold for more general endowment processes.

However, characterizing the optimal contract's short-run dynamics is more challenging for $n \geq 3$ types. The main difficulty, as pointed out by Battaglini and Lamba (2014) in the context of monopolistic screening, is that the pattern of binding constraints can be quite complicated under the optimal contract.[32]

Our recursive approach makes this problem especially clear: at each step, we can view the agent's type as a vector $t_i := (\theta_i, \mathbf{f}_i)$ and the allocation as $a_i := (x_i, \mathbf{w}_i)$. The agent's interim utility when type i and reporting type j is then $u(i,j) := \theta_i x_j + \mathbb{E}^{\mathbf{f}_i}\left[\mathbf{w}_j\right]$. When shocks are iid (so that $\mathbf{p} = \mathbf{f}_1 = \mathbf{f}_2$), the continuation term is independent of the agent's actual type and therefore cancels out of the incentive constraint [IC*]. This means that (i) the allocation is effectively one-dimensional, (ii) the agent's utility function $u(i,j)$ satisfies a single-crossing condition, and (iii) continuation utility acts like a *common numeraire*, or "money," for both types of agent. With positive persistence, by contrast, we lose single crossing on the type set $\{t_1, t_2\}$ because the transition probability vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ are *not* comparable (in the product order on $\mathbb{R}^2$ whereby $f_{11} > f_{21}$ and $f_{12} < f_{22}$). As a result, ex ante continuation utility no longer acts like "money" and the principal's recursive problem is effectively a sequence of (linked) static multidimensional screening problems.[33] These problems are difficult to analyze precisely because the pattern of binding constraints is determined endogenously by the chosen

(31) In Appendix J, we use a change of variables to best exploit the scale invariance implied by CARA utility, and even obtain a value function separable in two state variables. Nevertheless, obtaining a one-dimensional representation remains intractable, at least via our seemingly natural change of variables. However, it is certainly true that these homogeneity properties make numerical computations more efficient.

(32) In our setting with risk aversion and no monetary transfers, this problem is likely only exacerbated.

(33) As noted in Section 3.3 of Pavan, Segal and Toikka (2014), the closest analogy is to static screening problems with unidimensional types and multidimensional allocations.

mechanism.

On "no hidden borrowing": We maintain throughout the assumption of *no hidden borrowing*, whereby the agent cannot over-report his endowment. This is a natural technological assumption in our setting. Essentially, it requires that the agent does not have access to an external capital market or that, if he does, the principal is able to monitor his trading activity.[34] In the iid setting, Thomas and Worrall (1990) show that it is without loss to only consider these "downward" incentive constraints; like us, Williams (2011) directly imposes *no hidden borrowing* in a continuous-time analogue of our model with persistent private information.[35]

In the Markovian setting, it is theoretically possible that the optimal contract we derive is not robust to the possibility of hidden borrowing, though we do not know of any examples where this is the case. Namely, it will fail to be robust if and only if it violates the discrete-type analogue of the "integral monotonicity" condition of Pavan, Segal and Toikka (2014). General conditions on primitives that guarantee integral monotonicity are not well understood in the literature, even in quasi-linear settings; this is closely related to the above discussion on the challenges of multidimensional screening. It is beyond the scope of the present paper to characterize these conditions. However, we emphasize that *no hidden borrowing* is distinct from the *first-order approach* often taken in the literature. No hidden borrowing is a technological assumption that rules out certain reporting strategies for the agent. Given this constraint, we consider *all global (downward)* incentive constraints in the principal's problem (ie, study the full problem). The first-order approach, on the other hand, considers only *local* deviations by the agent (ie, looks at a relaxed problem), and often does not provide analytical verification that the solution to this relaxed problem is, in fact, a solution to the full problem.

6.2. On the Differential Martingale

The martingale approach can help shed light on the dynamics of the optimal mechanism in a variety of settings. Intuitively, the differential martingale measures the gradient of the Pareto frontier[36] which is determined, in effect, by the degree to which utility is transferable between the principal and the agent. In turn, the degree of transferability depends on (i) what types of monetary transfers are allowed, and (ii) risk aversion.

Abstracting from risk aversion, the role of money can be understood by contrasting Battaglini (2005), Fu and Krishna (2014), and Guo and Hörner (2014), which feature, respectively, unrestricted transfers, transfers restricted by a limited liability constraint, and no transfers.

(34) Admittedly, it may be less appropriate if the agent's private information represents a pure taste or productivity parameter that is not easily verifiable.

(35) See, in particular, Section 6 of that paper for an analysis of dynamic insurance with hidden income.

(36) Namely, the slope in direction $(1,1)$, which corresponds to varying ex ante promised utility but holding information rents (both iid and Markovian components) constant.

Money serves two related roles. First, it allows the principal to deliver many different levels of promised utility with the same allocation by varying the payment scheme, thereby minimizing backloaded allocative distortions. Second, it allows the principal to deliver promised utility before the agent's future private information has been realized.[37]

Using our recursive framework, the differential martingale is constant[38] in the setting of Battaglini (2005). Allocative distortions are induced only by the agent's *initial* private information and decrease over time; they are determined entirely by the evolution of the Lagrange multiplier μ on the incentive compatibility constraint. In the setting of Fu and Krishna (2014), the agent earns rents for private information that materializes after the initial date and incentives must be partially supplied through allocative distortions. The differential martingale is constant only in the region of the recursive domain where allocative distortions vanish and incentives are provided entirely through monetary transfers.[39] Outside of this region, where allocative distortions are required, the differential martingale must "split" as in our Sections 4 and 5. Finally, in Guo and Hörner (2014), where allocative distortions are the only instrument to provide incentives, the differential martingale never settles down; this corresponds to their long-run inefficiency results, whereby the agent converges either to misery or to bliss.

Our insurance model, adapted from Thomas and Worrall (1990), is most similar to Guo and Hörner (2014) in the sense that the principal only has one instrument — namely, consumption utiles — with which to provide incentives. However, the presence of risk aversion mechanically introduces a cost-smoothing motive for the principal that is not present in the above models, which is known to alter the dynamics of the optimal contract (and of the differential martingale) even when the principal has additional instruments.[40]

Taken together, our results and this discussion suggest that the differential martingale (and the principal's dual variables, more generally) can, through the kind of Martingale Convergence Theorem arguments used in this paper, help provide a unified perspective on the implications of environmental primitives on the long-run dynamics and allocative distortions of optimal contracts across a wide range of settings — especially those in which closed-form solutions are not available. We believe that this is an advantage of the recursive approach that warrants further exploration.

(37) In the extreme case of unrestricted transfers, Esö and Szentes (2017) show that this means the agent doesn't receive *any* rents for his future private information.

(38) In particular, it is equal to ± 1, depending on the sign convention.

(39) In their model, this region corresponds to a firm operating at the efficient scale.

(40) This is the case, for example, in dynamic taxation models that feature an agent whose utility is additively separable in labor and consumption. In these models, the agent's marginal utility for consumption is type-independent, meaning that consumption utiles act as a kind of "common numeraire" for all types of the agent; see footnote 7.

7. Averting Misery through Growth

Our results suggest that the long-run distributional implications of private information — namely, the classic immiseration results — are robust to details of the *information structure*. We conclude by pointing out that they are, however, sensitive to the underlying *technological assumptions* of the workhorse models. We build on the seminal analysis of Atkeson and Lucas (1992), which is the natural "general equilibrium" analogue of our contracting model. When the economy-wide resource constraint *grows* over time, the long-run implications of the optimal contract are somewhat subtle: *inequality* in the cross-sectional distributions of consumption and promised utility increases without bound, but the *levels* of these variables converge to their *upper* bounds. Thus, immiseration occurs in a *relative*, but not *absolute*, sense in the presence of aggregate growth. Indeed, in an absolute sense, exactly the opposite happens: all agents converge to *infinite consumption*, or *bliss*.

We lay out the model in Section 7.1 and present our results in Section 7.2. A discussion of interpretation and implications is deferred to Section 7.3. The analysis is deliberately simple — we rely directly on results from Atkeson and Lucas (1992) and Phelan (1998) — and should be viewed merely as a proof of concept pointing to an important avenue for future research.

7.1. Constant Growth Economies

The model is a straightforward many-agent version of the one described in Section 2. Time is discrete and runs over an infinite horizon. There is a continuum of agents (with mass normalized to one), each of whom faces idiosyncratic endowment shocks in each period. There is no aggregate uncertainty. A social planner, the principal, wants to deliver a given distribution of lifetime utilities to the population, subject to the usual incentive compatibility constraints. Following Atkeson and Lucas (1992) transparent and keep the analysis tractable, we assume throughout Section 7 that shocks are iid over time and across the population.[41]

We focus on the class of *constant growth economies*, which feature the per-period resource constraint

$$\mathbb{E}\left[\int c^{(t)} dm\right] \leq e \cdot G^t \qquad \forall t \in \mathbb{N} \tag{7.1}$$

where $e \geq 0$ is the economy's initial endowment, $G \geq 1$ is the constant gross growth rate, and m is a probability measure on $\mathbb{R}_{--}$ representing the distribution of initial lifetime utility promises across the population. (Invoking a continuum law of large numbers, we may take $\mathbb{P}(h)$ to be the fraction of the population that experiences the sequence of shocks $h \in \mathscr{H}$.) The *efficiency problem*, as defined by Atkeson and Lucas (1992), is to find the smallest $e \geq 0$ for which it is possible to deliver the given distribution of lifetime utilities in an incentive-compatible way.[42]

(41) We ignore technical issues associated with modeling a continuum of pairwise independent random variables.

(42) Because agents are assumed to have CARA utility, it may be possible to deliver certain distributions of

The economy with growth is equivalent to an economy without growth in which resource constraints apply to *relative consumption* $\chi^{(t)} := c^{(t)}/G^t$. The feasibility constraint in this transformed economy is

$$\text{[Feas]} \qquad \mathbb{E}\left[\int \chi^{(t)} dm\right] \leq e \qquad \forall\, t \in \mathbb{N}$$

Theorem 2 in Atkeson and Lucas (1992) shows how to represent the continuum-agent efficiency problem as a collection of separate single-agent contracting problems, each of which can be analyzed with recursive methods.[43] The efficiency problem then reduces to finding some $q \in (0, 1)$ such that the solution to

$$\text{[7.2]} \qquad \min_{\{\chi_t\}} \sum_{t=0}^{\infty} q^t \mathbb{E}\left[\chi^{(t)}\right]$$

or, equivalently,

$$\text{[7.3]} \qquad \min_{\{c_t\}} \sum_{t=0}^{\infty} \hat{q}^t \mathbb{E}\left[c^{(t)}\right]$$

with $\hat{q} := q/G$, subject to sequential promise keeping and incentive compatibility constraints, also satisfies [**Feas**] with equality. Any such $q \in (0, 1)$, if it exists, is said to be a *price system*, and the corresponding optimal allocation is said to be *efficient* subject to the resource constraint.

7.2. Relative and Absolute Immiseration

Atkeson and Lucas (1992) show that a price system exists when all agents have identical CARA utility functions and $G = 1$, and their arguments extend in a straightforward manner to the class of constant growth economies with $G > 1$. We assume that $G > 1$ is sufficiently small (relative to the discount rate) to render the efficiency problem *nontrivial*, meaning that the cost of delivering finite lifetime utilities to the agents is bounded away from zero for the planner.[44] Our main result shows that immiseration occurs only in a relative sense.

Proposition 7.1. If all agents have identical CARA utility, then in any constant growth economy with $G > 1$ and a nontrivial efficiency problem:

initial promised utilities m with negative consumption goods, ie, with $e < 0$. To keep statements as simple as possible, we assume that m requires that $e \geq 0$ in any solution. Clearly, such initial distributions exist.

(43) Phelan (1998) also discusses this approach.

(44) If G is sufficiently large, it is possible to deliver infinite utility to every agent in the economy at essentially zero cost to the principal — ie, that there is some sequence $\{e_n\}_{n=1}^{\infty}$ such that each e_n is a feasible solution to the efficiency problem and $e_n \to 0$. This can be done in an incentive compatible way, by giving each agent a constant fraction of the aggregate endowment in each period. Such an efficiency problem is *trivial* because the lower bound, in terms of costs to the principal, is zero. Formally, a solution to the principal's problem does not exist in this case.

(a) There exists a price system;
(b) Individual consumption is unbounded above: for a full measure of agents, $c^{(t)} \to +\infty$;
(c) Instantaneous and promised utilities converge almost surely to 0, their upper bound;
(d) The variance of relative consumption increases without bound: $\mathbb{V}(\chi^{(t)}) \to +\infty$.

The proof of Proposition 7.1 is in Appendix F. Parts (b) and (c) mean that there is no *absolute* immiseration in the optimal mechanism. Even with private information, any strictly positive growth saves (almost) everyone from poverty. Indeed, exactly the opposite occurs: (almost) every agent converges to *bliss*. Yet, part (d) says that, even with growth, private information requires that the optimal mechanism induces *relative* immiseration. In the long run, inequality increases without bound so that the bottom tail of the wealth distribution becomes infinitely poorer than the top tail. Part (d) supplements the findings of Atkeson and Lucas (1992), who show that both absolute and relative immiseration occur when there is no growth. Proposition 7.1 shows that only the latter prediction is valid when there is growth, however slow.

Intuitively, an economy with positive growth is isomorphic to one with zero growth but where the principal is relatively *less patient*. To see why this is true, note that if $G > 1$, then the discount factors q and $\hat{q}$ defined, respectively, in [7.2] and [7.3] satisfy $q > \hat{q}$. In such a world, providing continuation utiles is cheap for the principal relative to providing instantaneous consumption, so she will optimally backload payments. This causes promised utility to drift up over time. Proposition 7.1 shows that, in a pathwise sense, this upward drift overpowers the tendency for promised utility to fan out that drives the usual immiseration result. On average, however, this fanning-out still means that inequality increases over time and without bound.

7.3. Interpretation and Future Work

Insofar as such a simple model can be taken seriously as a theory of distribution, these findings suggest that the extent of inequality is an inherently normative question while the level of poverty hinges on technological constraints. From a positive perspective, inequality, not poverty, is the predicted consequence of private information frictions over sensible macroeconomic time horizons. This is ultimately a prediction that must be taken to the data.[45]

Proposition 7.1 is complementary to the main findings of Farhi and Werning (2007). Both results suggest that absolute immiseration is a knife-edge result. We show that it is overturned if there is *any* growth; they show that it is overturned if the principal is even *slightly* more patient

(45) It is common in the literature to interpret this class of insurance models as describing multiple generations in a dynastic economy. But under this interpretation, each period corresponds to roughly 30 years of calendar time, and the long-run properties of the optimal mechanism must be understood as describing limiting behavior that takes place over hundreds or thousands of years. Certainly, aggregate resources are not constant over these very long time horizons. For the *positive* side of the theory, it is therefore important to understand the distributional implications of private information in growing economies.

than the agents. However, our optimal mechanism results in unbounded inequality, while Farhi and Werning's leads to a steady-state distribution and an asymptotic mixing condition, which they refer to as "social mobility." The intuition is straightforward: even with growth, if the principal only places positive Pareto weight on the initial generation, then it is optimal to deliver punishments and rewards (as is manifested through the fanning-out of continuation utilities) *permanently*.

Proposition 7.1 is also suggestive of a more general normative point. Namely, welfare tradeoffs can be very different when the principal is able to control not only allocations, but also the rate at which the economy's aggregate resources grow over time. Consider, for example, the extreme case where the principal has a Rawlsian objective. When aggregate resources are constant over time, the principal will optimally redistribute resources so as to move the utility distribution towards perfect equality (subject to incentive constraints). But the absence of absolute immiseration in Proposition 7.1 — and, indeed, the fact that parts (b) and (c) show all agents converging to bliss in the long run — suggests that, if redistribution were to decrease the rate of growth, even such a Rawlsian principal may find it optimal to induce an allocation that results in relative immiseration, and hence infinite inequality in the long run.

The simple model presented here assumes that growth is exogenous. It is important to better understand the endogenous interaction between growth and the dynamics of distribution under private information, and the dependence of this interaction on the underlying technological environment. Such questions are central to the design of optimal policy in macroeconomics and public finance, and are also related to the dynamics of inequality and hierarchy within organizations. We are actively pursuing these directions in ongoing work.

Appendices

The appendices contain proofs of propositions stated in the text, and much else besides. We provide here a guide to their organization. Appendix A formally describes sequential contracts and the sequential version of the principal's problem [**SP**]. Appendix B provides a construction of the domain for our recursive contracts, by showing that the set V in Theorem 1 is the largest fixed point of an APS-style operator. While there are many recursive contracts that do not satisfy [**TVC**], Appendix G provides an explicit construction of a recursive contract that satisfies [**TVC**], and thereby provides a useful upper bound for the value function P.

Appendices C and D describe, respectively, the full information value function and the principal's value function with private information. Appendix D also details properties of the value function P and those of the optimal contract, culminating in Appendix D.5, which brings together earlier results to prove Theorem 2 and Proposition 3.2. Appendix E proves that the process $\mathsf{D}_{(1,1)}\, P(\mathbf{v}, s)$ is a martingale and also its long run properties, thereby establishing Theorems 3 and 4. Appendix H describes further short-run properties of the optimal contract, some of which are the content of Proposition 5.1 and some of which provide the characterization in terms of the principal's dual variables, as discussed in Sections 5 and 6. Proposition 7.1 is proved in Appendix F.

A more general model, with n types and more general utility functions, is considered in Appendix I, where it is shown that our results extend to this case too. Appendix J revisits the model with binary types, and considers a change of variables, to polar coordinates. While this makes the value function separable in two variables, it does not lead to an explicit solution. Indeed, it highlights why the Markovian case makes explicit calculations difficult. Finally, Appendix K collects some miscellaneous properties of Markov chains.

A. Sequential Contracts

In this appendix, we formally define sequential contracts, provide the details of the principal's sequence problem, and formalize the connection between this sequence problem and the recursive problem [**RP**] defined in Section 3 of the main text.

Intuitively, a sequential contract specifies consumption allocations (transfers from the principal to the agent, or vice versa) conditional on histories of previous allocations and messages from the agent. By the Revelation Principle, it is without loss to consider direct revelation mechanisms, whereby the principal chooses allocations conditional on histories of reported endowments. Let $\mathcal{G}$ denote the space of *private* histories, which are sequences of *realized* endowment types of the form $g = (s^{(0)}, s^{(1)}, \dots) \in S^{\infty}$. Similarly, $\mathcal{H}$ is the space of *public* histories, which are sequences of *reported* endowment types $h = (\hat{s}^{(0)}, \hat{s}^{(1)}, \dots) \in S^{\infty}$. Let G^t and H^t denote the spaces of length-t private and public histories, respectively. A *(pure) reporting strategy* for the agent is a sequence of functions $s_t : G^t \times H^{t-1} \to S$, which map past realized shocks, the current realized shock, and past reports into a current reported shock type. Every reporting strategy $\mathfrak{s} = (s_t)_{t=0}^{\infty}$ induces a probability measure $\sigma \in \Delta(\mathcal{H})$, which may differ from the measure induced by the true transition probabilities. We will let $\sigma^* \in \Delta(\mathcal{H})$ denote the distribution induced by the *truthful strategy* defined by $s_t^*(\cdot) := s^{(t)}$.[46] We say that a strategy $\mathfrak{s}$ is *admissible* if $s_t(\cdot) = s^{(t)}$ for all private histories of the form $g^t = (s^{(0)}, \dots, s^{(t)} = 1) \in G^t$, which feature an $s = 1$ shock in the current period. We restrict attention to admissible strategies. That is, following the discussion in Section 2, the agent cannot over-report low endowment shocks. We denote the space of admissible reporting strategies by $\mathcal{S}$ and the space of induced measures over public histories by Σ; in the sequel, we will refer almost exclusively to Σ.

At time $t = 0$, the principal promises an agent with initial type $s \in S$ exactly $v_s \in \mathbb{R}_{--}$ lifetime utiles, summarized by the vector of *contingent* promised utilities $\mathbf{v} := (v_1, v_2)$. She has a prior belief $\mu \in (0, 1)$ that $\theta^{(0)} = \theta_1$, so that $v := \mu v_1 + (1 - \mu) v_2$ is the corresponding ex-ante promised utility for the agent.[47] By varying the ex ante promise v, we can trace out the entire constrained Pareto frontier.

Sequential contracts are naturally described in terms of flow utilities. Recall that a transfer of c from the principal to the agent, net of the agent's endowment, delivers flow utility $\theta_s u(c)$. Hence, any

(46) In the sequel, we will use the notation $\mathbb{P}^{\sigma}$ to denote these induced measures over joint public-private history pairs. Similarly, $\mathbb{E}^{\sigma}$ will denote the associated expectation operators. By construction, $\mathbb{E}^{\sigma^*} = \mathbb{E}$, where un-indexed expectations are with respect to the measure $\mathbb{P}$ induced by the true transition probabilities.

(47) We will mostly suppress explicit mention of the prior belief, though it is implicit in all probabilistic statements.

such transfer is equivalent to a flow utility allocation $x := u(c)$, so that the agent's utility in state s is given by $\theta_s x$. Furthermore, define $C(x) := u^{-1}(x)$ to be quantity of consumption good associated with utility transfer x. Thus, a *sequential contract* is a sequence of functions $x_t : \mathbb{R}^2_{--} \times H^t \to \mathbb{R}$ for $t = 0, 1, 2, \ldots$, which condition on all past and current reports and the initial vector of promises $\mathbf{v}$.[48] Every sequential contract $\tilde{x}$ induces a stochastic processes $(x^{(t)})_{t=0}^{\infty}$ on $\mathbb{R}_{--}$, the transition probabilities of which are determined by the agent's reporting strategy. When the agent follows the strategy σ, the cost to the principal of a sequential contract $\tilde{x}$ is

$$R(\tilde{x}, \sigma) := \mathbb{E}^{\sigma}\left[\sum_{t=0}^{\infty} \alpha^t C(x^{(t)})\right]$$

We restrict attention to contracts that induce well-defined cost $R(\tilde{x}, \sigma)$ (though values of $\pm\infty$ are permissible).[49] The set of all sequential contracts is $\mathcal{A}$. The agent has lifetime utility function $U : \mathcal{A} \times \Sigma \to \mathbb{R}$ given by

$$U(\tilde{x}, \sigma) := \mathbb{E}^{\sigma}\left[\sum_{t=0}^{\infty} \alpha^t \theta^{(t)} x^{(t)}\right]$$

We say that a sequential contract $\tilde{x}$ *implements* $\mathbf{v} \in \mathbb{R}^2_{--}$ if it satisfies

[S-PK$_1$] $$v_1 = U(\tilde{x}, \sigma^* | s^{(0)} = 1)$$

[S-PK$_2$] $$v_2 = U(\tilde{x}, \sigma^* | s^{(0)} = 2)$$

[S-IC] $$U(\tilde{x}, \sigma^*) \geq U(\tilde{x}, \sigma) \qquad \forall\, \sigma \in \Sigma$$

The first two constraints, [**S-PK$_1$**] and [**S-PK$_2$**], are the familiar *promise-keeping* conditions, which ensure that the principal actually delivers the appropriate level of contingent lifetime utility to the agent. The *incentive compatibility* constraint [**S-IC**] requires that truthtelling is an optimal reporting strategy for the agent.

The set of sequential contracts that implement $\mathbf{v}$ is $\Pi(\mathbf{v})$. The principal's objective is to minimize the expected, discounted cost of delivering $\mathbf{v}$ to the agent. Her value function is therefore

[SP] $$P^*(\mathbf{v}, \mu) := \inf_{\tilde{x} \in \Pi(\mathbf{v})} R(\tilde{x}, \sigma^*)$$

We refer to this as the principal's *sequence problem*. Every fixed sequential contract $\tilde{x} \in \Pi(\mathbf{v})$ induces a stochastic control problem for the agent, in which his choice set is the space of reporting strategies Σ and his optimal choice is σ^*, the truthtelling strategy.

With this terminology in hand, one can show that the sequence problem [**SP**] and recursive problem [**RP**] are equivalent in the following sense:[50]

(48) More generally, randomized allocations are possible, but will not be necessary in our setup due to standard convexity assumptions.

(49) This will be without loss of generality.

(50) We omit a formal proof, which follows almost exactly from arguments in the proofs of Lemmas 3.1 and 3.2 in Atkeson and Lucas (1992). The only gap is that recursive contracts, as we have defined them, are *stationary*, while Atkeson and Lucas allow them to depend explicitly on time. But this is without loss of generality in the present setting; see footnote 16.

Proposition A.1. For every implementable sequential contract $\tilde{x} \in \Pi(\mathbf{v})$, there is an incentive compatible recursive contract $\xi \in \Xi(\mathbf{v})$ that is weakly less costly for the principal. Conversely, any [**TVC**]-implementable recursive contract $\xi \in \Xi(\mathbf{v})$ induces a unique implementable sequential contract $\tilde{x}_\xi \in \Pi(\mathbf{v})$.

However, we note that, in general, the sequence problem [**SP**] and the recursive problem [**RP**] are *not* truly equivalent. As the constraints of [**RP**] require only one-step promise keeping and incentive compatibility, they are relaxations of the global constraints from the sequence problem [**SP**]. Indeed, because the agent's utility function is unbounded below, this is a *strict* relaxation. Hence, we will always have $P \leq P^*$, and it is possible that $P(\mathbf{v}, s) < P^*(\mathbf{v}, s)$ for some $(\mathbf{v}, s) \in V \times S$. However, Proposition A.1 immediately yields a useful corollary: if the solution to [**RP**] is [**TVC**]-implementable (i.e., if there exists an *globally optimal* recursive contract), then the induced allocation solves the sequence problem [**SP**] and hence $P = P^*$, where P is the value function in problem [**RP**]. Part (g) of Theorem 2 and Proposition 3.2 in Section 3.4 state this connection formally.

B. Proof of Theorem 1

B.1. Outline

Mathematically, the problem of determining V is identical to the Abreu, Pearce and Stacchetti (1990) procedure for finding the largest self-generating set in repeated games. Our proof is constructive: we use Farkas' Lemma to pin down the boundaries of the appropriate convex cone and then iterate on an the appropriate APS-style operator. Since CARA utility is strictly negative and unbounded below, the set V must be both open and unbounded. A notable technical feature of the proof is that the iterative step effectively requires transfinite induction to accommodate openness. To ensure that (i) V is the appropriate domain for the full problem sequence [**SP**] (in the sense that $\mathbf{v} \in V$ if and only if $\Pi(\mathbf{v}) \neq \emptyset$), and (ii) there exists a [**TVC**]-implementable recursive contract (so that Assumption 1 is not vacuous), we explicitly construct a suboptimal [**TVC**]-implementable contract, called ζ, that self-generates at a single point after finitely-many steps. The construction of ζ is carried out in Appendix G.

B.2. Formalities

For any set $K \subset \mathbb{R}^2_-$, define the (APS) operators $\mathscr{B}, \mathscr{B}^\circ : 2^{\mathbb{R}^2_-} \setminus \varnothing \to 2^{\mathbb{R}^2_-} \setminus \varnothing$ as

$$\mathscr{B}(K) := \left\{ \mathbf{v} \in \mathbb{R}^2_- : (x_i, \mathbf{w}_i)_{i=1,2} \text{ implements } \mathbf{v} \text{ and } x_i \leq 0,\ \mathbf{w}_i \in K \right\}$$
$$\mathscr{B}^\circ(K) := \left\{ \mathbf{v} \in \mathbb{R}^2_- : (x_i, \mathbf{w}_i)_{i=1,2} \text{ implements } \mathbf{v} \text{ and } (x_i, \mathbf{w}_i) \ll \mathbf{0},\ \mathbf{w}_i \in K \right\}$$

We will consider two methods of implementing $\mathbf{v} \in \mathbb{R}^2_-$. The first (corresponding to the operator $\mathscr{B}$) will only require $x_i \leq 0$ and $\mathbf{w}_i \leq 0$, while the second (corresponding to the operator $\mathscr{B}^\circ$) will require that $x_i < 0$ and $\mathbf{w}_i \ll \mathbf{0}$. (Notice that the second requirement precludes us from implementing $\mathbf{0}$.)

We will use the following version of Farkas' Lemma in the sequel.[51]

Theorem 5 (Farkas' Lemma)**.** *Let A be an $m \times n$ matrix, B be an $\ell \times n$ matrix, $a^* \in \mathbb{R}^m$, and $b^* \in \mathbb{R}^\ell$. Exactly one of the following alternatives holds. Either there exists $x \in \mathbb{R}^n$ such that*

$$\begin{aligned} Ax &= a^* \\ Bx &\leq b^* \\ x &\leq 0 \end{aligned} \qquad \text{[B.1]}$$

or else there exists $p \in \mathbb{R}^m$ and $q \in \mathbb{R}^\ell$ such that

$$\begin{aligned} pA + qB &\leq 0 \\ q &\geq 0 \\ p \cdot a^* + q \cdot b^* &< 0 \end{aligned} \qquad \text{[B.2]}$$

For each $t > 0$, define $K_t := \left\{\mathbf{v} \in \mathbb{R}^2_- : v_2 \geq t v_1\right\}$. Notice that K_t is a closed convex cone and contains the origin.

Lemma B.1. Let $t := f_{22}/f_{12} > 1$. Then, $\mathcal{B}(\mathbb{R}^2_-) = K_t$.

Proof. It is easy to see that $\mathbf{v} = \mathbf{0}$ is implemented by $(x_i, \mathbf{w}_i) = \mathbf{0} \in \mathbb{R}^2_-$. Moreover, if $v_2 = 0$, then we must have $(x_2, \mathbf{w}_2) = \mathbf{0}$. Therefore, in what follows, we shall assume that $v_2 \neq 0$.

We want to show that the system of linear inequalities

$$\theta_1 x_1 + \alpha f_{11} w_{11} + \alpha f_{12} w_{12} = v_1 \qquad [\textbf{PK}_1]$$

$$\theta_2 x_1 + \alpha f_{21} w_{11} + \alpha f_{22} w_{12} \leq v_2 \qquad [\textbf{IC}]$$

$$(x_i, \mathbf{w}_i) \leq \mathbf{0} \qquad [\textbf{Feas}]$$

has a solution. Notice that we have ignored [$\textbf{PK}_2$] because the variables $(x_2, \mathbf{w}_2)$ can be chosen independently of the solution to the above equations.

These inequalities do not have a solution if, and only if, there exist $p, q \in \mathbb{R}$ such that the dual system for Farkas' Lemma (Theorem 5) has a solution:

$$p\theta_1 + q\theta_2 \leq 0 \qquad [\textbf{B.3}]$$

$$pf_{11} + qf_{21} \leq 0 \qquad [\textbf{B.4}]$$

$$pf_{12} + qf_{22} \leq 0 \qquad [\textbf{B.5}]$$

$$q \geq 0 \qquad [\textbf{B.6}]$$

$$pv_1 + qv_2 < 0 \qquad [\textbf{B.7}]$$

Because $q \geq 0$, it follows from [**B.3**] that $p \leq -q(\theta_2/\theta_1) \leq 0$. Notice that if $q > 0$, then $p < 0$. If $q = 0$, then [**B.7**] would require that $p > 0$ and [**B.5**] would require that $p \leq 0$, which is impossible.

(51) See Vohra (2004) for a discussion of Farkas' Lemma, and Proposition 38 on p. 17 of Border (2013) for exactly this version.

Hence, $q > 0$ for any solution of this dual system, and consequently $p < 0$. In particular, if $v_2 < 0$, then [B.7] if, and only if, $q > (-p)\,|v_1|\,/\,|v_2| > 0$.

We shall find necessary and sufficient conditions for there to exist such $p, q \in \mathbb{R}$ that satisfy the dual system [B.3]–[B.7] does not possess a solution.

Conditions [B.3]–[B.5] can be rewritten as

$$q \leq (-p)(\theta_1/\theta_2)$$
$$q \leq (-p)(f_{11}/f_{21})$$
$$q \leq (-p)(f_{12}/f_{22})$$

Notice that $-p > 0$ and $f_{12}/f_{22} < 1 < \min[f_{11}/f_{21}, \theta_1/\theta_2]$. Then, if $|v_1|\,/\,|v_2| = f_{12}/f_{22}$, it is clear that we cannot find a $q \in \mathbb{R}$ that simultaneously satisfies $q > (-p)\,|v_1|\,/\,|v_2| > 0$ and $q \leq (-p)(f_{12}/f_{22})$. (Notice that f_{12}/f_{22} is the smallest value of $|v_1|\,/\,|v_2|$ for which the dual does not possess a solution. Therefore, $|v_1|\,/\,|v_2| = f_{12}/f_{22}$ is the weakest necessary condition for the dual to not possess a solution and so $|v_1|\,/\,|v_2| \geq f_{12}/f_{22}$ is also a sufficient condition.)

Therefore, if $|v_1|\,/\,|v_2| \geq f_{12}/f_{22}$, the dual system does not have a solution. But this conditions is equivalent to $-v_1 f_{22} \geq -v_2 f_{12}$, which is equivalent to requiring that $v_2 \geq t v_1$, where $t = f_{22}/f_{12} > 1$. In other words, $\mathcal{B}(\mathbb{R}^2_-) = K_t$, as claimed. □

The next lemma computes the iterates of K_t under $\mathcal{B}$.

Lemma B.2. Let $t > 1$ and define $t' := (f_{21} + t f_{22})/(f_{11} + t f_{12})$. Then, $\mathcal{B}(K_t) = K_{t'}$.

Proof. Once again, $\mathbf{v} = \mathbf{0}$ is implemented by $(x_i, \mathbf{w}_i) = \mathbf{0} \in \mathbb{R}^2_-$. Moreover, if $v_2 = 0$, then we must have $(x_i, \mathbf{w}_i) = \mathbf{0}$. Therefore, as before, we shall assume that $\mathbf{v} \leq 0$.

The requirement that $\mathbf{w}_i \in K_t$ amounts to requiring that $\alpha w_{12} \geq \alpha t w_{11}$. Therefore, we want to show that the system of linear inequalities

[PK$_1$]	$\theta_1 x_1 + \alpha f_{11} w_{11} + \alpha f_{12} w_{12} = v_1$
[IC]	$\theta_2 x_1 + \alpha f_{21} w_{11} + \alpha f_{22} w_{12} \leq v_2$
[Feas-K_t]	$\alpha t w_{11} - \alpha w_{12} \leq 0$

has a solution. Notice that we have ignored [PK$_2$] because the variables $(x_2, \mathbf{w}_2)$ can be chosen independently of the solution to the above equations.

These inequalities do not have a solution if, and only if, there exist $p, q_1, q_2 \in \mathbb{R}$ such that the dual system for Farkas' Lemma (Theorem 5) has a solution:

[B.8]	$p\theta_1 + q_1\theta_2 \leq 0$
[B.9]	$p f_{11} + q_1 f_{21} + q_2 t \leq 0$
[B.10]	$p f_{12} + q_1 f_{22} - q_2 \leq 0$
[B.11]	$q_1, q_2 \geq 0$
[B.12]	$p v_1 + q_1 v_2 < 0$

Because $q_1 \geq 0$, it follows from [**B.8**] that $p \leq -q_1(\theta_2/\theta_1) \leq 0$. Therefore, if $q_1 > 0$, then $p < 0$. Moreover, if $q_1 = 0$, we must necessarily have $p > 0$, or else we would violate [**B.12**]. Therefore, we must necessarily have $p > 0$. Also, $q_1 = 0$ is impossible, for [**B.12**] would imply $p > 0$ and [**B.8**] would imply $p \leq 0$. Hence $q_1 > 0$ and $p < 0$ at any solution to this dual system. In particular, if $v_2 < 0$, then [**B.12**] if, and only if, $q_1 > (-p)\,|v_1|\,/\,|v_2| > 0$.

We shall find necessary and sufficient conditions for there to exist such $p, q \in \mathbb{R}$ that satisfy the dual system [**B.3**]–[**B.7**] does not possess a solution.

Conditions [**B.8**]–[**B.10**] can be rewritten as

$$\begin{aligned} q_1 &\leq (-p)(\theta_1/\theta_2) \\ q_1 f_{21} + q_2 t &\leq (-p) f_{11} \\ q_1 f_{22} - q_2 &\leq (-p) f_{12} \end{aligned}$$

The last two of these inequalities has a solution if, and only if, $q_1 \leq (-p)(f_{11} + t f_{12})/(f_{21} + t f_{22})$. (This is easily seen if you draw a picture. Draw the q_1-q_2 plane for a fixed p.) Moreover, it is easy to see that $1/t' := (f_{11} + t f_{12})/(f_{21} + t f_{22}) < 1$ because $t > 1$.

Notice also that $\theta_1/\theta_2 > 1$, so if $q_1 \leq (-p)(1/t')$, then $q_1 \leq (-p)(\theta_1/\theta_2)$ is automatically satisfied.

This implies that if $|v_1|\,/\,|v_2| \geq 1/t'$, it is clear that we cannot find a $q_1 \in \mathbb{R}$ that simultaneously satisfies [**B.12**], ie, $q_1 > (-p)\,|v_1|\,/\,|v_2| > 0$, and $q_1 \leq (-p)(1/t')$. (Notice that $1/t'$ is the smallest value of $|v_1|\,/\,|v_2|$ for which the dual does not possess a solution. Therefore, $|v_1|\,/\,|v_2| = 1/t'$ is the weakest necessary condition for the dual to not possess a solution and so $|v_1|\,/\,|v_2| \geq 1/t'$ is also a sufficient condition.)

Therefore, if $|v_1|\,/\,|v_2| \geq 1/t'$, the dual system does not have a solution. But this conditions is equivalent to $-v_1 t' \geq -v_2$, which is equivalent to requiring that $v_2 \geq t' v_1$, where $t = (f_{21} + t f_{22})/(f_{11} + t f_{12}) > 1$. In other words, $\mathcal{B}(\mathbb{R}^2_-) = K_t$, as claimed.

Finally, we shall show that $t' < t$. To see this, notice that $t' < t$ if, and only if, $f_{12} t^2 + t(f_{11} - f_{22}) - f_{21} > 0$. Let us define the quadratic polynomial $\varphi(\lambda) := f_{12}\lambda^2 + \lambda(f_{11} - f_{22}) - f_{21}$. Then, $t' < t$ if, and only if, $\varphi(t) > 0$.

Let us observe the following properties of φ: $\varphi(1) = 0$ and $\varphi'(1) = 2f_{12} + (f_{11} - f_{22}) = 1 - \beta > 0$ because $\beta = f_{22} - f_{12} > 0$. Therefore, $\varphi(t) > 0$ because φ is quadratic and $t > 1$. But this implies $t' < t$, as claimed. Thus, $\mathcal{B}(K_t) = K_{t'} \subset K_t$. ☐

Let us inductively define $\mathcal{B}^n(\cdot) := \mathcal{B}\left(\mathcal{B}^{n-1}(\cdot)\right)$ as the n-fold iteration of $\mathcal{B}$. Consider the iteration $\lim_{n\to\infty} \mathcal{B}^n(\mathbb{R}^2_-)$ and observe that because $\mathcal{B}(K_t) \subsetneq K_t$ whenever $t > 1$, we have $\lim_{n\to\infty} \mathcal{B}^n(\mathbb{R}^2_-) = K_1$. We show next that K_1 is a fixed point of $\mathcal{B}$.

Lemma B.3. The cone K_1 satisfies $\mathcal{B}(K_1) = K_1$.

Proof. The claim follows immediately from the proof of Lemma B.2 once we notice that the map $t \mapsto (f_{21} + t f_{22})/(f_{11} + t f_{12})$ maps 1 to 1. That is, $(f_{21} + t f_{22})/(f_{11} + t f_{12}) = 1$ if $t = 1$. ☐

For a set $K \subset \mathbb{R}^2$, let int K denote the interior of the set.

Lemma B.4. The operator $\mathscr{B}^\circ$ satisfies $\mathscr{B}^\circ(\operatorname{int} K_1) = \operatorname{int} K_1$.

Proof. Consider, again, the equations

$$[\mathbf{PK}_1] \qquad \theta_1 x_1 + \alpha f_{11} w_{11} + \alpha f_{12} w_{12} = v_1$$

$$[\mathbf{IC}] \qquad \theta_2 x_1 + \alpha f_{21} w_{11} + \alpha f_{22} w_{12} \leq v_2$$

Multiply [$\mathbf{PK}_1$] by $t = \theta_2/\theta_1 < 1$ and subtract it from [**IC**]. This results in

$$[\mathbf{IC}^\star] \qquad \alpha w_{11}(f_{21} - t f_{11}) + \alpha w_{12}(f_{22} - t f_{12}) \leq v_2 - t v_1$$

Notice that [$\mathbf{PK}_1$] and [$\mathbf{IC}^\star$] together imply [**IC**]. To see this, multiply [$\mathbf{PK}_1$] by $t = \theta_2/\theta_1 < 1$ and add it to [$\mathbf{IC}^\star$] to obtain [**IC**].

Let $\eta_2 := f_{22} - t f_{12}$ and $\eta_1 := f_{21} - t f_{11}$ and notice that $\eta_2 > 0$ because $f_{22} > f_{12}$. We shall consider two cases that depend on the sign of η_1.

Case 1: $\eta_1 \geq 0$. Then, any choice of $\mathbf{w}_1 \in \operatorname{int} K_1$, which entails $w_{11} < w_{12} < 0$ will result in $\alpha w_{11}\eta_1 + \alpha w_{12}\eta_2 < 0$, ie, [$\mathbf{IC}^\star$] holds as a strict inequality. Notice that $\alpha f_{11} w_{11} + \alpha f_{12} w_{12} < 0$ though it may be smaller than $v_1 < 0$. Nevertheless by scaling any such $\mathbf{w}_1$, we can find $\mathbf{w}_1^* \in K$ so that $\alpha f_{11} w_{11}^* + \alpha f_{12} w_{12}^* > v_1$. Then, choose x_1 so that [$\mathbf{PK}_1$] holds with equality. Thus, we have found $(x_1, \mathbf{w}_1^*)$ that satisfies [$\mathbf{PK}_1$] and [**IC**] and also satisfies $(x_1, \mathbf{w}_1^*) \ll \mathbf{0}$. Of course, we can always find $(x_2, \mathbf{w}_2) \in \operatorname{int} K_1$ that satisfies [$\mathbf{PK}_2$]. This shows that $\mathscr{B}^\circ(\operatorname{int} K_1) = \operatorname{int} K_1$.

Case 2: $\eta_1 < 0$. Notice that $\eta_2 < 1$. We claim that $\eta_1 + \eta_2 > 0$. To see this, observe that $\eta_2 + \eta_1 = f_{22} + f_{21} - t(f_{12} + f_{11}) = 1 - t > 0$ because $t = \theta_2/\theta_1 < 1$ by definition. This also implies that $1 > \eta_2 > -\eta_1 > 0$, and in particular, $0 < (-\eta_1)/\eta_2 < 1$.

Notice that the cone $\{\mathbf{w}_1 \in K_1 : w_{11} < w_{12} < w_{11}(-\eta_1)/\eta_2\}$ is open and convex. Observe that for any $\mathbf{w}_1$ in this cone, we have $\alpha f_{11} w_{11} + \alpha f_{12} w_{12} < 0$. Now, choose $\mathbf{w}_1$ in this cone so that $\alpha f_{11} w_{11} + \alpha f_{12} w_{12} > v_1$. Such a choice is always possible by scaling $\mathbf{w}_1$ suitably. Let $\mathbf{w}_1^*$ be such a choice and notice that it satisfies [$\mathbf{IC}^\star$] as a strict inequality. Let $\theta_1 x_1 := v_1 - \big(\alpha f_{11} w_{11} + \alpha f_{12} w_{12}\big)$ so that $x_1 < 0$. As before, we can choose $(x_2, \mathbf{w}_2) \in \operatorname{int} K_1$ that satisfies [$\mathbf{PK}_2$]. This shows that $\mathscr{B}^\circ(\operatorname{int} K_1) = \operatorname{int} K_1$ in this case too, which proves the claim. □

Here is a more direct proof that K_1 is a fixed point of $\mathscr{B}$.

Lemma B.5. The operator $\mathscr{B}$ satisfies $\mathscr{B}(K_1) = K_1$. Moreover, $\mathscr{B}^\circ(K_1) = \operatorname{int} K_1$.

Proof. We have already seen in Lemma B.4 that every $\mathbf{v} \in \operatorname{int} K_1$ is in $\mathscr{B}(K_1)$. If $\mathbf{v} = \mathbf{0}$, it is easy to see that $(x_i, \mathbf{w}_i) = \mathbf{0} \in K_1$ implements $\mathbf{0}$. Moreover, $\mathbf{0} \notin \mathscr{B}^\circ(K_1)$.

That $\mathscr{B}^\circ(K_1) = \operatorname{int} K_1$ is not difficult to show. An explicit construction is provided in Lemmas G.4, G.5, and G.6, culminating in Proposition G.7. □

We now conclude by proving that $V = \operatorname{int} K_1$ is the relevant domain.

Lemma B.6. The set $\operatorname{int} K_1$ is the largest fixed point of $\mathscr{B}^\circ$. The set K_1 is the largest fixed point of $\mathscr{B}$.

Proof. For any $K \subset \mathbb{R}^2_-$, it is easy to see that $\mathscr{B}^\circ(K) \subset \mathscr{B}(K)$. Lemma B.1 shows that $\mathscr{B}(\mathbb{R}^2_-) \subsetneq \mathbb{R}^2_-$. But the monotonicity of the operators implies that this holds upon iterations too, ie, we have $\mathscr{B}^{\circ n}(\mathbb{R}^2_-) \subset \mathscr{B}^n(\mathbb{R}^2_-)$. More importantly, we have

$$\mathscr{B}^\circ\left(\mathscr{B}^n(\mathbb{R}^2_-)\right) \subset \mathscr{B}\left(\mathscr{B}^n(\mathbb{R}^2_-)\right)$$

Lemma B.2 shows that $\mathscr{B}^n(\mathbb{R}^2_-)$ is a strictly decreasing sequence. The limit of this sequence is K_1 and Lemma B.5 shows it is a fixed point of $\mathscr{B}$. However, Lemma B.5 also shows that K_1 cannot be a fixed point of $\mathscr{B}^\circ$. Indeed, it shows that $\mathscr{B}^\circ(K_1) = \operatorname{int} K_1$. This establishes that $\operatorname{int} K_1$ is the largest fixed point of $\mathscr{B}^\circ$, which completes the proof. □

We have established that $V = \operatorname{int} K_1$ is the domain for our analysis.

C. The Full-Information Contract

Consider the case in which the principal can directly observe the agent's endowment shocks. This is the *first-best (full information) problem*. We will begin by studying the *recursive* version of this problem, which is given by

[RP-FB]
$$\hat{Q}^*(\mathbf{v}) := \inf_{\xi \in \Xi^{\mathrm{FB}}(\mathbf{v})} R(\tilde{x}_\xi)$$

where the space of recursive contracts $\Xi^{\mathrm{FB}}(\mathbf{v})$ is defined analogously to $\Xi(\mathbf{v})$ in Section 3, minus the appropriate incentive compatibility constraints. In this appendix, we will establish useful properties of the value function $\hat{Q}^*$, show that [**RP-FB**] has a unique solution, and show that this solution also solves the sequence version of the first-best problem.

C.1. Some Useful Bounds

In this section, we establish some useful bounds and limit properties of $\hat{Q}^*$ and P that will be used extensively in later sections.

In working with the recursive version of the first-best problem, ie, with [**RP-FB**], for convenience we will sometimes drop the ξ-dependence of the processes $(x^{(t)})$ and $(\mathbf{v}^{(t)})$.

Using the recursive promise-keeping constraints

[PK$_1$]
$$v_1 = \theta_1 x_1 + \alpha \mathbb{E}^{\mathbf{f}_1}[\mathbf{w}_1]$$

[PK$_2$]
$$v_2 = \theta_2 x_2 + \alpha \mathbb{E}^{\mathbf{f}_2}[\mathbf{w}_2]$$

to eliminate the instantaneous allocations x_i, we may therefore write the cost of contract $\tilde{x}$ along path $h \in \mathscr{H}$ as

$$R(\tilde{x}; h) = \sum_{t=0}^{\infty} \alpha^t C\left(\left|\frac{v_i^{(t)}(h) - \alpha \mathbb{E}^{\mathbf{f}_i}\left[\mathbf{w}_i^{(t)}(h)\right]}{\theta_i}\right|\right)$$

where, recall, $C(x) = -\log(-x)$. The expected cost of this contract is then $R(\tilde{x}) := \mathbb{E}\left[R(\tilde{x}; h)\right]$.

Lemma C.1. For any $\mathbf{v} \in V$, the first-best value function is finite, ie, $\hat{Q}^*(\mathbf{v}) \in \mathbb{R}$.

Proof. Let us begin by establishing a uniform lower bound for the pathwise lifetime cost function. Examining [**PK**$_1$] and [**PK**$_2$], it follows immediately from the negativity of the agent's utility that $w_{ij} > v_i/\alpha f_{ij}$ or, equivalently, $|w_{ij}| < \eta |v_i|$ where $\eta := 1/(\alpha \cdot \min_{ij} f_{ij})$. This bounds the growth rate of $\mathbf{v}^{(t)}$ along any feasible path.

Fix initial promise $\mathbf{v} \in V$, recursive contract $\xi \in \Xi^{\mathrm{FB}}(\mathbf{v})$, and path $h \in \mathcal{H}$. Along the path h, we have

$$C(x_i^{(t)}(h)) = -\log\left(\left|\frac{v_i^{(t)}(h) - \alpha \mathbb{E}^{\mathbf{f}_i}\left[\mathbf{w}_i^{(t)}(h)\right]}{\theta_i}\right|\right)$$
$$\geq -\log\left(\frac{1}{\theta_i} \cdot \left[|v_i^{(t)}(h)| + |w_{i1}^{(t)}(h)| + |w_{i2}^{(t)}(h)|\right]\right) \geq \log(\theta_i) - \log\left(\eta^t \cdot \max_i |v_i^{(0)}| \cdot [1 + 2\eta]\right)$$

where the first inequality follows from the triangle inequality and monotonicity of the $\log(\cdot)$ function, and the second inequality follows from iterating the bounds from the previous paragraph t times. This simplifies to

$$C(x_i^{(t)}(h)) \geq \log(\theta_i) - t\log(\eta) - \log(\max_i |v_i|) - \log(1 + 2\eta)$$

Note that the RHS is independent of the path $h \in \mathcal{H}$. Taking the α-discounted sum, it is easy to see that there exist $\kappa_1, \kappa_2 \in \mathbb{R}$ such that

$$\text{[C.1]} \qquad \inf_{\xi \in \Xi^{\mathrm{FB}}(\mathbf{v}),\ h \in \mathcal{H}} R(\tilde{x}_\xi; h) \geq \kappa_1 + \kappa_2 \log(\max_i |v_i|)$$

which is the desired uniform lower bound on pathwise costs. Because this lower bound holds for any feasible full-information contract, it follows immediately that $\hat{Q}^*(\mathbf{v}, s) > -\infty$. Moreover, we have shown in Proposition G.7 that the value function $P(\mathbf{v}, s)$ is bounded above by the real-valued function $Q^\xi(\mathbf{v}, s)$. Then, because $\Xi(\mathbf{v}) \subset \Xi^{\mathrm{FB}}(\mathbf{v})$, we have the string of inequalities $+\infty > P(\mathbf{v}, s) \geq \hat{Q}^*(\mathbf{v}, s) > -\infty$ for all $(\mathbf{v}, s) \in V \times S$, completing the proof. □

The proof of Lemma C.1 immediately yields the following corollary:

Corollary C.2. For all $(\mathbf{v}, s) \in V \times S$, we have $P(\mathbf{v}, s) \geq \hat{Q}^*(\mathbf{v}) > -\infty$.

Next, we bound the limiting behavior of the value functions:

Lemma C.3. For any initial promise $\mathbf{v} \in V$ and any recursive contract $\xi \in \Xi^{\mathrm{FB}}(\mathbf{v})$, we have

$$\liminf_{t \to \infty} \alpha^t \left[\inf_{h \in \mathcal{H}} \hat{Q}^*(\mathbf{v}^{(t)}(h), s^{(t)}(h))\right] \geq 0$$

Similarly, for any initial promise $\mathbf{v} \in V$ and any recursive contract $\xi \in \Xi(\mathbf{v})$, we have

$$\liminf_{t \to \infty} \alpha^t \left[\inf_{h \in \mathcal{H}} P(\mathbf{v}^{(t)}(h), s^{(t)}(h))\right] \geq 0$$

Proof. Because $\hat{Q}^*$ lies pointwise below P, it suffices to prove the limit condition for $\hat{Q}^*$. From equation [C.1] in the proof of Lemma C.1, we have

$$\hat{Q}^*(\mathbf{v}, s) \geq \kappa_1 + \kappa_2 \log(\max_i |v_i|)$$

for some $\kappa_1, \kappa_2 \in \mathbb{R}$ independent of $(\mathbf{v}, s) \in V \times S$. Let a path $h \in \mathscr{H}$ be given. The above inequality holds for all $(\mathbf{v}^{(t)}(h), s^{(t)}(h))$, ie,

$$\hat{Q}^*(\mathbf{v}^{(t)}(h), s^{(t)}(h)) \geq \kappa_1 + \kappa_2 \log(\max_i |v_i^{(t)}(h)|)$$

Iterating on the same growth-rate bounds as in the proof of Lemma C.1, we may bound from below the RHS of the above inequality by

$$\kappa_1 + \kappa_2 \log(\max_i |v_i^{(t)}(h)|) \geq \kappa_1 + \kappa_2 \cdot t \cdot \log(\eta) + \kappa_2 \cdot \log(\max_i |v_i^{(0)}|)$$

so that the RHS is independent of the path $h \in \mathscr{H}$. Then, using the fact that $\alpha^t \cdot t \to 0$, we have

$$\liminf_{t\to\infty} \alpha^t \left[\inf_{h \in \mathscr{H}} \hat{Q}^*(\mathbf{v}^{(t)}(h), s^{(t)}(h)) \right] \geq \liminf_{t\to\infty} \alpha^t \left[\kappa_1 + \kappa_2 \cdot t \cdot \log(\eta) + \kappa_2 \cdot \log(\max_i |v_i^{(0)}|) \right] = 0$$

which is the desired condition. □

C.2. Optimal Full-Information Contract

We may now characterize the unique full information optimal (recursive) contract. First, we define the feasible set for the full information problem. Namely, for each $\mathbf{v} \in \mathbb{R}^2_{--}$, we have

[**C.2**] $$\Gamma^{FB}(\mathbf{v}) := \left\{ (x_i, \mathbf{w}_i) \in (\mathbb{R}_{--} \times \mathbb{R}^2_{--})^S : [\mathbf{PK}_1],\ \ [\mathbf{PK}_2] \text{ are satisfied} \right\}$$

We begin with an important technical lemma.

Lemma C.4. The first-best value function $\hat{Q}^* : \mathbb{R}^2_{--} \times S \to \mathbb{R}$ satisfies the functional equation

[**C.3**] $$\hat{Q}^*(\mathbf{v}, s) = \inf_{(x_i, \mathbf{w}_i) \in \Gamma^{FB}(\mathbf{v})} \sum_{i=1,2} f_{si} \left[C(x_i) + \alpha \hat{Q}^*(\mathbf{w}_i, i) \right]$$

and each $\hat{Q}^*(\cdot, s)$ is convex and continuously differentiable. Namely, the infimum in [**C.3**] is attained at each $(\mathbf{v}, s) \in \mathbb{R}^2_{--} \times S$.

Proof. We omit the details, which are completely analogous to the arguments used to prove Theorem 2 presented in Appendix D. See, in particular, Lemma D.1 for the functional equation, Lemma D.3 for convexity, Lemmas D.4 and D.5 for attainment of the infimum, and Lemma D.8 for continuous differentiability. □

With Lemma C.4 in hand, problem [RP-FB] reduces to a smooth, convex, finite-dimensional minimization problem. Letting $\psi_1 \in \mathbb{R}$ and $\psi_2 \in \mathbb{R}$, respectively, denote the multipliers on the constraints [$\mathbf{PK}_1$] and [$\mathbf{PK}_2$], the Lagrangian is

[C.4]

$$\mathscr{L}^{FB} = \sum_{i=1,2} f_{si}\left[C(x_i) + \alpha \hat{Q}^*(\mathbf{w}_i, i)\right] + \psi_1\left(v_1 - \theta_1 x_1 - \alpha \mathbb{E}^{\mathbf{f}_1}[\mathbf{w}_1]\right) + \psi_2\left(v_2 - \theta_2 x_2 - \alpha \mathbb{E}^{\mathbf{f}_2}[\mathbf{w}_2]\right)$$

The (necessary and sufficient) optimality conditions consist of the envelope conditions

[$\mathbf{FB\text{-}Env}_1$] $$\hat{Q}_1^*(\mathbf{v}, s) = \psi_1$$

[$\mathbf{FB\text{-}Env}_2$] $$\hat{Q}_2^*(\mathbf{v}, s) = \psi_2$$

the first-order conditions for instantaneous utilities

[$\mathbf{FB\text{-}FOC}x_1$] $$f_{s1} C'(x_1) = \theta_1 \psi_1$$

[$\mathbf{FB\text{-}FOC}x_2$] $$f_{s2} C'(x_2) = \theta_2 \psi_2$$

and the first-order conditions for continuation utilities

[$\mathbf{FB\text{-}FOC}w_{11}$] $$f_{s1} \hat{Q}_1^*(\mathbf{w}_1, 1) = f_{11}\psi_1$$

[$\mathbf{FB\text{-}FOC}w_{12}$] $$f_{s1} \hat{Q}_2^*(\mathbf{w}_1, 1) = f_{12}\psi_1$$

[$\mathbf{FB\text{-}FOC}w_{21}$] $$f_{s2} \hat{Q}_1^*(\mathbf{w}_2, 2) = f_{21}\psi_2$$

[$\mathbf{FB\text{-}FOC}w_{22}$] $$f_{s2} \hat{Q}_2^*(\mathbf{w}_2, 2) = f_{22}\psi_2$$

It is easy to see that [$\mathbf{FB\text{-}FOC}w_{11}$] and [$\mathbf{FB\text{-}FOC}w_{12}$] combine to deliver

[C.5] $$\frac{\hat{Q}_1^*(\mathbf{w}_1, 1)}{f_{11}} = \frac{\hat{Q}_2^*(\mathbf{w}_1, 1)}{f_{12}}$$

and, similarly, that [$\mathbf{FB\text{-}FOC}w_{21}$] and [$\mathbf{FB\text{-}FOC}w_{22}$] combine to deliver

[C.6] $$\frac{\hat{Q}_1^*(\mathbf{w}_2, 2)}{f_{21}} = \frac{\hat{Q}_2^*(\mathbf{w}_2, 2)}{f_{22}}$$

These optimality conditions give us the following characterization of the full-information optimum.

Lemma C.5. Fix any initial promise $\mathbf{v} \in \mathbb{R}^2_{--}$. There exists a unique recursively optimal recursive contract. If the first shock is $s^{(0)} = s \in S$, the optimal contract satisfies

[C.7] $$x^{(t)}(\mathbf{v}, s^{(0)} = s) = \frac{(1-\alpha)\cdot v_s}{\theta^{(t)}}$$

[C.8] $$\mathbf{w}_i^{(t)}(\mathbf{v}, s^{(0)} = s) = (v_s, v_s) \qquad \text{for } i \in S$$

Moreover, the value function $\hat{Q}^*(\cdot, s)$ is strictly convex for each $s \in S$.

Proof. Fix $\mathbf{v} \in \mathbb{R}^2_{--}$ and define $\mathbf{v}^{(0)} = \mathbf{v}$. We will first show that there exists a contract of the form [C.8] that attains the infimum in [C.3] at each step. Suppose first that $s^{(0)} = 1$; the proof for $s^{(0)} = 2$ is completely symmetric. Combining [C.5] with [$\mathbf{FB\text{-}Env}_1$] and [$\mathbf{FB\text{-}Env}_2$] in the next period, we have

$$\psi_1^{(1)}(s^{(0)} = 1) = \frac{f_{11}}{f_{12}} \psi_2^{(1)}(s^{(0)} = 1)$$

Plugging this into [**FB-FOC**x_1] and [**FB-FOC**x_2] in period $t = 1$, we obtain

$$\frac{C'(x_1^{(1)}(s^{(0)} = 1))}{\theta_1} = \frac{C'(x_2^{(1)}(s^{(0)} = 1))}{\theta_2}$$

Now, if $s^{(1)} = 1$, then it is easy to see that $\psi_i^{(2)} = \psi_i^{(1)}$ for $i \in S$, so that the equations in the above displays will continue to hold in period $t = 2$. If instead $s^{(1)} = 2$, then it is easy to see that

$$\psi_1^{(2)} = \frac{f_{21}}{f_{12}}\psi_2^{(1)} = \frac{f_{21}}{f_{11}}\psi_1^{(1)}$$

and, similarly,

$$\psi_2^{(2)} = \frac{f_{22}}{f_{12}}\psi_2^{(1)}$$

Plugging these expressions into [**FB-FOC**x_1] and [**FB-FOC**x_2] at period $t = 2$ gives us the conditions

$$\frac{C'(x_1^{(2)})}{\theta_1} = \frac{1}{f_{11}}\psi_1^{(1)}$$

and

$$\frac{C'(x_1^{(2)})}{\theta_2} = \frac{1}{f_{12}}\psi_2^{(1)}$$

Combining the above expressions and iterative forward, it is straightforward to see that, regardless of the sequence of shocks and for any $t \geq 0$ and $k \geq 0$, we always have

[**C.9**]
$$\frac{C'(x_i^{(t)}(s^{(0)} = 1))}{\theta_i} = \frac{C'(x_j^{(t+k)}(s^{(0)} = 1))}{\theta_j}$$

Using the fact that $C(x) = -\log(x)$, it follows that, conditional on $s^{(0)} = 1$, the process $\theta^{(t)}x^{(t)}$ is constant. Plugging this into the promise-keeping constraints then immediately implies that, conditional on $s^{(0)} = 1$, the process $\mathbf{v}^{(t)}$ is constant for $t \geq 1$ and, in particular, $v_i^{(t)} = v_1^{(0)}$. Using these facts and the logarithmic form of the cost function then delivers the form of the optimal contract for $s = 1$. As noted above, the proof for $s = 2$ is symmetric.

Now, uniqueness will follow from strict convexity of the value function. To prove strict convexity, notice that the contract described by [**C.8**] is *stationary*, in the sense that the continuation promises self-generate at a single point. Therefore, the same argument as in the proof of Lemma D.11 shows that $\hat{Q}^*$ is strictly convex.

Thus, there exists a unique contract that attains the infimum in [**C.3**] at each step. To show that this contract is recursively optimal, in the sense that it attains the infimum in [**RP-FB**], we may follow the same argument as in the proof of Lemma D.6. □

It remains to show (i) that $\hat{Q}^*$ is strictly increasing the direction $(1, 1)$, and (ii) that the optimal contract and value function from [**RP-FB**] coincide with the optimal contract and value function from the *full-information sequence problem*

[**SP-FB**]
$$Q^*(\mathbf{v}) := \inf_{\tilde{x} \in \Pi^{\mathrm{FB}}(\mathbf{v})} R(\tilde{x})$$

where $\Pi^{FB}(\mathbf{v})$ is defined analogously to $\Pi(\mathbf{v})$ from [**SP**], minus the incentive constraint [**S-IC**].

Lemma C.6. The value functions for [**SP-FB**] and [**RP-FB**] are equal, ie, $Q^* = \hat{Q}^*$. They are finite on $\mathbb{R}^2_{--}$, strictly convex, and are strictly increasing in direction $(1,1)$, ie, for any $(\mathbf{v}, s) \in \mathbb{R}^2_{--} \times S$ and $\varepsilon > 0$,

$$[\textbf{C.10}] \qquad Q^*(\mathbf{v}, s) > Q^*(\mathbf{v} - (\varepsilon, \varepsilon), s).$$

Moreover, the recursively optimal contract from Lemma C.5 solves [**SP-FB**], ie, it is globally optimal.

Proof. First, we show that the optimal recursive contract attains the infimum in [**SP-FB**]. Because the optimal recursive contract induces a promised utility process that self-generates at a single point, it satisfies [**TVC**]. It follows from the Bounded Convergence Theorem that the induced allocation process $(x^{(t)})$ satisfies [**S-PK$_1$**] and [**S-PK$_2$**], and thus is an elements of $\Pi^{FB}(\mathbf{v})$. Thus, it must attain the infimum in [**SP-FB**], as it is feasible in that problem and solves the relaxed problem, [**RP-FB**]. That the value functions coincide then follows immediately, and Q^* inherits all the properties from $\hat{Q}^*$ and vice versa.

To see that $Q^*(\cdot, s)$ is strictly increasing in the direction $(1,1)$, notice that increasing $\mathbf{v}$ to $\mathbf{v}' + \varepsilon(1,1)$ for $\varepsilon > 0$ leads to a strict increase of all the contractual variables in [**C.8**]. Because $Q^*(\cdot, s)$ is strictly convex, it follows that $\mathsf{D}_{(1,1)} Q^*(\mathbf{v}, s) > 0$ for all $(\mathbf{v}, s) \in \mathbb{R}^2_{--} \times S$. □

D. Properties of the Value Function and Optimal Contract

We will break the proofs of Theorem 2 and Proposition 3.2 into five blocks. First, Appendix D.1 establishes basic properties of the value function and the recursively optimal contract. Second, Appendix D.2 establishes additional properties that hold under Assumption 1. Third, Appendix D.3 establishes homogeneity properties of the value function and its partial derivatives, which follow from the assumption that the agent has CARA utility. Fourth, additional properties of the recursively optimal contract are established in Appendix D.4 by studying the *interim* version of the principal's optimization problem. (The intermediate results established in this section will also be used to prove Propositions H.1 and 5.1 in Appendix H.) Finally, Appendix D.5 puts all the preceding pieces together and proves the theorem and proposition.

D.1. Basic Properties

We will establish the relevant properties through a series of lemmas.

It is useful to define the Bellman operator $T : \mathbb{R}^{V\times S} \to \overline{\mathbb{R}}^{V\times S}$ by

$$[\textbf{T}] \qquad TQ(\mathbf{v}, s) := \inf_{(x_i, \mathbf{w}_i)_{i=1,2} \in \Gamma(\mathbf{v})} \sum_{i=1,2} f_{si} \left[C(x_i) + \alpha Q(\mathbf{w}_i, i) \right]$$

where $\mathbb{R}^{V\times S} := \{f : V \times S \to \mathbb{R}\}$ and similarly for $\overline{\mathbb{R}}^{V\times S}$, where $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. For fixed $\xi(\mathbf{v}, s) = (x_i, \mathbf{w}_i)_{i=1,2} \in \Gamma(\mathbf{v})$, define

$$[\textbf{H}] \qquad H(Q, \xi)(\mathbf{v}, s) := \sum_{i=1,2} f_{si} \left[C(x_i) + \alpha Q(\mathbf{w}_i, i) \right]$$

Lemma D.1. The principal's value function, P, satisfies the Bellman equation

[**BE**] $$P = TP$$

Proof. We have $P \leq Q^{\xi}$ by construction, Proposition shows that G.7 shows that $Q^{\xi} < +\infty$, and Corollary C.2 gives us $P > -\infty$. Hence, P is real-valued on $V \times S$. The lemma then follows from standard Principle of Optimality arguments. □

Lemma D.2. P is the smallest real-valued fixed point of the functional equation [**FE**] in the order interval $[Q^*, Q^{\xi}]$.

Proof. This follows from Lemma C.3 and an elementary adaptation of Proposition 2.1 from Kamihigashi (2014). In particular, Lemma C.3 implies that the analogue of Kamihigashi's condition 2.17 holds, with the upper bound $\overline{v}$ replaced by the lower bound Q^* and the sign of the limit term reversed, as is appropriate given our focus on minimization and his focus on maximization. □

Lemma D.3. $P(\cdot, s)$ is convex and continuous for each $s \in S$.

Proof. By Lemma C.1, we know that $P(\mathbf{v}, s) \in \mathbb{R}$ for all $(\mathbf{v}, s) \in V \times S$. In problem [**RP**], the constraint correspondence $\Xi : V \to \Gamma(V)$ has a convex graph, and the return function $C(\cdot)$ is strictly convex. Together, these facts imply that P is convex. Because V is open, this also implies that for each $s \in S$, $P(\cdot, s)$ is continuous. □

Lemma D.4. Fix $s \in S$. For any convergent sequence $\{\mathbf{v}^n\} \subset V$ with limit point $\mathbf{v} \in \partial V$, the boundary of V, we have $P(\mathbf{v}^n, s) \to +\infty$.

Proof. Let $s \in S$ be fixed throughout. The boundary of V is the union of two rays, $R_1 := \{\mathbf{v} \in \mathbb{R}^2_- : v_2 = 0\}$ and $R_2 := \{\mathbf{v} \in \mathbb{R}^2_- : v_1 = v_2\}$. (Note that $R_1 \cap R_2 = \{\mathbf{0}\}$.) Consider any sequence of contractual variables that are feasible along the sequence, ie, $\xi(\mathbf{v}^n, s) \in \Gamma(\mathbf{v}^n)$ for all $n \in \mathbb{N}$. We may write $\xi(\mathbf{v}^n, s) = (x_1^n, x_2^n, \mathbf{w}_1^n, \mathbf{w}_2^n) \leq \mathbf{0}$.

First, suppose the limit point $\mathbf{v} \in R_1$. Because each $\xi^n(\mathbf{v}^n, s) \in \Gamma(\mathbf{v}^n)$, the constraint [**PK$_2$**] must be satisfied for each n, namely

$$v_2^n = \theta_2 x_2^n + \alpha \mathbb{E}^{\mathbf{f}_2}\left[\mathbf{w}_2^n\right]$$

which implies that $v_2^n < \theta_2 x_2^n$ for all n, by negativity of the control variables. Hence, we must have $x_2^n \to 0$ along this sequence, which implies that the flow cost $C(x^n) \to +\infty$. Because the sequence is convergent, all of the control variables along the sequence must lie in a bounded set, and it is easy to see that this implies that the sequence $H(P, \xi(\mathbf{v}^n, s))$ is bounded below. (Refer to, eg, equation [**C.1**] in the proof of Lemma C.1.) It follows that $H(P, \xi(\mathbf{v}^n, s)) \to +\infty$. Now, because this is true for any sequence of feasible contractual variables, it follows that $TP(\mathbf{v}^n, s) \to +\infty$ as well. By Lemma D.1, we have $P(\mathbf{v}^n, s) \to +\infty$.

Second, suppose the limit point $\mathbf{v} \in R_2$. Because each $\xi^n(\mathbf{v}^n, s) \in \Gamma(\mathbf{v}^n)$, the constraint [**IC***] must be satisfied for each n, namely

$$v_2^n - v_1^n \geq (\theta_2 - \theta_1)x_1^n + \alpha\beta(w_{12}^n - w_{11}^n)$$

Because $\mathbf{w}_1^n \in V$ for all n from the definition of $\Gamma(\cdot)$ (see equation [3.1]) and $(\theta_2 - \theta_1)x_1^n > 0$ for all n, it follows that $x_1^n \to 0$ and $\mathbf{w}_1^n \to R_2$. Thus, $C(x_1^n) \to +\infty$; the remainder of the proof is analogous to the previous paragraph. □

Lemma D.5. For each $(\mathbf{v}, s) \in V \times S$, the infimum in [BE] is attained. That is, there exists a *conserving*[52] recursive contract ξ^*.

Proof. Let $(\mathbf{v}, s) \in V \times S$ be given. With continuity of P established in Lemma D.3, it suffices to show that we can restrict attention to a compact set. Existence of a minimizer will then follow immediately.

First, we use the promise-keeping constraints [$\mathbf{PK}_1$] and [$\mathbf{PK}_2$] to establish lower bounds on the control variables. Consider the constraint [$\mathbf{PK}_1$], and use negativity of x_1 and each w_{1i} to conclude that

$$f_{1i} w_{1i} > \langle \mathbf{w}_1, \mathbf{f}_1 \rangle > v_1/\alpha \qquad \text{for} \quad i \in \{1, 2\} \tag{D.1}$$

Similarly, [$\mathbf{PK}_2$] implies that

$$f_{2i} w_{2i} > \langle \mathbf{w}_2, \mathbf{f}_2 \rangle > v_2/\alpha \qquad \text{for} \quad i \in \{1, 2\} \tag{D.2}$$

We may analogously use these constraints and negativity of the $\mathbf{w}_i$ vectors to show that

$$\theta_i x_i > v_i \qquad \text{for} \quad i \in \{1, 2\} \tag{D.3}$$

For each $(\mathbf{v}, s) \in V \times S$, define the open, bounded rectangle

$$M(\mathbf{v}, s) := \left\{ (x_i, \mathbf{w}_i)_{i=1,2} \in \left(\mathbb{R}_{--} \times \mathbb{R}^2_{--} \right)^S \ : \ [\text{D.1}], [\text{D.2}], [\text{D.3}] \text{ hold} \right\}$$

By the above arguments, this rectangle contains the feasible set, ie, $\Gamma(\mathbf{v}, s) \subseteq M(\mathbf{v}, s)$. Hence, $\Gamma(\mathbf{v}, s)$ is bounded for each $(\mathbf{v}, s) \in V \times S$.

Now, define the set $K(\mathbf{v}, s) := \overline{\Gamma(\mathbf{v}, s) \cap L(\mathbf{v}, s)}$, which is the closure of $\Gamma(\mathbf{v}, s) \cap L(\mathbf{v}, s)$, where

$$L(\mathbf{v}, s) := \left\{ \xi(\mathbf{v}, s) := (x_i, \mathbf{w}_i)_{i=1,2} : H(P, \xi(\mathbf{v}, s)) \leq Q^{\zeta}(\mathbf{v}, s) \right\}$$

is the set of controls that are weakly better than the contract ζ defined in Appendix G. Clearly $K(\mathbf{v}, s)$ is compact, as it is closed by definition and we established above that $\Gamma(\mathbf{v}, s)$ is bounded. It is also nonempty because $\zeta(\mathbf{v}, s) \in K(\mathbf{v}, s)$ by construction. Hence, $H(P, \cdot)$, which is continuous, has a minimizer in $K(\mathbf{v}, s)$. It remains to verify that

$$\underset{\zeta(\mathbf{v},s) \in K(\mathbf{v},s)}{\arg\min} \quad H(P, \zeta(\mathbf{v}, s)) \subseteq \Gamma(\mathbf{v}, s)$$

If not, then it must be that some minimizer $\zeta^*(\mathbf{v}, s) \in \overline{\Gamma(\mathbf{v}, s)} \backslash \Gamma(\mathbf{v}, s)$. But this implies that at least one of the x_i or w_{ij} is equal to zero, while the other control variables are uniformly bounded below because $\zeta^*(\mathbf{v}, s) \in M(\mathbf{v}, s)$. If one of the $x_i = 0$, then $H(P, \zeta^*(\mathbf{v}, s)) = +\infty$ because $C(0) = 0$. If one of the $w_{ij} = 0$, then $H(P, \zeta^*(\mathbf{v}, s)) = +\infty$ by Lemma D.4. In either case, this implies that $\zeta^*(\mathbf{v}, s) \notin L(\mathbf{v}, s)$ because, as established in Proposition G.7, $Q^{\zeta}(\mathbf{v}, s) < +\infty$ for all $(\mathbf{v}, s) \in V \times S$. □

(52) Following Kreps (2012), we say a contract is *conserving* if it attains the optimum in Bellman's equation at each step.

Lemma D.6. Any conserving contract ξ^* is recursively optimal in the principal's recursive problem [**RP**]. Hence, there exists a recursively optimal contract.

Proof. Lemma D.5 shows that a conserving contract exists, so it suffices to show that any such contract is recursively optimal for the principal. But recursively optimality is a direct consequence of Lemma C.3 and an easy adaptation of Theorem 4.5 in Stokey, Lucas and Prescott (1989). In particular, the limit condition for P described in Lemma C.3 is exactly the analogue of the hypothesis in SLP's Theorem 4.5, with limsup changed to liminf and the sign of the limit term flipped, to reflect our focus on minimization (versus their focus on maximization). □

Lemma D.7. If ξ^* is a globally optimal recursive contract, then $\tilde{x}_{\xi^*}$ solves [**SP**].

Proof. Let ξ^* be a globally optimal recursive contract. By the definition of incentive compatibility, truth-telling is an *unimprovable* strategy for the agent, ie, it cannot be improved up by any one-shot deviations. By the definition of sequential optimality, ξ^* is [**TVC**]-implementable. Thus, the utility process for the agent induced by ξ^* is *lower convergent*, and thus any unimprovable strategy is in fact optimal for the agent (see Proposition A6.7 in Kreps (2012)). Hence, truth-telling is an optimal strategy for the agent, which is equivalent to saying that $\tilde{x}_{\xi^*}$ satisfies [**S-IC**]. Moreover, a straightforward application of the Bounded Convergence Theorem implies that [**S-PK$_1$**] and [**S-PK$_2$**] also hold. Hence, $\tilde{x}_{\xi^*} \in \Pi(\mathbf{v})$, and so is feasible in [**SP**]. Optimality in [**SP**] follows from the fact that [**RP**] is a relaxation of [**SP**]. □

Lemma D.8. $P(\cdot, s)$ is continuously differentiable for each $s \in S$.

Proof. From Lemma D.5, the infimum in the Bellman equation is attained. The claim then follows from a straightforward adaptation of the theorem of Benveniste and Scheinkman (see, eg, Theorems 4.10 and 9.10 in Stokey, Lucas and Prescott (1989)). □

Lemma D.9. The directional derivative $\mathsf{D}_{(1,1)}P(\cdot, s)$ is non-negative for each $s \in S$.

Proof. If $P(\cdot, s)$ were non-decreasing in the direction $(1, 1)$, non-negativity of the directional derivative would follow directly from the definition

$$\mathsf{D}_{(1,1)}P(\mathbf{v}, s) := \lim_{\varepsilon \downarrow 0} \frac{P(\mathbf{v} + (\varepsilon, \varepsilon), s) - P(\mathbf{v}, s)}{\varepsilon}$$

(Note that Lemma D.8 implies that the directional derivative exists.) Hence, it suffices to show that $P(\cdot, s)$ is non-decreasing in direction $(1, 1)$. Corollary C.6 shows that the first-best value function Q^* is non-decreasing (indeed, strictly increasing) in this direction. We will show that P inherits this property from Q^*. The proof is order-theoretic. (In particular, it does *not* rely on convergence of $T^n Q^*$, the n-fold iterate of the Bellman operator [**T**] on Q^*, to P in countably-many steps; we are not able to show that such convergence takes place.)

Let $[Q^*, Q^\xi]$ denote the order interval (in the pointwise order) of functions $Q : V \times S \to \mathbb{R}$ that lie weakly above Q^* and weakly below Q^ξ. Let $\Phi := \{Q \in [Q^*, Q^\xi] : Q(\mathbf{v}, s) \geq Q(\mathbf{v} - (\varepsilon, \varepsilon), s)\ \forall \varepsilon > 0\}$. That is, Φ consists of all functions in the order interval $[Q^*, Q^\xi]$ with the property that they are non-decreasing in the direction $(1, 1)$.

Claim 1: Φ is a lattice in the pointwise order.

Proof. It is easy to see that if $F, G \in \Phi$, then $F \vee G, F \wedge G \in [Q^*, Q^\xi]$. Now, fix $(\mathbf{v}, s) \in V \times S$ and $\varepsilon > 0$. If F and G are ordered the same way at $(\mathbf{v}, s)$ and $(\mathbf{v} + (\varepsilon, \varepsilon), s)$, there is nothing left to prove. So suppose, without loss of generality, that $F(\mathbf{v}, s) \geq G(\mathbf{v}, s)$ and $G(\mathbf{v} + (\varepsilon, \varepsilon), s) \geq F(\mathbf{v} + (\varepsilon, \varepsilon), s)$. Then,

$$(F \wedge G)(\mathbf{v} + (\varepsilon, \varepsilon), s) = F(\mathbf{v} + (\varepsilon, \varepsilon), s) \geq F(\mathbf{v}, s) \geq (F \wedge G)(\mathbf{v}, s)$$

Similarly,

$$(F \vee G)(\mathbf{v} + (\varepsilon, \varepsilon), s) \geq F(\mathbf{v} + (\varepsilon, \varepsilon), s) \geq F(\mathbf{v}, s) = (F \vee G)(\mathbf{v}, s)$$

which concludes the proof. □

Claim 2: The lattice Φ is complete.

Proof. Let $F \subseteq \Phi$ be nonempty and define $\overline{f}(\mathbf{v}, s) := \sup_{f \in F} f(\mathbf{v}, s)$ and $\underline{f}(\mathbf{v}, s) := \inf_{f \in F} f(\mathbf{v}, s)$ for each $(\mathbf{v}, s) \in V \times S$. We show that $\overline{f} \in F$; the proof for $\underline{f}$ is symmetric. Suppose towards a contradiction that there exists $(\mathbf{v}, s) \in V \times S$ and some $\varepsilon > 0$ such that $(\mathbf{v}', s) \in V \times S$, where $\mathbf{v}' = \mathbf{v} + (\varepsilon, \varepsilon)$, and $\overline{f}(\mathbf{v}, s) > \overline{f}(\mathbf{v}', s)$. Because every function in Φ is bounded by $[Q^*, Q^\xi]$, both of which are finite, this implies that there exists some $\delta > 0$ such that

$$\overline{f}(\mathbf{v}, s) - \delta \geq \overline{f}(\mathbf{v}', s)$$

By definition of the supremum, there exists some $f \in F$ such that $f(\mathbf{v}, s) > \overline{f}(\mathbf{v}, s) - \delta$. Combined with the above display and the definition of Φ, this implies that

$$f(\mathbf{v}, s) > \overline{f}(\mathbf{v}'s,) \geq f(\mathbf{v}', s)$$

which contradicts the fact that f is non-decreasing in the direction $(1, 1)$ by virtue of $f \in F \subset \Phi$. □

Claim 3: $T : \Phi \to \Phi$ is well-defined and monotone.

Proof. Monotonicity is standard. It is easy to see that for $Q \in \Phi$, $TQ \in [Q^*, Q^\xi]$. All that remains is to show that $TQ \in \Phi$.

To see this, fix $(\mathbf{v}, s) \in V \times S$ and $\delta > 0$, and let $(x_s, \mathbf{w}_s)_{s=1,2}$ be a δ-optimal pair for the Bellman operator. Then,

$$\begin{aligned}
\delta + TQ(\mathbf{v}, s) &\geq \sum_{i=1,2} f_{si}\big[C(x_s) + \alpha Q(\mathbf{w}_s, s)\big] \\
&\geq \sum_{i=1,2} f_{si}\big[C(x_s) + \alpha Q(\mathbf{w}_s - \frac{\varepsilon}{\alpha}(1, 1), s)\big] \\
&\geq TQ(\mathbf{v} - (\varepsilon, \varepsilon), s)
\end{aligned}$$

where the second first inequality uses the fact that $Q \in \Phi$ and the second inequality follows because $(x_s, \mathbf{w}_s - \varepsilon/\alpha(1, 1)) \in \Gamma(\mathbf{v} - (\varepsilon, \varepsilon))$. It follows that $TQ \in \Phi$, since $\delta > 0$ was arbitrary. This proves that T is well defined. □

Now, because Φ is a complete lattice by Claims 1 and 2 and T is well-defined and monotone on Φ by Claim 3, it follows from Tarski's Fixed Point Theorem that the Bellman operator T has a fixed point in Φ. Let $\hat{P}$ be the smallest fixed point of T in Φ. It is easy to see that $\hat{P}$ is also the smallest fixed point of T in $[Q^*, Q^\xi]$, as $Q^* \in \Phi$ by Corollary C.6. Lemma D.2 therefore implies that $\hat{P} = P$, so that $P \in \Phi$. □

D.2. Strict Properties

Lemma D.10. Let ξ^* be globally optimal in [**RP**]. Then, ξ^* *delivers promises*, ie, the process $(\mathbf{v}^{(t)})$ which it induces satisfies

$$\text{[DP]} \qquad \lim_{t\to\infty} \alpha^t \mathbb{E}\left[\mathbf{v}^{(t)}\right] = 0$$

Proof. Under the assumption of [**TVC**]-implementability (which follows from sequential optimality), we may apply the Bounded Convergence Theorem to the process $(\mathbf{v}^{(t)})$. Property [**DP**] follows immediately. □

Lemma D.11. Let ξ^* be recursively optimal in [**RP**], and suppose that it satisfies [**DP**]. Then, the value function $P(\cdot, s)$ is strictly convex for each $s \in S$. In particular, this implies that $P(\cdot, s)$ is strictly convex under Assumption 1.

Proof. The part of the lemma concerning Assumption 1 follows immediately from the first part and Lemma D.10, so it suffices to establish the first part. Let ξ^* be a recursively optimal contract that satisfies [**DP**]. Moreover, by Lemma D.7, ξ^* attains the infimum in [**BE**] at each step.

We will show strict convexity by expanding the right-hand side of [**BE**]. For fixed $\eta \in (0,1)$, define $\mathbf{v}'' := \eta\mathbf{v} + (1-\eta)\mathbf{v}'$. Define $(x^{(t)}, \mathbf{w}_1^{(t)}, \mathbf{w}_2^{(t)})_{t\geq 1}$ and $(x'^{(t)}, \mathbf{w}_1'^{(t)}, \mathbf{w}_2'^{(t)})_{t\geq 1}$ to be the policies induced by ξ^* starting from $\mathbf{v}$ and $\mathbf{v}'$, respectively. Let $(x''^{(t)}, \mathbf{w}_1''^{(t)}, \mathbf{w}_2''^{(t)})_{t\geq 1}$ denote the history-wise convex combination of those two processes. Because ξ^* is incentive compatible and satisfies [**DP**] and Γ has a convex graph, this new process is also incentive compatible and satisfies [**DP**]. Moreover, $\mathbf{v} \neq \mathbf{v}'$ and the fact that ξ^* satisfies [**DP**] together imply that there exists some T and public history $h^T := (s^{(0)}, \ldots, s^{(T)})$ such that $x^{(T)}(h^T) \neq x'^{(T)}(h^T)$. That is, after the sequence of shocks h^T, the induced allocations must differ. Clearly, every such h^T has positive probability because the Markov process for endowments is fully connected.

Now, expanding the right-hand side of [**BE**] for T periods, we have

$$\begin{aligned}
\eta P(\mathbf{v}, s) + (1-\eta)P(\mathbf{v}', s) &= \sum_{t=0}^{T} \alpha^t \mathbb{E}\left[\eta C(x^{(t)}) + (1-\eta)C(x^{(t)}) \middle| s\right] \\
&\quad + \alpha^T \mathbb{E}\left[\eta P(\mathbf{w}^{(T)}, s^{(T)}) + (1-\eta)P(\mathbf{w}'^{(T)}, s^{(T)}) \middle| s\right] \\
&> \sum_{t=0}^{T} \alpha^t \mathbb{E}\left[C(x''^{(t)}) \middle| s\right] + \alpha^T \mathbb{E}\left[P(\mathbf{w}''^{(T)}, s^{(T)}) \middle| s\right] \geq P(\mathbf{v}'', s)
\end{aligned}$$

The strict inequality follows from convexity of $P(\cdot, s)$ and strict convexity of $C(\cdot)$, and the weak inequality follows from the definition of $P(\cdot, s)$. This establishes strict convexity. □

Lemma D.12. Under Assumption 1, there exists a unique optimal contract, ξ^* (in both the recursive and sequential sense). It is continuous in $(\mathbf{v}, s)$.

Proof. Assumption 1 gives existence on a globally optimal recursive contract. Because $P(\cdot, s)$ is strictly convex from Lemma D.11, this optimal recursive contract must be unique. Hence, there can exist at most one globally optimal contract. Continuity in $(\mathbf{v}, s)$ follows from Berge's Theorem of the Maximum (see, eg, Chapter 3 of Stokey, Lucas and Prescott (1989)). □

Lemma D.13. Under Assumption 1, the directional derivative $\mathsf{D}_{(1,1)}P(\cdot,\cdot)$ is strictly positive on $V \times S$.

Proof. Lemma D.9 implies that the directional derivative is non-negative. To see that $\mathsf{D}_{(1,1)}P$ is strictly positive under Assumption 1, suppose there exists $(\mathbf{v}, s) \in V \times S$ such that $\mathsf{D}_{(1,1)}P(\mathbf{v}, s) = 0$. For this fixed state, define the function $f : \mathbb{R}_+ \to \mathbb{R}$ by $f(y) := P((v_1 - y, v_2 - y), s)$. It is strictly concave because P is strictly convex under Assumption 1 by Lemma D.11, and is non-increasing by Lemma D.9. Hence, the hypothesis that $\mathsf{D}_{(1,1)}P(\mathbf{v}, s) = 0$ implies that $f(y) \equiv f(0)$ for all $y \in \mathbb{R}_+$. But this contradicts strict convexity of P and is therefore impossible. It follows that $\mathsf{D}_{(1,1)}P$ is strictly positive on $V \times S$. □

D.3. Homogeneity Properties

We state here the main homogeneity properties of the value function, its derivatives, and the recursively optimal policy.

Lemma D.14. For all $a \in \mathbb{R}$:

(a) The principal's value function satisfies the homogeneity property

[**D.4**] $$P(e^{-a}\mathbf{v}, s) = P(\mathbf{v}, s) + \frac{a}{1-\alpha}$$

(b) For fixed $s \in S$, the recursively optimal contract is homogenous of degree one in $\mathbf{v}$

[**D.5**] $$\xi(e^{-a}\mathbf{v}, s) = e^{-a}\xi(\mathbf{v}, s)$$

Proof. Item (a) is immediate from the fact that the return function satisfies

[**D.6**] $$C(e^{-a}x) = -\log(-e^{-a}x) = -\log(-x) - \log(e^{-a}) = C(x) + a$$

Part (b) follows from part (a) and the fact that gr Γ is a convex cone. □

We also have the following useful homogeneity property for the derivatives of the value function:

Lemma D.15. For all $a \in \mathbb{R}$:

(a) The partial derivatives P_i are homogenous of degree -1

[**D.7**] $$P_i(e^{-a}\mathbf{v}, s) = e^{a}P_i(\mathbf{v}, s)$$

(b) The differential martingale $\mathsf{D}_{(1,1)}P$ is homogenous of degree -1

[**D.8**] $$\mathsf{D}_{(1,1)}P(e^{-a}\mathbf{v}, s) = e^{a}\mathsf{D}_{(1,1)}P(\mathbf{v}, s)$$

(c) The efficiency sets E_1 and E_2 defined in Proposition 5.1 are rays in V.

Proof. For part (a), let $\varepsilon > 0$ and consider the following argument for $P_1(\cdot, s)$; the result for $P_2(\cdot, s)$ follows analogously. We have:

$$\begin{aligned}
P_1(e^{-a}\mathbf{v}, s) &= \lim_{\varepsilon \to 0} \frac{P(e^{-a}\mathbf{v} + (\varepsilon, 0), s) - P(e^{-a}\mathbf{v}, s)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{P(e^{-a}\left[\mathbf{v} + (e^{a}\varepsilon, 0)\right], s) - P(e^{-a}\mathbf{v}, s)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{P(\mathbf{v} + (e^{a}\varepsilon, 0), s) - P(\mathbf{v}, s)}{\varepsilon} \qquad \text{by part (a) of Lemma D.14} \\
&= e^{a} \cdot \lim_{\varepsilon \to 0} \left[\frac{P(\mathbf{v} + (e^{a}\varepsilon, 0), s) - P(\mathbf{v}, s)}{e^{a}\varepsilon}\right] = e^{a} P_1(\mathbf{v}, s)
\end{aligned}$$

Parts (b) and (c) are immediate corollaries. □

D.4. The Interim Contracting Problems

In this appendix, we study the *interim* version of the principal's problem. This perspective allows us to more easily prove several results about the recursively optimal contract and the relevant Lagrange multipliers. The results contained herein are used in Appendix D.5 to complete the proofs of Theorem 2 and Proposition 3.2, in Appendix H to complete the proofs of Propositions H.1 and 5.1.

§ D.4.1. *Formalities*

Because we are using interim state and control variables, the principal's optimization can be carried out at the interim stage. Consider the *interim problems*

[FE-Q^1] $\qquad Q^1(\mathbf{v}) = \min_{(x_1, \mathbf{w}_1)} \left[C(x_1) + \alpha P(\mathbf{w}_1, 1)\right]$

[PK$_1$-Q^1] $\qquad$ subject to $\quad v_1 = \theta_1 x_1 + \alpha \mathbb{E}^{\mathbf{f}_1}\left[\mathbf{w}_1\right]$

[IC*-Q^1] $\qquad v_2 - v_1 \geq (\theta_2 - \theta_1)x_1 + \alpha\beta(w_{12} - w_{11})$

and

[FE-Q^2] $\qquad Q^2(v_2) = \min_{(x_2, \mathbf{w}_2)} \left[C(x_2) + \alpha P(\mathbf{w}_2, 2)\right]$

[PK$_2$-Q^2] $\qquad$ subject to $\quad v_2 = \theta_2 x_2 + \alpha \mathbb{E}^{\mathbf{f}_2}\left[\mathbf{w}_2\right]$

These problems have a close connection to the recursive problem

[FE] $\qquad P(\mathbf{v}, s) = \min_{(x_i, \mathbf{w}_i)_{i=1,2} \in \Gamma(\mathbf{v})} \sum_{i=1,2} f_{si}\left[C(x_i) + \alpha P(\mathbf{w}_i, i)\right]$

specified in Theorem 2, as we see in the following lemmas.

Lemma D.16. The functions Q^i defined in [**FE-Q^1**] and [**FE-Q^2**] satisfy $P(\mathbf{v}, s) = f_{s1}Q^1(\mathbf{v}) + f_{s2}Q^2(v_2)$ for all $(\mathbf{v}, s) \in V \times S$. Moreover, a tuple $(x_i, \mathbf{w}_i)_{i=1,2}$ is a minimizer in [**FE**] at $(\mathbf{v}, s)$ if and only if $(x_1, \mathbf{w}_1)$ is a minimizer in [**FE-Q^1**] at $\mathbf{v}$ and $(x_2, \mathbf{w}_2)$ is a minimizer in [**FE-Q^2**] at v_2.

Proof. Note that the constraints [**PK**$_1$] and [**IC***] are independent of $(x_2, \mathbf{w}_2)$ and, similarly, the constraint [**PK**$_2$] is independent of $(x_1, \mathbf{w}_1)$. Hence, we may re-write [**FE**] as

$$
\begin{aligned}
P(\mathbf{v}, s) &= \sum_{i=1,2} f_{si} \min_{(x_i, \mathbf{w}_i) \in \Gamma_i(\mathbf{v})} [C(x_i) + \alpha P(\mathbf{w}_i, i)] \\
&= f_{s1} Q^1(\mathbf{v}) + f_{s2} Q^2(v_2)
\end{aligned}
$$

where $\Gamma_1(\mathbf{v}) := \{(x_1, \mathbf{w}_1) \in \mathbb{R}_{--} \times V : [\mathbf{PK}_1] \text{ and } [\mathbf{IC}^*] \text{ are satisfied}\}$ and $\Gamma_2(\mathbf{v}) := \{(x_2, \mathbf{w}_2) \in \mathbb{R}_{--} \times V : [\mathbf{PK}_2] \text{ is satisfied}\}$. The lemma follows immediately. □

As an immediate consequence of Lemma D.16, we have the following fact about the recursively optimal contract.

Lemma D.17. The recursively optimal contract $\xi^* : (\mathbf{v}, s) \mapsto \Gamma(\mathbf{v})$ does not depend on $s \in S$. The policy functions $x_2 : V \to \mathbb{R}_{--}$ and $\mathbf{w}_2 : V \to V$ depend on $\mathbf{v} \in V$ only through v_2, its second component.

Proof. To see the recursively optimal contract does not depend on the previous report $s \in S$, notice that neither the objective functions nor the feasible sets in the problems [**FE-**Q^1] and [**FE-**Q^2] depend on this variable. Similarly, the feasible set in [**FE-**Q^2] depends on $\mathbf{v}$ only through its second component, v_2, so the function $\xi^*(\mathbf{v}, 2)$ depends only on this component. □

Another consquence of Lemma D.16 is the following characterization of the interim value functions Q^i.

Lemma D.18. The interim value functions Q^i are convex and continuously differentiable. Moreover, $Q^1(\cdot)$ is strictly increasing in v_1 and weakly decreasing in v_2, while $Q^2(\cdot)$ is strictly increasing in v_2. Under Assumption 1, each of the Q^i is strictly convex and $Q^1(\cdot)$ is strictly decreasing in v_2.

Proof. Convexity and continuous differentiability follow from the (assumed) strict convexity and continuous differentiability of $C(\cdot)$, and the convexity and continuous differentiability of $P(\cdot, s)$ established, respectively, in Lemmas D.3 and D.8, and standard dynamic programming arguments. The strict convexity under Assumption 1 follows similarly from Lemma D.11.

To see that $Q^2(\cdot)$ is strictly increasing in v_2, suppose that $(x_2^*, \mathbf{w}_2^*)$ is a minimizer in [**FE-**Q^2] at v_2. For any $\varepsilon > 0$ small enough, the policy $(x_2^* - \frac{\varepsilon}{\theta_2}, \mathbf{w}_2^*)$ is feasible in the problem [**FE-**Q^2] at $\hat{v}_2 := v_2 - \varepsilon$. (We can clearly take $\varepsilon > 0$ small enough so that $\hat{\mathbf{v}} \in V$.) Because this policy costs strictly less than $(x_2^*, \mathbf{w}_2^*)$, it follows that $Q^2(\hat{v}_2) < Q^2(v_2)$ for all $v_2 > \hat{v}_2$.

To see that $Q^1(\cdot)$ is strictly increasing in v_1, suppose that $(x_1^*, \mathbf{w}_1^*)$ is a minimizer in [**FE-**Q^1] at $\mathbf{v}$. For any $\varepsilon > 0$, the policy $(x_1^* - \frac{\varepsilon}{\theta_1}, \mathbf{w}_1^*)$ is feasible in the problem [**FE-**Q^1] at $\hat{\mathbf{v}} := (v_1 - \varepsilon, v_2)$. Clearly, this new policy satisfies [**PK**$_1$] at $\hat{\mathbf{v}}$. To see that it satisfies [**IC***], note that the change from $\mathbf{v}$ to $\hat{\mathbf{v}}$ increases the LHS of [**PK**$_1$] by $\varepsilon > 0$, while the change in policies increases the RHS of [**PK**$_1$] by $\varepsilon \cdot (\theta_1 - \theta_2)/\theta_1 < \varepsilon$. Because the new policy costs strictly less, the desired strict monotonicity follows.

Now, it is easy to see that $Q^1(\cdot)$ is weakly decreasing in v_2, as increasing v_2 enlarges the constraint set in problem [**FE-**Q^1]. As for strict monotonicity, suppose towards a contradiction that Assumption 1 holds and $Q^1(\cdot)$ is not strictly decreasing in v_2 at some $\mathbf{v} = (v_1, v_2) \in V$. The convexity of $Q^1(\cdot)$

then implies that $Q^1((v_1, v_2 - \varepsilon)) = Q^1(\mathbf{v})$ for all $\varepsilon > 0$ such that $(v_1, v_2 - \varepsilon) \in V$. But we have shown above that $Q^1(\cdot)$ is strictly convex under Assumption 1, delivering the desired contradiction. $\square$

For the interim problems, we will denote η_1 and σ to be the multipliers on [**PK**$_1$] and [**IC***], respectively, and let η_2 to be the multiplier on [**PK**$_2$]. The Lagrangians for problems [**FE-**Q^1] and [**FE-**Q^2], respectively, are

$$\begin{aligned} \mathscr{L}_1(\mathbf{v}) = C(x_1) + \alpha f_{11} Q^1(\mathbf{w}_1) + \alpha f_{12} Q^2(w_{12}) + \eta_1 \Big[v_1 - \theta_1 x_1 - \alpha \mathbb{E}^{\mathbf{f}_1}[\mathbf{w}_1] \Big] \\ -\sigma \Big[v_2 - v_1 - (\theta_2 - \theta_1) x_1 - \alpha\beta(w_{12} - w_{11}) \Big] \end{aligned} \tag{L-Q^1}$$

and

$$\mathscr{L}_2(v_2) = C(x_2) + \alpha f_{21} Q^1(\mathbf{w}_2) + \alpha f_{22} Q^2(w_{22}) + \eta_2 \Big[v_2 - \theta_2 x_2 - \alpha \mathbb{E}^{\mathbf{f}_2}[\mathbf{w}_2] \Big] \tag{L-Q^2}$$

where we have used Lemma D.16 to write P in terms of the Q^i. The following fact about the multipliers is useful:

Lemma D.19. Suppose Assumption 1 holds. For all $\mathbf{v} \in V$, at the optimum $(\eta_1(\mathbf{v}), \eta_2(v_2), \sigma(\mathbf{v})) \gg \mathbf{0}$.

Proof. Suppose $\eta_1(\mathbf{v}) \leq 0$ at the optimum. Consider the perturbation $\mathbf{w}_1 - (\varepsilon, \varepsilon)$ for some small $\varepsilon > 0$. Formally, the directional derivative of the Lagrangian $\mathscr{L}_1(\mathbf{v})$ with respect to $\mathbf{w}_1$ in the direction $(1,1)$ is $\alpha(\eta_1 - \mathsf{D}_{(1,1)} P(\mathbf{w}_1, 1))$. (Note that this perturbation does not affect the [**IC***] term.) By Lemma D.13, $D_{(1,1)} P(\mathbf{w}_1, 1) > 0$ under Assumption 1. Thus, the term $\alpha(\eta_1 - \mathsf{D}_{(1,1)} P(\mathbf{w}_1, 1))$ is strictly negative, contradicting optimality. Thus, $\eta_1(\mathbf{v}) > 0$ at the optimum. The proof for $\eta_2(v_2) > 0$ is similar. To see that $\sigma(\mathbf{v}) > 0$ at the optimum, recall that Lemma D.18 says that $Q^1(v_1, \cdot)$ is strictly decreasing, strictly convex, and continuously differentiable under Assumption 1, so that $Q^1_2(\mathbf{v}) < 0$ for all $\mathbf{v} \in V$. Moreover, envelope condition for v_2 in [**L-**Q^1] is

$$Q^1_2(\mathbf{v}) = -\sigma(\mathbf{v}) \tag{Env$_2$-Q^1}$$

so that $\sigma(\mathbf{v}) > 0$. $\square$

Recall from Section H.1 the Lagrangian associated with the value function P, ie, the *ex ante* Lagrangian:

$$\begin{aligned} \mathscr{L}(\mathbf{v}, s) = \sum_{i=1,2} f_{si} \left[C(x_i) + \alpha P(\mathbf{w}_i, i) + \frac{\lambda_i}{f_{si}} \left(v_i - \theta_i x_i - \alpha \mathbb{E}^{\mathbf{f}_i}[\mathbf{w}_i] \right) \right] \\ -\mu \left(\theta_2 x_2 + \alpha \mathbb{E}^{\mathbf{f}_2}[\mathbf{w}_2] - \theta_2 x_1 - \alpha \mathbb{E}^{\mathbf{f}_2}[\mathbf{w}_1] \right) \end{aligned} \tag{L-P}$$

The optimality conditions for this ex ante problem (see also Section H.1) consist of the envelope conditions

$$P_1(\mathbf{v}, s) = \lambda_1 \tag{Env1}$$

$$P_2(\mathbf{v}, s) = \lambda_2 \tag{Env2}$$

the first-order conditions for instantaneous utilities

[**FOC**x_1] $$f_{s1}C'(x_1) = \theta_1\lambda_1 - \theta_2\mu$$

[**FOC**x_2] $$f_{s2}C'(x_2) = \theta_2\lambda_2 + \theta_2\mu$$

and the first-order conditions for contingent continuation utilities

[**FOC**w_{11}] $$f_{s1}P_1(\mathbf{w}_1, 1) = f_{11}\lambda_1 - f_{21}\mu$$

[**FOC**w_{12}] $$f_{s1}P_2(\mathbf{w}_1, 1) = f_{12}\lambda_1 - f_{22}\mu$$

[**FOC**w_{21}] $$f_{s2}P_1(\mathbf{w}_2, 2) = f_{21}\lambda_2 + f_{21}\mu$$

[**FOC**w_{22}] $$f_{s2}P_2(\mathbf{w}_2, 2) = f_{22}\lambda_2 + f_{22}\mu$$

We may write the system of optimality conditions for the interim problems as follows. For [**L-**Q^1] we have

[**Env**$_1$**-**Q^1] $$Q_1^1(\mathbf{v}) = \eta_1 + \sigma$$

[**Env**$_2$**-**Q^1] $$Q_2^1(\mathbf{v}) = -\sigma$$

[**FOC**x_1**-**Q^1] $$\frac{C'(x_1)}{\theta_1} = \eta_1 + \frac{\Delta}{\theta_1}\sigma$$

[**FOC**w_{11}**-**Q^1] $$Q_1^1(\mathbf{w}_1) = \eta_1 + \frac{\beta}{f_{11}}\sigma$$

[**FOC**w_{12}**-**Q^1] $$Q_2^1(\mathbf{w}_1) + \frac{f_{12}}{f_{11}}Q_2^2(w_{12}) = \frac{f_{12}}{f_{11}}\eta_1 - \frac{\beta}{f_{11}}\sigma$$

and for [**L-**Q^2] we have

[**Env**$_2$**-**Q^2] $$Q_2'(v_2) = \eta_2$$

[**FOC**x_2**-**Q^2] $$\frac{C'(x_2)}{\theta_2} = \eta_2$$

[**FOC**w_{21}**-**Q^2] $$Q_1^1(\mathbf{w}_2) = \eta_2$$

[**FOC**w_{22}**-**Q^2] $$Q_2^2(w_{22}) + \frac{f_{21}}{f_{22}}Q_2^1(\mathbf{w}_2) = \eta_2$$

We can use the optimality conditions for the Lagrangians [**L-**Q^1], [**L-**Q^2], and [**L-P**] to relate the interim multipliers (η_1, η_2, σ) to the ex ante multipliers $(\lambda_1, \lambda_2, \mu)$.

Lemma D.20. At the optimum, the multipliers satisfy

[**D.9**] $$\frac{\lambda_1(\mathbf{v}, s)}{f_{s1}} = \eta_1(\mathbf{v}) + \sigma(\mathbf{v})$$

[**D.10**] $$\frac{\lambda_2(\mathbf{v}, s)}{f_{s2}} = \eta_2(v_2) - \frac{f_{s1}}{f_{s2}}\sigma(\mathbf{v})$$

[**D.11**] $$\frac{\mu(\mathbf{v}, s)}{f_{s1}} = \sigma(\mathbf{v})$$

As a consequence, under Assumption 1, $\mu(\mathbf{v}, s) > 0$ for all $(\mathbf{v}, s) \in V \times S$ at the optimum.

Proof. Lemma D.16 gives the identity $P(\cdot, s) = f_{s1} Q^1(\cdot) + f_{s2} Q^2(\cdot)$. Thus, it follows that

$$P_1(\mathbf{v}, s) = f_{s1} Q_1^1(\mathbf{v}) \quad \textbf{[D.12]}$$

$$P_2(\mathbf{v}, s) = f_{s1} Q_2^1(\mathbf{v}) + f_{s2} Q_2^2(v_2) \quad \textbf{[D.13]}$$

Now, combining [**D.12**] with the envelope conditions [**Env1**] and [**Env$_1$-Q^1**] yields

$$\lambda_1(\mathbf{v}, s) = f_{s1} \left(\eta_1(\mathbf{v}) + \sigma(\mathbf{v})\right)$$

which immediately delivers [**D.9**]. Similarly, combine [**D.13**] with the envelope conditions [**Env2**], [**Env$_2$-Q^1**], and [**Env$_2$-Q^2**] to get

$$\lambda_2(\mathbf{v}, s) = -f_{s1}\sigma(\mathbf{v}) + f_{s2}\eta_2(v_2)$$

which immediately delivers [**D.10**]. Finally, we may combine the FOCs [**FOCx_2**] and [**FOCx_2-Q^2**] to get

$$f_{s2}\eta_2(v_2) = \lambda_2(\mathbf{v}, s) + \mu(\mathbf{v}, s)$$

and then substitute [**D.10**] into this expression to get [**D.11**]. The assertion that μ is strictly positive on $V \times S$ follows directly from [**D.11**] and Lemma D.19. □

D.5. Putting the Pieces Together

We now have all the requisite pieces in place to prove Theorem 2 and Proposition 3.2.

Proof of Theorem 2. That P satisfies the functional equation [**FE**] follows from Lemmas D.1 and D.5. Convexity and continuous differentiability of $P(\cdot, s)$ follow from, respectively, Lemma D.3 and Lemma D.8. Part (a) of the theorem follows from Lemma D.2. The (non)-monotonicity results in part (b) follow from Lemmas D.16 and D.18. The behavior of P near the boundary of V in part (b) of the theorem follows from Lemma D.4. Part (c) follows from Lemma D.9. Parts (d) and (e) follow from Lemmas D.14 and D.15.

The different pieces of part (f) follows from several lemmas. The existence of an recursively optimal contract follows from Lemmas D.5 and D.6. That this contract can be taken to be non-random follows from convexity of $C(\cdot)$ and $P(\cdot, s)$. Lemma D.17 gives us independence of the recursively optimal contract from $s \in S$ and, for the components $(x_2, \mathbf{w}_2)$, independence from v_1.

Finally, part (g) of the theorem follows from Lemma D.7. □

Proof of Proposition 3.2. Part (a) follows from Lemma D.12. Part (b) is essentially a restatement of part (g) in Theorem 2, and thus follows from Lemma D.7. Part (c) follows from Lemma D.11, and part (d) follows from Lemma D.13. Part (e) follows from Lemma D.20. □

E. Proofs from Section 4

E.1. Proof of Theorem 3

Adding [**Env1**] and [**Env2**] delivers

$$\mathsf{D}_{(1,1)} P(\mathbf{v}, s) = \lambda_1 + \lambda_2$$

Combining [**FOC**w_{11}] with [**FOC**w_{12}] and [**FOC**w_{21}] with [**FOC**w_{22}] gives us, respectively,

$$f_{s1} \cdot \mathsf{D}_{(1,1)} P(\mathbf{w}_1, 1) = -\mu + \lambda_1$$

and

$$f_{s2} \cdot \mathsf{D}_{(1,1)} P(\mathbf{w}_2, 2) = \mu + \lambda_2$$

The martingale property follows immediately from the above expressions. Integrability follows from non-negativity (Lemma D.9) and the fact that the set of states S is finite. This complete the proof.

E.2. Proof of Theorem 4

We begin with a technical lemma.

Lemma E.1. Let $Y := \mathsf{D}P(V, 2) \subseteq \mathbb{R}^2$ be the image of V under the map $\mathsf{D}P(\cdot, 2)$. Then, under Assumption 1, Y is homeomorphic to V. In particular, Y is an open convex cone.

Proof. We first show that $\mathsf{D}\,P(\cdot, 2)$ is injective. To see this, notice first that because P is strictly convex, by Theorem 7.21 of Van Tiel (1984), $\mathsf{D}\,P(\cdot, 2)$ is *strictly monotone* in the sense that for all $\mathbf{v}, \mathbf{v}' \in V$ such that $\mathbf{v} \neq \mathbf{v}'$, $\langle \mathsf{D}\,P(\mathbf{v}, 2) - \mathsf{D}\,P(\mathbf{v}', 2), \mathbf{v} - \mathbf{v}' \rangle > 0$. But now suppose $\mathsf{D}\,P(\cdot, 2)$ is not injective, so that there are $\mathbf{v}, \mathbf{v}' \in V$ distinct such that $\mathsf{D}\,P(\mathbf{v}, 2) = \mathsf{D}\,P(\mathbf{v}', 2)$. But this would imply that $0 = \langle \mathbf{0}, \mathbf{v} - \mathbf{v}' \rangle = \langle \mathsf{D}\,P(\mathbf{v}, 2) - \mathsf{D}\,P(\mathbf{v}', 2), \mathbf{v} - \mathbf{v}' \rangle > 0$, which is a contradiction. It now follows Brouwer's Invariance of Domain Theorem — see, for instance, Theorem 2B.3 on p. 172 Hatcher (2001) — that V is homeomorphic to $\mathsf{D}(P, 2) = Y$.

By part (e) of Theorem 2, $\mathsf{D}P(\mathbf{v}, 2) = a\,\mathsf{D}\,P(a\mathbf{v}, 2)$. Therefore, if $\mathsf{D}(\mathbf{v}, 2) = \mathbf{y} \in Y$ for some $\mathbf{v} \in V$, then for all $a > 0$, $(1/a)\mathbf{y} \in Y$, which implies Y is an open cone.

We now prove that Y is convex. Towards this end, fix $\mathbf{y}_0, \mathbf{y}_1 \in Y$, and for each $s \in [0, 1]$, define $\mathbf{y}_s := (1 - s)\mathbf{y}_0 + s\mathbf{y}_1$, and $F_s := \{\lambda \mathbf{y}_s : \lambda > 0\}$. Because Y is a cone, $F_0, F_1 \subset Y$. Moreover, for all $s \in (0, 1)$, $F_s \subset \operatorname{conv}\{F_0, F_1\}$, the convex hull of F_0 and F_1.

Recall also that V is convex and hence path connected. Because Y is homeomorphic to V, Y is also path connected. This means that there exists a path $g : [0, 1] \to Y$ in Y such that $g(0) = \mathbf{y}_0$, $g(1) = \mathbf{y}_1$, and for all $t \in (0, 1)$, $g(t) \in Y$. Moreover, for all $s \in (0, 1)$, there exists $t \in (0, 1)$ such that $g(t) \in F_s$. To see this last fact, notice that by part (b) of Theorem 2, Y is contained in the open half-space $\{(x_1, x_2) \in \mathbb{R}^2 : x_1 > 0\}$. Thus, the set $Y \setminus \operatorname{conv}\{F_0, F_1\}$ is the union of two disjoint open sets and therefore not path connected. It follows that the path g must traverse $\operatorname{conv}\{F_0, F_1\}$, or else we would arrive at the contradiction that $[0, 1]$ is not path connected. Taken together, this implies that $F_s \subset Y$ for all $s \in (0, 1)$, which implies that $\mathbf{y}_s \in Y$ for all $s \in (0, 1)$, ie, Y is convex. □

Proof of Theorem 4. Theorem 3 shows that $\mathsf{D}_{(1,1)}P$ is a strictly positive martingale. By Doob's Martingale Convergence Theorem (see Theorem 2 on p. 517 of Shiryaev (1995)), it must converge $\mathbb{P}$-a.s. to a non-negative, $\mathbb{P}$-integrable random variable. We may discard the set of paths along which there are only finitely-many instances of consecutive θ_2 shocks, as this set is $\mathbb{P}$-null. (See Proposition K.2 for a proof.)

So fix an arbitrary path $(s^{(t)})_{t=0}^{\infty} \in S^{\infty}$ along which there are infinitely-many instances of consecutive θ_2 shocks and suppose, towards contradiction, that $\mathsf{D}_{(1,1)}P(\mathbf{v}^{(t)}, s^{(t)}) \to C > 0$ along this path. For any fixed $\varepsilon \in (0, C)$, there exists $T \in \mathbb{N}$ such that for all $t \geq T$, we have $\mathsf{D}_{(1,1)}P(\mathbf{v}^{(t)}, s^{(t)}) \in [C-\varepsilon, C+\varepsilon]$. By part (c) of Proposition H.1, which states that for all $\mathbf{v} \in V$ we have $\mathsf{D}_{(1,1)}P(\mathbf{v}, 2) \in E_2$, this implies that there exists a compact set $F \subset E_2$ such that $\mathsf{D}P(\mathbf{v}^{(t)}, s^{(t)}) \in F$ for all $t \geq T$ along this path for which $s^{(t)} = 2$. By compactness and properties of the chosen path, there exists a further subset of dates $(\tau_t)_{t=T}^{\infty}$ such that $s^{(\tau_t)} = s^{(\tau_t+1)} = 2$ and the subsequence $\mathsf{D}P(\mathbf{v}^{(\tau_t)}, s^{(\tau_t)}) \to \mathbf{a} \in F$.

Now, because F is compact, under Assumption 1 Lemma E.1 implies that that $\mathbf{v}^{(t)} \in G \subset V$ for some compact G, and thus $\mathbf{v}^{(\tau_t)} \to \mathbf{v} \in G \subset V$. Moreover, by continuity of the policy functions (see Proposition 3.2), we have $\mathbf{v}^{(\tau_t+1)} = \xi^c((\mathbf{v}^{(\tau_t)}, 2), 2) \to \xi^c((\mathbf{v}, 2), 2) =: \mathbf{w}_2$. But this implies that $\mathsf{D}_{(1,1)}P(\mathbf{v}, 2) = \mathsf{D}_{(1,1)}P(\mathbf{w}_2, 2)$, which is impossible by part (c) of Proposition H.1.

Hence, we must have $\mathsf{D}_{(1,1)}P(\mathbf{v}^{(t)}, s^{(t)}) \to 0$ along $\mathbb{P}$-a.e. path. By part (a) of Proposition H.1, this implies that, for this generic set of paths, $\mathsf{D}P(\mathbf{v}^{(n_t)}, s^{(n_t)}) \to (0,0)$ along the subsequence of dates (n_t) for which $s^{(n_t)} = 2$. We will show that convergence in fact occurs along the entire sequence. By Theorem 3, the directional derivative $\mathsf{D}_{(1,1)}P$ is strictly positive on $V \times S$. But part (b) of Proposition H.1 says that $\mathsf{D}P(\mathbf{w}_1, 1)$ always satisfies $P_1(\mathbf{w}_1, 1)/f_{11} \geq P_2(\mathbf{w}_1, 1)/f_{12}$. This implies that $P_1(\mathbf{w}_1, 1) > 0$. Now, recall the first-order condition

[FOCw_{11}] $$f_{s1}P_1(\mathbf{w}_1, 1) = f_{11}\lambda_1 - f_{21}\mu$$

Because $\mu > 0$ by part (e) of Proposition 3.2 (see also Lemma D.20), [**FOC**w_{11}] implies that

$$P_1(\mathbf{w}_1, 1) < \frac{f_{11}}{f_{s1}} \cdot \lambda_1(\mathbf{v}, s)$$

Indeed, starting from any date n_t at which $s^{(n_t)} = 2$, and for any number k of subsequent consecutive $s = 1$ shocks, we may iterate forward on the above display to get

$$P_1(\mathbf{v}^{(n_t+k)}, s^{(n_t+k)}) < \frac{f_{11}}{f_{21}} P_1(\mathbf{v}^{(n_t)}, s^{(n_t)})$$

Thus, the fact that $\lambda_1(\mathbf{v}^{(n_t)}, s^{(n_t)}) \to 0$ (recall from [**Env1**] that $\lambda_1(\mathbf{v}, s) = P_1(\mathbf{v}, s)$) and the above display together imply that $\lambda_1(\mathbf{v}^{(t)}, s^{(t)}) \to 0$ along this generic set of paths. Because $\mathsf{D}_{(1,1)}P$ is strictly positive and $P_1(\mathbf{v}^{(t)}, s^{(t)}) \to 0$, we also see that $P_2(\mathbf{v}^{(t)}, s^{(t)}) \to 0$ along these paths, which establishes part (i). From the envelope conditions [**Env1**] and [**Env2**], we have $(P_1, P_2)^{(t)} = (\lambda_1, \lambda_2)^{(t)} \to 0$. Finally, combining these observations with [**PW**] implies that $\mu(\mathbf{v}^{(t)}, s^{(t)}) \to 0$ as well.

Hence, for this generic set of paths, the entire vector of multipliers converges to zero, ie, $(\lambda_1, \lambda_2, \mu)^{(t)} \to (0,0,0)$. Recall the first-order conditions

[FOCx_1] $$f_{s1}C'(x_1) = \theta_1\lambda_1 - \theta_2\mu$$

[FOCx_2] $$f_{s2}C'(x_2) = \theta_2\lambda_2 + \theta_2\mu$$

Convergence of the multipliers and [**FOC**x_1] together imply that $x_1^{(t)} \to -\infty$. Similarly, convergence of the multipliers and [**FOC**x_2] together imply that $x_2^{(t)} \to -\infty$, establishing part (iii). Because the agent's utility function is bounded above, we must in turn have that $v_1^{(t)} \to -\infty$ and $v_2^{(t)} \to -\infty$, which proves part (iii). Finally, observe that by [**IC***], $v_2^{(t)} - v_1^{(t)} \geq (\theta_2 - \theta_1)x_1^{(t)} + \alpha\beta(w_{12}^{(t)} - w_{11}^{(t)}) \geq (\theta_2 - \theta_1)x_1^{(t)}$. Because $x_1^{(t)} \to -\infty$, it follows immediately that $(\theta_2 - \theta_1)x_1^{(t)} \to \infty$, which establishes part (iv), completing the proof. □

F. Proof of Proposition 7.1

Part (a) — ie, existence of a price system — follows from the existence result for CARA utility in Atkeson and Lucas (1992) and the definition of a non-trivial efficiency problem. We may then directly appeal to Lemma 1 in Phelan (1998). He proves that, assuming a price system exists, the optimal contract for an agent with initial promise $v^{(0)}$ is described by the following stochastic process $(c^{(\tau)})_{\tau=0}^{\infty}$ for consumption (including whatever the agent receives through his stochastic endowment):

[F.1]
$$c^{(\tau)} = -\log(-v^{(0)}) - \sum_{t=1}^{\tau-1} \log(-v(s^{(t)})) + \omega_{s^{(\tau)}} + a(s^{(\tau)})$$

where $a : S \to \mathbb{R}$ maps endowment shocks into units of consumption and, recall, $\omega_s \in \mathbb{R}$ denotes the endowment of an agent in state $s \in S$. Hence, consumption follows a random walk with drift. The first term is a constant and the latter two terms are iid innovations. It is the sum that determines the drift, which must be positive because $G > 1$ implies that $\mathbb{E}[c^{(\tau+1)}] - \mathbb{E}[c^{(\tau)}] = (G-1)\mathbb{E}[c^{(\tau)}] > 0$. Now, a random walk with positive drift converges to plus infinity with probability one, so $c^{(\tau)} \to +\infty$ a.s. By continuity of the utility function $u(\cdot)$, we also have $u(c_\tau) \to \lim_{c\to\infty} u(c) = 0$ a.s. and therefore also $v^{(\tau)} \to 0$ a.s. This gives us parts (b) and (c) of the proposition.

For part (d), consider the dynamics of $\chi^{(\tau)} := c^{(\tau)} \cdot G^{-\tau}$. The net growth rate of $\chi^{(\tau)}$ is given by

[F.2]
$$\frac{\mathbb{E}[\chi^{(\tau+1)} - \chi^{(\tau)}]}{\mathbb{E}[\chi^{(\tau)}]} = \frac{\mathbb{E}[c^{(\tau+1)}/G] - \mathbb{E}[c^{(\tau)}]}{\mathbb{E}[c^{(\tau)}]} = \frac{\mathbb{E}[c^{(\tau)}] - \mathbb{E}[c^{(\tau)}]}{\mathbb{E}[c^{(\tau)}]} = 0$$

so that the $\chi^{(\tau)}$ process has zero drift, as it must from the definition. Hence, $\chi^{(\tau)}$ is a driftless random walk, so its variance increases without bound. That is, $\mathbb{V}(\chi^{(t)}) \to +\infty$ a.s.

G. A Useful Suboptimal Contract

In this appendix, we construct a suboptimal, [**TVC**]-implementable contract called ζ. It is used to complete the proof of Lemma B.5, in many places throughout Appendices D and E, and ensures that Assumption 1 is not vacuous.

G.1. Suboptimal Contracts

It is clear that there are suboptimal contracts. We shall now show that there exists a suboptimal contract $\zeta = (\zeta^i, \zeta^c)$ that induces a value function Q^ζ (for this suboptimal contract) that is finite everywhere in V. This is non-trivial because instantaneous returns — the cost of providing the agent x utiles — are unbounded both above and below. It goes without saying that Q^{ξ} is pointwise greater than the principal's value function P.

G.2. Some Derived Parameters and Relations Between Them

Let $t_0 := \theta_2/\theta_1$. It is easy to see that $t_0 \in (0, 1)$. Define $\eta_1 := f_{21} - t_0 f_{11}$ and $\eta_2 := f_{22} - t_0 f_{12}$. It is easy to see that $\eta_2 > 0$. However, we do not have sufficient information to sign η_1. Let $\lambda_* := \alpha f_{21}/(1 - \alpha f_{22})$ and $\lambda^* := (\alpha\eta_1 + t_0)/(1 - \alpha\eta_2)$.

We shall now record some facts about these newly defined parameters.

Lemma G.1. We have $\eta_1 + t_0\eta_2 > 0$. In particular, the quadratic $\psi(t) = -f_{12}t^2 + (f_{22} - f_{11})t + f_{21} > 0$ for all $t \in (0, 1)$, and $\psi(t_0) = \eta_1 + t_0\eta_2$.

Proof. Of course, if $\eta_1 \geq 0$, then the inequality always holds. We shall show that it holds even if $\eta_1 < 0$.

Notice that

$$\begin{aligned} \eta_1 + t_0\eta_2 &= f_{21} - t_0 f_{11} + t_0(f_{22} - t_0 f_{12}) \\ &= -f_{12}t_0^2 + (f_{22} - f_{11})t_0 + f_{21} \\ &=: \psi(t_0) \end{aligned}$$

where $\psi(x)$ is the quadratic polynomial defined above. Notice that $\psi(1) = 0$, while $\psi(0) = f_{21} > 0$. Also observe that $\psi'(t) = -2f_{12}t + (f_{22} - f_{11})$, which implies that $\psi'(1) = -2f_{12} + f_{22} - f_{11} = \beta - 1 < 0$.

Because ψ is a quadratic, it can have at most two real zeros. But 1 is a zero of ψ, which implies that the other zero is real and (strictly) negative. Therefore, for all $t \in [0, 1)$, $\psi(t) > 0$, which completes the proof. □

Lemma G.2. The following inequalities hold:

(a) $0 < \lambda_*, \lambda^* < 1$.
(b) $t_0 < \lambda^*$.
(c) $\lambda_* < \lambda^*$.

Proof. Part (a) follows immediately from basic assumptions in the model, so we shall only establish parts (b) and (c).

To see (b), notice that $t_0 \leq \lambda^*$ if, and only if, $t_0 - t_0\alpha\eta_2 \leq \alpha\eta_1 + t_0$, which holds if, and only if, $\alpha(\eta_1 + t_0\eta_2) > 0$, which is true by Lemma G.1.

To see (c), notice that $\alpha f_{21}/(1 - \alpha f_{22}) < (\alpha\eta_1 + t_0)/(1 - \alpha\eta_2)$ if, and only if, $\alpha f_{21} - \alpha^2 f_{21}\eta_2 < \alpha\eta_1 + t_0 - \alpha^2 f_{22}\eta_1 - \alpha t_0 f_{22}$. This is equivalent to requiring that

$$\begin{aligned} \xi(\alpha) &:= \alpha^2(f_{21}\eta_2 - f_{22}\eta_1) + \alpha(\eta_1 - f_{21} + f_{22}t_0) + t_0 \\ &= \alpha^2 t_0(f_{22}f_{11} - f_{12}f_{21}) + \alpha t_0(f_{22} - f_{11}) + t_0 > 0 \end{aligned}$$

It is easy to see that $f_{22}f_{11} > f_{12}f_{21}$. We also have $\alpha t_0(f_{22} - f_{11}) + t_0 = t_0\,[1 + \alpha(f_{22} - f_{11})] > 0$ because $1 + \alpha(f_{22} - f_{11}) > 0$. This implies $\xi(\alpha) > 0$ for all $\alpha \in (0, 1)$, as claimed. □

Let $\mu^* := (f_{21} + \lambda^* f_{22})/(f_{11} + \lambda^* f_{12})$. It is easy to see that $\mu^* > 0$. We claim

Lemma G.3. With μ^* and λ^* defined as above, $\mu^* > \lambda^*$.

Proof. $\mu^* = (f_{21} + \lambda^* f_{22})/(f_{11} + \lambda^* f_{12}) > \lambda^*$ if, and only if, $f_{21} + \lambda^* f_{22} > f_{11}\lambda^* + \lambda^{*2} f_{12}$. This holds if, and only if, $\psi(\lambda^*) > 0$, where ψ is the quadratic defined in Lemma G.1. But this is true because $\lambda^* \in (0, 1)$ and $\psi > 0$ on $(0, 1)$. □

Consider, again, the equations

$$\text{[PK}_1\text{]} \qquad \theta_1 x_1 + \alpha f_{11} w_{11} + \alpha f_{12} w_{12} = v_1$$

$$\text{[IC]} \qquad \theta_2 x_1 + \alpha f_{21} w_{11} + \alpha f_{22} w_{12} \le v_2$$

Multiply [PK_1] by $t_0 = \theta_2/\theta_1 < 1$ and subtract it from [IC]. This results in

$$\alpha w_{11}(f_{21} - t f_{11}) + \alpha w_{12}(f_{22} - t f_{12}) \le v_2 - t_0 v_1$$

which we rewrite as

$$\text{[IC}^\star\text{]} \qquad \alpha \eta_1 w_{11} + \alpha \eta_2 w_{12} \le v_2 - t_0 v_1$$

Notice that [PK_1] and [$\text{IC}^\star$] together imply [IC]. To see this, multiply [PK_1] by $t = \theta_2/\theta_1 < 1$ and add it to [$\text{IC}^\star$] to obtain [IC].

G.3. Suboptimal Strategy – Part 1

Let $V_1 := \{\mathbf{v} \in V : \mathbf{v} = (v_1, \lambda v_1),\ \lambda_* < \lambda \le \lambda^*\}$ be a cone in the domain. We shall now show that every point in V_1 is self-generating in the sense that there exists a feasible strategy such that given a starting point $\mathbf{v} \in V_1$, we can satisfy promise keeping and incentive compatibility while staying at $\mathbf{v}$ in all subsequent periods.

For $\mathbf{v} \in V_1$, define the contract $\zeta(\mathbf{v}, s) = (\zeta^i(\mathbf{v}, s) = x_s, \zeta^c(\mathbf{v}, s) = \mathbf{v})$.

It suffices to show the following:

Lemma G.4. Let $\mathbf{v} \in V_1$. Then, there exist $x_s < 0$ such that $\zeta(\mathbf{v}, s) = (\zeta^i(\mathbf{v}, s) = x_s, \zeta^c(\mathbf{v}, s) = \mathbf{v})$ solves [PK_2], [PK_1], and [IC].

Proof. The contract $\zeta(\mathbf{v}, s)$ satisfies [PK_2] if, and only if,

$$\alpha(f_{21} + \lambda f_{22})v_1 > \lambda v_1$$

which, in turn, holds if, and only if, $\lambda > \alpha f_{21}/(1 - \alpha f_{22}) = \lambda_*$, which is true by assumption.

Similarly, $\zeta(\mathbf{v}, s)$ satisfies [PK_1] if, and only if,

$$\alpha(f_{11} + \lambda f_{12})v_1 > v_1$$

which, in turn, holds if, and only if, $\lambda < (1 - \alpha f_{11})/\alpha f_{22}$, which is always true because $\lambda < 1 < (1 - \alpha f_{11})/\alpha f_{22}$.

Finally, $\zeta(\mathbf{v}, s)$ satisfies [**IC**] if, and only if, it satisfies [**IC***], which requires that

$$\alpha(\eta_1 + \lambda\eta_2)v_1 \leq (\lambda - t_0)v_1$$

Because $v_1 < 0$, this holds if, and only if, $\alpha(\eta_1 + \lambda\eta_2) \geq (\lambda - t_0)$ which is equivalent to $\lambda \leq (\alpha\eta_1 + t_0)/(1 - \alpha\eta_2) = \lambda^*$. This holds because $(v_1, \lambda v_2) \in V_1$. Therefore, $\zeta^c(\mathbf{v}, s) = \mathbf{v}$ is feasible and implements itself, as claimed. □

G.4. Suboptimal Strategy — Part 2

Recall μ^* defined above in section G.2, and define

$$V_2 := \left\{\mathbf{v} \in V : \mathbf{v} = (v_1, \mu v_1),\, 0 < \mu < \mu^*\right\}$$

We will now show that it is possible to implement $\mathbf{v} \in V$ such that $\mathbf{w}_i \in V$. More precisely, we have

Lemma G.5. Let $\mathbf{v} \in V_2$. Then, there exists $(x_i, \mathbf{w}_i)_{i=1,2} \ll \mathbf{0}$ that implements $\mathbf{v}$ such that $\mathbf{w}_i \in V_1$ for $i = 1, 2$.

Proof. From Lemma G.4, we can implement all $\mathbf{v} \in V_1$. Therefore, it suffices to consider $\mathbf{v} \in V_2 \setminus V_1$. We can readily choose $\mathbf{w}_2 \in V_1$ and $x_2 < 0$ such that $\theta_2 x_2 + \langle \mathbf{f}_2, \mathbf{w}_2\rangle = v_2$, ie, such that [**PK$_2$**] holds. Notice that the choice of $(x_2, \mathbf{w}_2)$ does not affect the choice of $(x_1, \mathbf{w}_1)$, because these variables don't enter [**PK$_1$**] and [**IC**].

Let $\mathbf{v} = (v_1, v_2 = \mu v_1)$. In order to prove the lemma, we need to find $w_{11} < 0$ and $\lambda \in (\lambda_*, \lambda^*]$ such that

$$\begin{aligned}
\alpha(f_{11} + \lambda f_{12})w_{11} &> v_1 \\
\alpha(\eta_1 + \lambda\eta_2)w_{11} &\leq v_2 - t_0 v_1 = (\mu - t_0)v_1
\end{aligned}$$

where the first inequality is merely [**PK$_1$**] while the second reflects [**IC***]. This system has a solution if, and only if,

$$\frac{1}{\alpha(f_{11} + \lambda f_{12})}v_1 < \frac{\mu - t_0}{\alpha(\eta_1 + \lambda\eta_2)}v_1$$

which has a solution if, and only if,

$$\frac{1}{\alpha(f_{11} + \lambda f_{12})} > \frac{\mu - t_0}{\alpha(\eta_1 + \lambda\eta_2)}$$

This has a solution if, and only if,

$$\begin{aligned}
\mu &< \frac{\eta_1 + \lambda\eta_2 + t_0(f_{11} + \lambda f_{12})}{(f_{11} + \lambda f_{12})} \\
&= \frac{f_{21} - t_0 f_{11} + \lambda(f_{22} - t_0 f_{12}) + t_0 f_0 - t_0\lambda f_{12}}{(f_{11} + \lambda f_{12})} \\
&= \frac{f_{21} + \lambda f_{22}}{(f_{11} + \lambda f_{12})} \\
&\leq \mu^* \quad \text{whenever } \lambda \leq \lambda^*
\end{aligned}$$

which is true by assumption. Therefore, we can choose any $\lambda \in (\lambda_*, \lambda^*]$ for which there exists $w_{11} < 0$ so that [**PK$_1$**] and [**IC***] (or equivalently [**IC**]) hold.

Notice that given $\mathbf{w}_{11} \in V_1$, we can always find x_1 such that [**PK$_1$**] holds with equality. This completes the proof. □

Let $(x_i, \mathbf{w}_i)_{i=1,2}$ be the solution described in Lemma G.5. Then, setting $\zeta(\mathbf{v}, i) = (\zeta^i(\mathbf{v}, i) = x_i, \zeta^c(\mathbf{v}, i) = \mathbf{w}_i)$ gives us a suboptimal contract defined on V_2 (recall that $V_1 \subset V_2$ because $\lambda^* < \mu^*$, as proved in Lemma G.3).

G.5. Suboptimal Strategy — Part 3

Let $V_3 := \{\mathbf{v} \in V : \mathbf{v} = (v_1, \mu v_1),\ \mu^* \leq \mu < 1\}$, so that $V = V_2 \cup V_3$. we shall now show that it is possible to construct a contract so that for any $\mathbf{v} \in V_3$, the continuation utilities reach V_1 in finitely many steps almost surely. In particular, this time can be taken to be independent of the sequence of shocks under consideration.

Let $W \subset V$ be a fixed subset and fix an initial $\mathbf{v} \in V$ and $s^{(0)}$. A contract is W-*amenable* at $(\mathbf{v}, s^{(0)})$ if there is a finite natural number N (that depends on W and $(\mathbf{v}, s^{(0)})$) such that for any sequence of shocks $s^{(1)}, \ldots, s^{(N)} \in S^N$, the continuation utility reaches W in at most N steps.

Lemma G.6. Let $\mathbf{v} \in V_3$ and fix $s^{(0)}$. Then, there exists a contract that is V_1-amenable at $(\mathbf{v}, s^{(0)})$.

Proof. Let $\mathbf{v} = (v_1, v_2 = \mu_0 v_1) \in V_3$ where $\mu_0 \in [\mu^*, 1)$. We will show that linear inequalities derived from the constraints possess suitable solutions; any collection of such solutions will serve as the desired V_1-amenable contract.

First, consider the case where $s^{(1)} = 2$. Let $\mathbf{w}_2 = (w_{21}, \lambda w_{21})$, so that we can re-write [**PK$_2$**] as

$$\alpha(f_{21} + \lambda f_{22})w_{21} > \mu v_1$$

It is immediate that we can choose $\lambda \in (\lambda_*, \lambda^*]$ and $w_{21} < 0$ that satisfies this inequality, and hence [**PK$_2$**]. Thus, if the first shock is $s^{(1)} = 2$, then the continuation utility $\mathbf{w}_2 \in V_1$.

Next, consider the case where $s^{(1)} = 1$. Let $\mathbf{w}_1 = (w_{11}, \mu_1 w_{11})$, so that [**PK$_1$**] and [**IC***] can (respectively) be re-written as

$$\alpha(f_{11} + \mu_1 f_{12})w_{11} > v_1$$
$$\alpha(\eta_1 + \mu_1 \eta_2)w_{11} \leq (\mu_0 - t_0)v_1$$

We can find $w_{11} < 0$ that solves the above inequalities if, and only if,

$$\frac{1}{f_{11} + \mu_1 f_{12}} > \frac{\mu_0 - t_0}{\eta_1 + \mu_1 \eta_2}$$

which, in turn, holds if, and only if,

$$\mu_1(\eta_2 - f_{12}(\mu_0 - t_0)) > f_{11}(\mu_0 - t_0) - \eta_1$$

It is easy to see that $\eta_2 - f_{12}(\mu_0 - t_0) = f_{22} - f_{12}\mu_0 > 0$. Moreover, $f_{11}(\mu_0 - t_0) - \eta_1 = f_{11}\mu_0 - f_{21}$.

Notice that $f_{11}\mu_0 - f_{21}$ can be either positive or negative. If it is negative (or zero), then we can always find $\mu_1 \in (\lambda_*, \lambda^*]$ such that $\mu_1(\eta_2 - f_{12}(\mu_0 - t_0)) > 0$, which implies that continuation utility $\mathbf{w}_1 \in V_1$ in the following period.

However, if $f_{11}\mu_0 - f_{21} > 0$, we can set $\mu_1 := (f_{11}\mu_0 - f_{21})/(f_{22} - f_{12}\mu_0)$. Then,

$$\begin{aligned}\mu_0 - \mu_1 &= \mu_0 - (f_{11}\mu_0 - f_{21})/(f_{22} - f_{12}\mu_0) \\ &= \frac{1}{f_{22} - f_{12}\mu_0}\left[-f_{12}\mu_0^2 + \mu_0(f_{22} - f_{11}) + f_{21}\right] \\ &= \frac{1}{f_{22} - f_{12}\mu_0}\psi(\mu_0) \\ &> 0\end{aligned}$$

because $\psi > 0$ on $(0, 1)$. Thus, $\mu_0 > \mu_1$. Proceeding iteratively, we can find a decreasing sequence where $\mu_n := (f_{11}\mu_{n-1} - f_{21})/(f_{22} - f_{12}\mu_{n-1})$. Because ψ is a quadratic, the sequence $\psi(\mu_n)$ lies in $[\psi(0), \psi(\mu_0)]$. Therefore,

$$\begin{aligned}\mu_{n-1} - \mu_n &= \frac{1}{f_{22} - f_{12}\mu_{n-1}}\psi(\mu_{n-1}) \\ &\geq \frac{1}{f_{22}}\min\left[\psi(0), \psi(\mu_0)\right] > 0\end{aligned}$$

so the decreasing sequence has a difference uniformly bounded away from zero. Therefore, there exists a finite N such that $\mu_N < \mu^*$. Then, if continuation utility is in V_2, we can use the contract defined in Lemma G.5 that then implements a transition to V_1 in one more step.

Any contract constructed from the solutions to these linear inequalities is one we desired, completing the proof. □

G.6. Suboptimal Strategy — Summary

We shall now show that there exists a suboptimal strategy that has finite cost for the principal. This implies that the optimal strategy, if it exists, also has finite costs. More generally, the solution to the sequence problem (using lim inf or lim sup) also has a finite solution for every $\mathbf{v} \in V$.

Proposition G.7. There exists a sub-optimal contract ζ that, starting at $(\mathbf{v}, s)$, costs the principal $Q^{\zeta}(\mathbf{v}, s) \in \mathbb{R}$ for all $(\mathbf{v}, s) \in V \times S$.

Proof. The proof relies on Lemmas G.4, G.5, and G.6. Starting at $(\mathbf{v}, s)$, the contract defined in Lemmas G.5 and G.6 guides continuation promised utility to V_1 in finitely many steps, where the finite bound is independent of the sequence of shocks. Then, upon reaching V_1, continuation utility stays at the same point regardless of the shock – this is Lemma G.4. Since there are only finitely many different utility levels for any starting point $(\mathbf{v}, s)$, the cost to the principal of this contract is finite, which proves the claim. □

H. The Optimal Contract — Short-Run Properties

In this appendix, we characterize the short-run dynamics of the optimal contract. This is done in two steps. First, in Appendix H.1 presents the optimality conditions and characterizes the dynamics of the dual variables for the principal's problem, highlighting allocative distortions that arise due to the persistence of the agent's private information. This characterization does *not* rely on the assumption of CARA utility. Second, in Appendix H.3, we use these results to prove Proposition 5.1 from Section 5.

H.1. Short-Run Dynamics — Dual Variables

Theorem 2 delivers a smooth, finite-dimensional, convex minimization problem, so we may solve for the optimal contract by taking (necessary and sufficient) first-order conditions. Letting $\mu \geq 0$ be the multiplier on [**IC**] and λ_1 and $\lambda_2 \in \mathbb{R}$ be the multipliers on [$\mathbf{PK_1}$] and [$\mathbf{PK_2}$] respectively, we have the Lagrangian

$$\mathscr{L}(\mathbf{v}, s) = \sum_{i=1,2} f_{si} \left[C(x_i) + \alpha P(\mathbf{w}_i, i) + \frac{\lambda_i}{f_{si}} \left(v_i - \theta_i x_i - \alpha \mathbb{E}^{\mathbf{f}_i} [\mathbf{w}_i] \right) \right]$$
$$- \mu \left(\theta_2 x_2 + \alpha \mathbb{E}^{\mathbf{f}_2} [\mathbf{w}_2] - \theta_2 x_1 - \alpha \mathbb{E}^{\mathbf{f}_2} [\mathbf{w}_1] \right)$$

The optimality equations consist of the envelope conditions

$$[\mathbf{Env}_i] \qquad P_i(\mathbf{v}, s) = \lambda_i$$

the first-order conditions for instantaneous utilities

$$[\mathbf{FOC}x_i] \qquad f_{si} C'(x_i) = \theta_i \lambda_i + (-1)^i \theta_2 \mu$$

and the first-order conditions for contingent continuation utilities

$$[\mathbf{FOC}w_{1i}] \qquad f_{s1} P_i(\mathbf{w}_1, 1) = f_{1i} \lambda_1 - f_{2i} \mu$$
$$[\mathbf{FOC}w_{2i}] \qquad f_{s2} P_i(\mathbf{w}_2, 2) = f_{2i} \lambda_2 + f_{2i} \mu$$

for $i = 1, 2$.

In this section, we change variables by looking at the map $(\mathbf{v}, s) \mapsto \mathrm{D}P(\mathbf{v}, s) = (P_1(\mathbf{v}, s), P_2(\mathbf{v}, s))$, and are immediately able to provide a number of qualitative insights.[53] Thus, since the envelope condition [$\mathbf{Env}_i$] implies that $P_i(\mathbf{v}, s) = \lambda_i$, we study the dynamics of the *dual variables* in the principal's problem.

We will use the following notation. From any starting state $(\mathbf{v}, s)$, after a θ_2 shock the optimal contract moves to state $(\mathbf{w}_1, 1)$. From this state, the contract transitions to $(\mathbf{w}_i^{(2)}(\theta_2), i)$ if the next shock is $i \in \{1, 2\}$. Our first characterization result is as follows.

Proposition H.1. Under the optimal mechanism:

(53) Theorem 2 and Proposition 3.2 guarantee that the value function is strictly convex and smooth, so the map $(\mathbf{v}, s) \mapsto \mathrm{D}P(\mathbf{v}, s)$ is a bijection, and this change of variables is without loss of generality.

(a) The support of the process $\mathsf{D}P(\mathbf{v}^{(t)}, s^{(t)})$ is a subset of the half-space of (P_1, P_2)-space for which $\mathsf{D}_{(1,1)}P > 0$.

(b) When $s^{(t)} = 1$ and $\beta > 0$, $\mathsf{D}P(\mathbf{v}^{(t)}, s^{(t)})$ lies in a convex cone *strictly below* the ray

$$E_1 := \left\{(P_1, P_2) \in \mathbb{R}^2 : P_1 + P_2 > 0, \quad \frac{P_1}{f_{11}} = \frac{P_2}{f_{12}}\right\}$$

When $s^{(t)} = 1$ and $\beta = 0$, $\mathsf{D}P(\mathbf{v}^{(t)}, s^{(t)})$ lies *exactly on* this ray.

(c) When $s^{(t)} = 2$, $\mathsf{D}P(\mathbf{v}^{(t)}, s^{(t)})$ lies on the ray

$$E_2 := \left\{(P_1, P_2) \in \mathbb{R}^2 : P_1 + P_2 > 0, \quad \frac{P_2}{f_{22}} = \frac{P_1}{f_{21}}\right\}$$

regardless of the value of $\beta \geq 0$.

(d) When $\beta = 0$, the rays E_1 and E_2 coincide.

(e) For any degree of persistence $\beta \in [0, 1)$, we have *martingale splitting* after high endowment shocks:

$$\mathsf{D}_{(1,1)}P(\mathbf{w}_1^{(2)}(\theta_2), 1) < \mathsf{D}_{(1,1)}P(\mathbf{w}_2, 2) < \mathsf{D}_{(1,1)}P(\mathbf{w}_2^{(2)}(\theta_2), 2)$$

If $\beta = 0$, we also have martingale splitting after low endowment shocks.

H.2. Proof of Proposition H.1

We prove the various parts of the proposition in turn. Part (a) follows immediately from part (d) of Proposition 3.2. For part (b), recall from Section H.1 (or Appendix D.4) the FOCs

$$[\mathbf{FOC}w_{11}] \qquad f_{s1}P_1(\mathbf{w}_1, 1) = f_{11}\lambda_1 - f_{21}\mu$$

$$[\mathbf{FOC}w_{12}] \qquad f_{s1}P_2(\mathbf{w}_1, 1) = f_{12}\lambda_1 - f_{22}\mu$$

Rearranging [**FOC**w_{11}] and [**FOC**w_{12}] delivers

$$\frac{P_1(\mathbf{w}_1, 1)}{f_{11}} = \frac{1}{f_{s1}} \cdot \left[-\frac{f_{21}}{f_{11}}\mu + \lambda_1\right]$$
$$\frac{P_2(\mathbf{w}_1, 1)}{f_{12}} = \frac{1}{f_{s1}} \cdot \left[-\frac{f_{22}}{f_{12}}\mu + \lambda_1\right]$$

Taking differences of these two expressions yields

$$[\mathbf{H.1}] \qquad \frac{P_1(\mathbf{w}_1, 1)}{f_{11}} - \frac{P_2(\mathbf{w}_1, 1)}{f_{12}} = \frac{\mu}{f_{s1}} \cdot \left[\frac{f_{22}}{f_{12}} - \frac{f_{21}}{f_{11}}\right] \geq 0$$

Now, under Assumption 1, we have $\mu > 0$ by Lemma D.19. Moreover, the bracketed term is strictly positive if $\beta > 0$ and equal to zero if $\beta = 0$. This shows that $\mathsf{D}P(\mathbf{w}_1, 1)$ lies below the ray E_1. By a simple adaptation of Lemma E.1, it must lie in a convex cone. This completes the proof of part (b).

For part (c), recall from Section H.1 the FOCs

$$[\mathbf{FOC}w_{21}] \qquad f_{s2}P_1(\mathbf{w}_2, 2) = f_{21}\lambda_2 + f_{21}\mu$$

$$[\mathbf{FOC}w_{22}] \qquad f_{s2}P_2(\mathbf{w}_2, 2) = f_{22}\lambda_2 + f_{22}\mu$$

Dividing [**FOC**w_{22}] by [**FOC**w_{21}] (recall from Theorem 2 that $P_1(\mathbf{w}_2, 2) > 0$) gives us the set E_2.

Part (d) follows immediately from the fact that, when $\beta = 0$, we have $P(\mathbf{v}, 1) = P(\mathbf{v}, 2)$ for all $\mathbf{v} \in V$.

Finally, part (e) follows from the next lemma and the fact that $\mu > 0$ (under Assumption 1; see Lemma D.20):

Lemma H.2. On any path induced by the recursively optimal contract, the differential martingale $\mathsf{D}_{(1,1)}P$ evolves according to

[**H.2**] $$\mathsf{D}_{(1,1)}P(\hat{\mathbf{w}}_1, 1) = \mathsf{D}_{(1,1)}P(\mathbf{w}_1, 1) + \frac{\beta\mu(\mathbf{v}, s) - f_{s1}\mu(\mathbf{w}_1, 1)}{f_{s1} \cdot f_{11}}$$

[**H.3**] $$\mathsf{D}_{(1,1)}P(\hat{\mathbf{w}}_2, 2) = \mathsf{D}_{(1,1)}P(\mathbf{w}_1, 1) - \frac{\beta\mu(\mathbf{v}, s) - f_{s1}\mu(\mathbf{w}_1, 1)}{f_{s1} \cdot f_{12}}$$

[**H.4**] $$\mathsf{D}_{(1,1)}P(\tilde{\mathbf{w}}_1, 1) = \mathsf{D}_{(1,1)}P(\mathbf{w}_2, 2) - \frac{\mu(\mathbf{w}_2, 2)}{f_{21}}$$

[**H.5**] $$\mathsf{D}_{(1,1)}P(\tilde{\mathbf{w}}_2, 2) = \mathsf{D}_{(1,1)}P(\mathbf{w}_2, 2) + \frac{\mu(\mathbf{w}_2, 2)}{f_{22}}$$

where, for any $(\mathbf{v}, s) \in V \times S$, we define $\mathbf{w}_i := \xi^c((\mathbf{v}, s), i)$, $\hat{\mathbf{w}}_i := \xi^c((\mathbf{w}_1, 1), i)$ and $\tilde{\mathbf{w}}_i := \xi^c((\mathbf{w}_2, 2), i)$.

Proof. We refer to $(\mathbf{v}, s)$ as date $t-1$ and $(\mathbf{w}_i, i)$ as date t.

Consecutive θ_1 shocks: Take [**FOC**w_{11}] at date $t-1$ and apply [**Env1**] from dates $t-1$ and t to get

[**H.6**] $$f_{s1}\left[\lambda_1(\mathbf{w}_1, 1) - \mu(\mathbf{w}_1, 1)\right] = f_{11}\left[\lambda_1(\mathbf{v}, s) - \mu(\mathbf{v}, s)\right] + \beta\mu(\mathbf{v}, s) - f_{s1}\mu(\mathbf{w}_1, 1)$$

Add together [**FOC**w_{11}] and [**FOC**w_{12}] at dates $t-1$ and t, respectively, to get

[**H.7**] $$f_{s1}\mathsf{D}_{(1,1)}P(\mathbf{w}_1, 1) = -\mu(\mathbf{v}, s) + \lambda_1(\mathbf{v}, s)$$

and

[**H.8**] $$f_{11}\mathsf{D}_{(1,1)}P(\hat{\mathbf{w}}_1, 1) = -\mu(\mathbf{w}_1, 1) + \lambda_1(\mathbf{w}_1, 1)$$

Combining [**H.6**], [**H.7**], and [**H.8**] yields [**H.2**].

Consecutive θ_2 shocks: Take [**FOC**w_{22}] at date $t-1$ and apply [**Env2**] from dates $t-1$ and t to get

[**H.9**] $$f_{s2}\left[\lambda_2(\mathbf{w}_2, 2) + \mu(\mathbf{w}_2, 2)\right] = f_{22}\left[\lambda_2(\mathbf{v}, s) + \mu(\mathbf{v}, s)\right] + f_{s2}\mu(\mathbf{w}_2, 2)$$

Add together [**FOC**w_{21}] and [**FOC**w_{22}] at dates $t-1$ and t, respectively, to get

[**H.10**] $$f_{s2}\mathsf{D}_{(1,1)}P(\mathbf{w}_2, 2) = \mu(\mathbf{v}, s) + \lambda_2(\mathbf{v}, s)$$

and

[**H.11**] $$f_{22}\mathsf{D}_{(1,1)}P(\tilde{\mathbf{w}}_2, 2) = \mu(\mathbf{w}_2, 2) + \lambda_2(\mathbf{w}_2, 2)$$

Combining [**H.9**], [**H.10**], and [**H.11**] yields [**H.5**].

θ_1 shock followed by θ_2 shock: Combine the martingale property at state $(\mathbf{w}_1, 1)$ with [**H.2**] to obtain [**H.3**].

θ_2 shock followed by θ_1 shock: Combine the martingale property at state $(\mathbf{w}_2, 2)$ with [**H.5**] to obtain [**H.4**]. ☐

H.3. Proof of Proposition 5.1

We prove the various parts of Proposition 5.1 in turn:

Proof of part (a). That the rays E_1 and E_2 correspond to rays $\tilde{E}_1$ and $\tilde{E}_2$ in V follows immediately from the definition of the sets E_1 and E_2 and the homogeneity properties of the partial derivatives P_i (see Lemma D.15 or part (e) of Theorem 2). To see that $\tilde{E}_1$ lies strictly above $\tilde{E}_2$, we will use the following lemma. Recall the definition of the interim multipliers (η_1, η_2, σ) from Appendix D.4.

Lemma H.3. We have the following:

(a) $\mathbf{v} \in \tilde{E}_1$ if and only if

[**H.12**] $$\eta_2(v_2) = \eta_1(\mathbf{v}) + \frac{\sigma(\mathbf{v})}{f_{12}}$$

(b) Fix $\mathbf{v} \in V$. At the optimum, we have

[**H.13**] $$\eta_2(w_{12}) = \eta_1(\mathbf{w}_1) + \frac{\sigma(\mathbf{w}_1)}{f_{12}} - \frac{\beta}{f_{11}}\frac{\sigma(\mathbf{v})}{f_{12}}$$

(c) $\mathbf{v} \in \tilde{E}_2$ if and only if

[**H.14**] $$\eta_2(v_2) = \eta_1(\mathbf{v}) + \frac{\sigma(\mathbf{v})}{f_{22}}$$

Proof. Parts (a) and (c) follow immediately from the definition of the $\tilde{E}_i$ and [**D.9**] and [**D.10**] in Lemma D.20. Part (b) follows from [**D.11**] in Lemma D.20 and [**H.1**] (which itself follows from the FOCs displayed in Section H.1). ☐

Notice that if $\beta = 0$, then $\tilde{E}_1 = \tilde{E}_2$. Suppose $\beta > 0$. Then, since $f_{22} > f_{12}$ and $\sigma(\mathbf{v}) > 0$ (see Lemma D.19), it is easy to see from Lemma H.3 that $\mathbf{v} \in \tilde{E}_2$ implies that $\eta_2(v_2) < \eta_1(\mathbf{v}) + \frac{\sigma(\mathbf{v})}{f_{12}}$. This is equivalent to the condition that $P_1(\mathbf{v}, 1)/f_{11} > P_2(\mathbf{v}, 1)/f_{12}$. This implies that $\tilde{E}_1$ lies above $\tilde{E}_2$ because convexity and continuous differentiability of P (see Theorem 2) imply that the lower contour sets of $P(\cdot, 1)$ (ie, sets of the form $L_K := \{\mathbf{v} \in V : P(\mathbf{v}, 1) \leq K\}$ for $K \in \mathbb{R}$) are convex, and tangent to the vector $\mathbf{f}_1^{\perp} := (1/f_{11}, -1/f_{12})$ exactly at points $\mathbf{v} \in \tilde{E}_1$.

Proof of part (b). This follows immediately from part (b) of Lemma H.3 and the above arguments for part (a).

Proof of part (c). Part (e) of Proposition 5.1 tells us that the differential martingale must strictly increase after consecutive θ_2 shocks. From the definition of the ray E_2 (part (c) of Proposition H.1) and the fact that the partial derivatives of P are homogenous of degree -1 (part (e) of Theorem 2), it follows that $\mathbf{w}_2$ increases after consecutive θ_2 shocks. To see that this implies that x_2 increases as well, note that [**Env**$_2$-Q^2] and [**FOC**x_2-Q^2] imply that $Q_2^2(v_2) = C'(x_2)/\theta_2$ at the optimum, and recall from Lemma D.18 that Q^2 is convex. The under-insurance property follows from the following lemma:

Lemma H.4. Suppose $\mathbf{v} \in E_2$. Then

$$C'(x_2)/\theta_2 > C'(x_1)/\theta_1$$

Proof. From [**FOC**x_1-Q^1] and [**FOC**x_2-Q^2] we have the above inequality if and only if

$$\eta_2(\mathbf{v}) > \eta_1(\mathbf{v}) + \frac{\theta_1 - \theta_2}{\theta_1}\sigma(\mathbf{v})$$

But if $\mathbf{v} \in E_2$, [**H.14**] lets us write

$$\eta_2(\mathbf{v}) = \eta_1(\mathbf{v}) + \frac{\sigma(\mathbf{v})}{f_{22}}$$

Since $1/f_{22} > 1$ and $(\theta_1 - \theta_2)/\theta_1 < 1$, and $\sigma > 0$, the claim follows. □

This completes the proof of part (c).

Proof of part (d). This follows immediately from part (e) of Proposition H.1, part (a) of Proposition 5.1, and the homogeneity of the partial derivatives of P (part (e) of Theorem 2). This proves the proposition.

I. General Model

In this appendix, we develop the extension to $n \geq 2$ types and to more general utility functions, as discussed in Section 6. This development culminates in Theorem 8 in Appendix I.5, which extends our immiseration result (Theorem 4 in the main text) to this more general setting.

For the sake of brevity, we merely sketch arguments that closely parallel those from the main binary-type model. Moreover, establishing certain properties of the value function requires arguments that depend on the special case under consideration (such as whether the consumption domain is bounded or closed, or whether the utility function satisfies growth conditions). In these instances, we state sufficient conditions that can be verified on a case-by-case basis.

I.1. Model

The agent has a strictly increasing, strictly concave, and continuously differentiable utility function $u : \mathcal{C} \to \mathbb{R}$, where $\mathcal{C} \subseteq \mathbb{R}$ is the domain of consumption which we assume is convex. Let $\mathcal{U} := u[\mathcal{C}]$, the image of $\mathcal{C}$ under u. Define $\underline{c} := \inf \mathcal{C}$ and $\overline{c} := \sup \mathcal{C}$, and similarly $\underline{u} := \inf \mathcal{U}$ and $\overline{u} := \sup \mathcal{U}$. We make the following regularity assumptions.

Assumption 2. The agent's utility function $u(\cdot)$ satisfies:

(a) **Inada conditions:** $\lim_{c \to \overline{c}} u'(c) = 0$ and $\lim_{c \to \underline{c}} u'(c) = \infty$.

(b) **DARA:** The function $\phi : \mathscr{C} \to \mathbb{R}$ defined by $\phi(c) := -\log(u'(c))$ is (weakly) concave.

Part (a) of Assumption 2 is essential for establishing the immiseration result, as pointed out by Phelan (1998). Part (b), which is also used in Thomas and Worrall (1990), states that the utility function has (weakly) decreasing absolute risk aversion (DARA). (To see this, notice that if $u(\cdot)$ is twice differentiable, then $\phi'(c) = -u''(c)/u'(c)$, the coefficient of absolute risk aversion. Concavity guarantees that this is weakly decreasing.) It is needed to ensure that the principal's domain and constraint correspondence are convex, allowing us to restrict attention to optimal contracts that do not involve randomization. Note that CARA and CRRA utility functions fall into this class.

The utility function has inverse $C : \mathscr{U} \to \mathscr{C}$, which is strictly increasing, strictly convex, and continuously differentiable. Assumption 2 implies that $\lim_{x\to \underline{u}} C'(x) = 0$ and $\lim_{x\to\overline{u}} C'(x) = \infty$.

The agent's random endowment in period t is $\omega^{(t)} \in \mathscr{W} = \{\omega_1, \ldots, \omega_n\} \subset \mathscr{C}$ where $n \geq 2$ and $\omega_n > \cdots > \omega_1$. Equivalently, the agent's *type* in period t is $s^{(t)} \in S = \{1, \ldots, n\}$. The type process $(s^{(t)})_{t\geq 0}$ is Markovian with transition probabilities

$$\mathbb{P}\left(s^{(t+1)} = j \middle| s^{(t)} = i\right) = f_{ij}$$

For simplicity, we assume that the agent's type process is *fully connected*, so that $f_{ij} > 0$ for all $i, j \in S$. We emphasize that we do *not* need to make any assumptions about the serial correlation structure beyond connectedness; our focus on positive serial correlation in the text was motivated by empirical relevance.

The recursive constraints may be written as

$$[\mathbf{PK}_i] \qquad v_i = x_i + \alpha \mathbb{E}^{\mathbf{f}_i}[\mathbf{w}_i]$$

$$[\mathbf{IC}_{ij}] \qquad v_i \geq \psi_{ij}(x_j) + \alpha \mathbb{E}^{\mathbf{f}_i}\left[\mathbf{w}_j\right]$$

where we require that [$\mathbf{PK}_i$] hold for all $i \in S$ and that [$\mathbf{IC}_{ij}$] hold for all $i, j \in S$ such that $i > j$. Thus, we assume that there is *no hidden borrowing*, so that the agent can only under-report his endowment.

Remark 6. Notice that the notation here is slightly different from that in the main body. Previously, the variables x_i denoted the agent's flow utility, net of that from his endowment, so that his total flow utility was $\theta_i x_i$. Here, x_i denotes the agent's flow utility *inclusive of* that from his endowment. In particular, if the principal gives consumption c_i an agent of type i, then the agent's total flow utility is $x_i = u(c_i + \omega_i)$.

The functions $\psi_{ij} : \mathscr{U} \to \mathscr{U}$ are defined for all $i, j \in S$ such that $i > j$, and represent the flow utility that a type i agent receives from consuming the flow allocation meant for type j. Formally,

$$\psi_{ij}(x) := u(\omega_i + C(x) - \omega_j)$$

We have the following important property.

Lemma I.1. Let $i, j \in S$ with $i > j$. Under Assumption 2, the functions ψ_{ij} are convex.

Proof. Let $x \in \text{int}(\mathscr{C})$ and define $\Delta_{ij} := \omega_i - \omega_j > 0$. Differentiating, and using the Inverse Function Theorem, we get

$$\psi'_{ij}(x) = \frac{u'(\Delta + C(x))}{u'(C(x))} > 0$$

Thus, using the function $\phi(\cdot)$ defined in Assumption 2, we have

$$\log\left(\psi'_{ij}(x)\right) = \phi\left(C(x)\right) - \phi\left(\Delta + C(x)\right)$$

(Notice that $\phi(\cdot)$, as defined, is increasing.) Because $C(\cdot)$ is strictly increasing, Assumption 2 guarantees that the RHS of the above equation is an increasing function of x, implying that $\psi'_{ij}(\cdot)$ is increasing as well. Thus, $\psi_{ij}(\cdot)$ is convex. □

I.2. Recursive Domain and Value Function

A largest recursive domain, $\varnothing \neq V \subseteq \mathbb{R}^n$, is easily seen to exist by using arguments similar to those in Appendix B. Namely, starting from an appropriately large subset $D \subseteq \mathbb{R}^n$, define an APS-style operator $A : 2^D \to 2^D$ from the [$\mathbf{PK}_i$] and [$\mathbf{IC}_{ij}$] constraints. This operator will be monotone on the complete lattice $(2^D, \subseteq)$. Thus, by Tarski's Fixed Point Theorem, the collection of fixed points of A forms a (nonempty) complete sublattice. Hence, a largest fixed point exists. This largest fixed point, namely V, must be nonempty, as it is always possible to implement the contingent utility vector that corresponds to the agent's autarky endowment stream.

As before, define the principal's recursive constraint correspondence by

$$\Gamma(\mathbf{v}) = \Big\{(x_s, \mathbf{w}_s) \in (\mathcal{U} \times V)^S : (x_s, \mathbf{w}_s)_{s \in S} \text{ satisfies } [\mathbf{PK}_i]\ \forall\, i \in S \\ \text{and } [\mathbf{IC}_{ij}]\ \forall\, i, j \in S \text{ with } i > j\Big\}$$

As noted above, Assumption 2 ensures that the principal's domain and constraint sets are convex, so that it is without loss to restrict attention to contracts that do not involve extraneous randomization.

Proposition I.2. Under Assumption 2, the domain $V \subseteq \mathbb{R}^n$ is convex, and the constraint correspondence $\Gamma : V \to (\mathcal{U} \times V)^S$ is convex-valued.

Proof. Let $\mathbf{v}, \mathbf{v}' \in V$ and $\alpha \in (0,1)$. By definition, there exist $(x_s, \mathbf{w}_s)_{s\in S} \in \Gamma(\mathbf{v})$ and $(x'_s, \mathbf{w}'_s)_{s\in S} \in \Gamma(\mathbf{v}')$. Let $\mathbf{v}^\alpha := \alpha \mathbf{v} + (1-\alpha)\mathbf{v}'$ and $(x^\alpha_s, \mathbf{w}^\alpha_s)_{s\in S} := \alpha(x_s, \mathbf{w}_s)_{s\in S} + (1-\alpha)(x_s, \mathbf{w}'_s)_{s\in S}$, where the convex combination is defined component-wise. Clearly this new contract satisfies each of the [$\mathbf{PK}_i$] at $\mathbf{v}''$. As each of the $\psi_{ij}(\cdot)$ are convex under Assumption 2, each of the [$\mathbf{IC}_{ij}$] hold as well. Thus, since V is defined to be the largest domain, it is easy to see that it must contain $\mathbf{v}^\alpha$. Hence, V must be convex. Given convexity of V, the above argument also establishes that the correspondence $\Gamma(\cdot)$ is convex-valued. □

In order to characterize the principal's value function and establish existence of an optimal contract (ie, to state the analogues of Theorem 2 and Proposition 3.2), we require a few boundedness properties. Proofs of these properties from primitive conditions of the model — namely, the utility function $u(\cdot)$ and consumption domain $\mathcal{C}$ — require arguments adapted to the case at hand. For brevity, we state as assumptions fairly weak sufficient conditions that can be verified in particular examples and that, in particular, are proved explicitly for the baseline model in the text. The relevant sufficient conditions are:

Assumption 3. The principal's recursive problem satisfies the following finiteness properties, for every $(\mathbf{v}, s) \in V \times S$:

(a) There exists an incentive compatible recursive contract ζ such that $Q^{\zeta}(\mathbf{v}, s) < +\infty$.

(b) The first-best value function satisfies

[i] $Q^*(\mathbf{v}, s) > -\infty$, and

[ii] For any recursive contract $\xi \in \Xi^{FB}(\mathbf{v})$,

$$\liminf_{t\to\infty} \alpha^t \left[\inf_{h\in\mathcal{H}} Q^*(\mathbf{v}^{(t)}, s^{(t)})\right] \geq 0$$

The first two pieces of Assumption 3 guarantee that the value function P is finite, and the final piece ensures that an appropriate transversality condition holds in the principal's recursive problem. These pieces correspond to, respectively, Proposition G.7, Lemma C.1, and Lemma C.3 for the baseline model. Similar proofs work for CARA utility with $n > 2$ types, and they are trivially satisfied if the consumption domain $\mathscr{C}$ is bounded (above, for the first piece; below, for the latter two).

Assumption 4. For every $(\mathbf{v}, s) \in V \times S$, the infimum in [**FE**] is attained.

Assumption 4 corresponds to Lemma D.5 for the baseline model. A similar proof works for CARA or CRRA utility with $n \geq 2$ types. The assumption holds trivially if the constraint correspondence $\Gamma(\cdot)$ is compact-valued.

Assumption 5. There exists a globally optimal recursive contract.

This corresponds exactly to Assumption 1 in the baseline model. Conditional on Assumption 4, Assumption 5 is satisfied trivially whenever $\mathscr{C}$ is bounded below, as this guarantees that an appropriate transversality condition holds for the agent's reporting problem.

With the appropriate sufficient conditions in hand, the next two propositions together constitute the analogues of Theorem 2 and Proposition 3.2 from the baseline model.

Proposition I.3. Let Assumptions 2 and 3 hold. Then, the principal's value function $P : V \times S \to \mathbb{R}$ satisfies the functional equation

[**FE**] $$P(\mathbf{v}, s) = \inf_{(\mathbf{x}_i, \mathbf{w}_i)\in\Gamma(\mathbf{v})} \sum_{i\in S} f_{si} \left[C(x_i) - \omega_i + \alpha P(\mathbf{w}_i, i)\right]$$

and, for each $s \in S$, $P(\cdot, s)$ is convex. Moreover, P is the pointwise smallest solution of [**FE**] that lies pointwise above the full-information value function Q^*.

Proof. The proposition can be established using straightforward adaptations of Lemmas D.1 and D.3. □

Proposition I.4. Let Assumptions 2, 3, and 4 hold. Then:

(a) For each $s \in S$, $P(\cdot, s)$ is continuously differentiable.

(b) The directional derivative $\mathsf{D}_{\mathbf{1}} P(\mathbf{v}, s)$ is non-negative.

(c) There exists a recursively optimal contract ξ^*, which is non-random and independent of the previous report $s \in S$.

(d) If, in addition, Assumption 5 holds, then (i) for each $s \in S$, $P(\cdot, s)$ is *strictly convex*, (ii) $\mathsf{D}_1 P(\mathbf{v}, s)$ is *strictly* positive, (iii) ξ^* is the *unique* recursively optimal contract, (iv) ξ^* is continuous in $\mathbf{v}$, and (v) $\tilde{x}_{\xi^*}$ is the unique solution to the sequence problem, [**SP**].

Proof Sketch. Note that the final piece of Assumption 3 suffices to establish the analogue of Lemma D.6. Then, follow the same arguments outlined in Appendix D.5, with Assumption 4 taking the role of Lemma D.5. □

I.3. Optimality Conditions

By Propositions I.3 and I.4, we have reduced the principal's problem to a smooth, convex, finite-dimensional minimization problem. The Lagrangian for this problem is

$$\mathscr{L}(\mathbf{v}, s) = \sum_{i=1}^{n} f_{si} \left[C(x_i) - \omega_i + \alpha P(\mathbf{w}_i, i)\right] + \sum_{i=1}^{n} \left[\lambda_i \left(v_i - x_i - \alpha \mathbb{E}^{\mathbf{f}_i} [\mathbf{w}_i]\right)\right]$$
$$- \sum_{i=2}^{n} \sum_{j=1}^{i-1} \left[\mu_{ij} \left(x_i + \alpha \mathbb{E}^{\mathbf{f}_i} [\mathbf{w}_i] - \psi_{ij}(x_j) - \alpha \mathbb{E}^{\mathbf{f}_i} \left[\mathbf{w}_j\right]\right)\right]$$

where $\lambda_i \in \mathbb{R}$ is the multiplier on the constraint [**PK**$_i$] for all $i \in S$, and $\mu_{ij} \geq 0$ is the multiplier on the constraint [**IC**$_{ij}$] for $i, j \in S$ with $i > j$. For notational ease, we extend μ_{ij} to all pairs $i, j \in \mathbb{Z}$, with the understanding that $\mu_{ij} = 0$ if (i) $i = 1$, (ii) $j \geq i$, or (iii) $i \notin S$ or $j \notin S$.

The optimality equations consist of the envelope conditions

[**Env**$_i$] $$P_i(\mathbf{v}, s) = \lambda_i$$

the first-order conditions for instantaneous utilities

[**FOC**x_i] $$f_{si} C'(x_i) = \lambda_i + \sum_{k=1}^{i-1} \mu_{ik} - \sum_{k=i+1}^{n} \psi'_{ki}(x_i) \mu_{ki}$$

and the first-order conditions for contingent continuation utilities

[**FOCw**$_{ij}$] $$f_{si} P_j(\mathbf{w}_i, i) = f_{ij} \left(\lambda_i + \sum_{k=1}^{i-1} \mu_{ik}\right) - \sum_{k=i+1}^{n} f_{kj} \mu_{ki}$$

for $i, j = 1, \ldots, n$.

The following consequence of the FOCs, which is analogous to part (b) of Proposition 5.1, will be crucial in the sequel.

Lemma I.5. Under the optimal contract, the value function satisfies

$$\frac{P_1(\mathbf{w}_n, n)}{f_{n1}} = \cdots = \frac{P_n(\mathbf{w}_n, n)}{f_{nn}}$$

Proof. Simply rearrange the FOCs [**FOCw**$_{ij}$] for $i = n$ and $j \in S$, recalling that $\mu_{ij} := 0$ for $i > n$. □

I.4. The Differential Martingale

We define $\mathsf{D}_{\mathbf{1}} P(\cdot, s)$ to be the directional derivative of $P(\cdot, s)$ in direction $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^n$. Because P is continuously differentiable, $\mathsf{D}_{\mathbf{1}} P(\mathbf{v}, s) = \sum_{i=1}^{n} P_i(\mathbf{v}, s)$. At points $\mathbf{v} \in \partial V$, the boundary of V, these definitions are adapted by using appropriate one-sided partial derivatives. Then, Theorem 3 generalizes to:

Theorem 7. *Let Assumptions 2 – 5 hold. Under the optimal contract, the process* $\mathsf{D}_{\mathbf{1}} P(\mathbf{v}^{(t)}, s^{(t)})$ *is a strictly positive martingale.*

Proof. Strict positivity follows from part (d) of Proposition I.3. Adding the [$\mathbf{Env}_i$] delivers

$$\mathsf{D}_{\mathbf{1}} P(\mathbf{v}, s) = \sum_{i=1}^{n} \lambda_i$$

For fixed $i \in S$, summing the [$\mathbf{FOCw}_{ij}$] over $j \in S$ gives

$$f_{si} \cdot \mathsf{D}_{\mathbf{1}} P(\mathbf{w}_i, i) = \lambda_i + \sum_{k=1}^{i-1} \mu_{ik} - \sum_{k=i+1}^{n} \mu_{ki}$$

The martingale property follows from summing over $i \in S$ and noting that

$$\sum_{i=1}^{n} \sum_{k=1}^{i-1} \mu_{ik} = \sum_{i=1}^{n} \sum_{k=i+1}^{n} \mu_{ki}$$

Integrability follows from non-negativity and the fact that the set of states S is finite. Thus, the process is a martingale. □

The following lemma extends [**H.5**] from Lemma H.2 to the case of n types and, moreover, shows that the differential martingale splits upon consecutive high-endowment reports.

Lemma I.6. Starting from any $(\mathbf{v}, s) \in V \times S$, let $\mathbf{w}_n$ denote the policy function if a shock type n is reported, and $\tilde{\mathbf{w}}_n$ the policy if another n shock is reported starting from the state $(\mathbf{w}_n, n)$. Under the optimal contract,

$$\mathsf{D}_{\mathbf{1}} P(\tilde{\mathbf{w}}_n, n) = \mathsf{D}_{\mathbf{1}} P(\mathbf{w}_n, n) + \frac{\sum_{k=1}^{n-1} \mu_{nk}(\mathbf{w}_n, n)}{f_{nn}}$$

Moreover, $\sum_{k=1}^{n-1} \mu_{nk}(\mathbf{v}, n) > 0$ for $\mathbf{v} \in V$, and thus

$$\mathsf{D}_{\mathbf{1}} P(\tilde{\mathbf{w}}_n, n) > \mathsf{D}_{\mathbf{1}} P(\mathbf{w}_n, n)$$

Proof Sketch. The proof of the first part follows exactly the argument from the proof of [**H.5**] from Lemma H.2. The argument for the second part follows the same lines as the proofs of Lemmas D.19 and D.20. □

I.5. Immiseration

We conclude with an extension of the immiseration result to this more general setting. Before stating the theorem, we define $\underline{v} := \underline{u}/(1-\alpha)$, which is the lifetime utility that the agent gets from receiving minimal consumption in every period.

Theorem 8. *Let Assumptions 2, 3, and 5 hold. Under the optimal contract, with probability one: (i)* $\mathsf{D}P(\mathbf{v}^{(t)}, s^{(t)}) \to \mathbf{0}$ *and thus* $\mathsf{D}_{\mathbf{1}}P(\mathbf{v}^{(t)}, s^{(t)}) \to 0$ *, (ii)* $v_i^{(t)} \to \underline{v}$ *for all* $i \in S$*, and (iii)* $x_i^{(t)} \to \underline{u}$ *for all* $i \in S$.

Note that, as in Lemma E.1 from Appendix E.2, the map $\mathsf{D}P(\cdot, n) : V \to Y := \mathsf{D}P(V, n)$ is a homeomorphism; exactly the same proof works in this more general setting.

Proof of Theorem 8. The first half of the proof is essentially identical to the proof of Theorem 4 in Appendix E.2. Using the same arguments, one can show that $\mathsf{D}_{\mathbf{1}}P(\mathbf{v}^{(t)}, s^{(t)}) \to 0$ along $\mathbb{P}$-a.e. path, and moreover that $\mathsf{D}P(\mathbf{v}^{(N_t)}, s^{(N_t)}) \to \mathbf{0}$ along the subsequence of dates (N_t) for which $s^{(N_t)} = n$. In particular, Lemma I.6 generates the requisite *martingale splitting* at state n and Lemma I.5 establishes that $\mathsf{D}_{\mathbf{1}}P(\mathbf{v}^{(N_t)}, s^{(N_t)}) \to 0$ implies $\mathsf{D}P(\mathbf{v}^{(N_t)}, s^{(N_t)}) \to \mathbf{0}$. An adaptation of Proposition K.2 ensures that the set of paths on which ω_n occurs infinitely often have full measure.

The remainder of the proof elaborates on the proof of Theorem 4 with the following argument, which first does induction through the type space at a fixed date, and then does induction across dates.

Base step: Let $\mathbf{w}_i^{(N_t)}$ denote the policy function for continuation utility after a report of type i, given that the current state is $(\mathbf{v}^{(N_t)}, s^{(N_t)})$. The FOC for $w_{11}^{(N_t)}$ is

$$f_{n1} P_1(\mathbf{w}_1^{(N_t)}, 1) = f_{11}\left(\lambda_1^{(N_t)} + 0\right) - \sum_{k=2}^{n} f_{k1}\mu_{k1}^{(N_t)}$$

Since $\lambda_1^{(N_t)} \to 0$ and $\mu_{ij} \geq 0$, it follows that

$$P_1(\mathbf{w}_1^{(N_t)}, 1) \to 0$$

and

$$\sum_{k=2}^{n} f_{k1}\mu_{k1}^{(N_t)} \to 0$$

Now, the FOC for $w_{12}^{(N_t)}$ is

$$f_{n1} P_2(\mathbf{w}_1^{(N_t)}, 1) = f_{12}\left(\lambda_1^{(N_t)} + 0\right) - \sum_{k=2}^{n} f_{k2}\mu_{k1}^{(N_t)}$$

As before, we have

$$P_2(\mathbf{w}_1^{(N_t)}, 1) \to 0$$

and

$$\sum_{k=2}^{n} f_{k2}\mu_{k1}^{(N_t)} \to 0$$

Continuing in this way through the FOCs for w_{1j}, $j \in S$, we see that

$$\mathsf{D}P(\mathbf{w}_1^{(N_t)}, 1) \to \mathbf{0}$$

and, because the Markov chain is fully connected,

$$\mu_{*,1}^{(N_t)} := \left(\mu_{1,1}^{(N_t)}, \ldots, \mu_{n,1}^{(N_t)}\right) \to \mathbf{0}$$

Now, consider the FOC for $w_{21}^{N_t}$:

$$f_{n2} P_1(\mathbf{w}_2^{(N_t)}, 2) = f_{21}\left(\lambda_2^{(N_t)} + \mu_{21}^{(N_t)}\right) - \sum_{k=3}^{n} f_{k1}\mu_{k2}^{(N_t)}$$

We have just shown that

$$\lambda_2^{(N_t)} + \mu_{21}^{(N_t)} \to 0$$

and the $\mu_{ij} \geq 0$, so it follows that $P_1(\mathbf{w}_2^{(N_t)}, 2) \to 0$ and $\sum_{k=3}^{n} f_{k1}\mu_{k2}^{(N_t)} \to 0$. Proceeding in this way iteratively through the FOCs for $w_{21}^{(N_t)}, \ldots, w_{2n}^{(N_t)}, w_{31}^{(N_t)}, \ldots, w_{3n}^{(N_t)}, \ldots, w_{nn}^{(N_t)}$, we see that

$$\mathsf{D}P(\mathbf{w}_s^{(N_t)}, s) \to \mathbf{0} \qquad \text{for all } s \in S$$

and

$$\mu_{*,s}^{(N_t)} := \left(\mu_{1,s}^{(N_t)}, \ldots, \mu_{n,s}^{(N_t)}\right) \to \mathbf{0} \qquad \text{for all } s \in S$$

Thus, it follows that, for any path under consideration,

$$\mathsf{D}P(\mathbf{v}^{(N_t+1)}, s^{(N_t+1)}) \to \mathbf{0}$$

Inductive step: Suppose we've shown that $\lambda_s^{(N_t+k)} \to 0$ and $\mu_{*,s}^{(N_t+k)} \to \mathbf{0}$ for all $s \in S$ and all $1 \leq k \leq M$. The proof for $N_t + M + 1$ follows exactly as in the base step.

Final step: We've shown that the vector of multipliers $\left(\lambda_s^{(t)}, (\mu_{s,j}^{(t)})_{j=1}^{s-1}\right) \to \mathbf{0}$ along the entire sequence. This completes part (i) of the theorem. Now, using [**FOC**x_i] and dropping time superscripts, we get

$$f_{si}C'(x_i) = \lambda_i + \underbrace{\sum_{k=1}^{i-1} \mu_{ik}}_{\to 0} - \underbrace{\sum_{k=i+1}^{n} \psi'_{ki}(x_i)\mu_{ki}}_{\geq 0}$$

By Assumption 2, this means that $x_i^{(t)} \to \underline{u}$, which is part (iii) of the theorem.

Part (ii) of the theorem then follows from part (iii) together with part (d) of Proposition I.4. Namely, because the induced allocation $\tilde{x}_{\xi^*}$ solves the sequence problem, it must satisfy the sequential version of promise-keeping. (See Appendix A for an exposition of the sequential problem in the context of the baseline model.) In particular, we have

$$v_i^{(t)} = x_i^{(t)} + \alpha\mathbb{E}\left[\sum_{k=1}^{\infty} \alpha^{k-1} x^{(t+k)} \middle| s^{(t)} = i\right]$$

All x terms on the RHS are converging to $\underline{u}$ by part (iii) of the present theorem, so it follows that $v_i^{(t)} \to \underline{v}$, which is part (ii). This completes the proof. □

J. Polar Coordinates

In this appendix, we revisit the baseline model and perform a change of variables to exploit the scale invariance implied by CARA utility and show that obtaining a one-dimensional representation of the problem remains intractable. See Section 6 in the main text for a motivating discussion.

Because of the homogeneity properties of CARA utility, the most natural candidate for a change of variables is polar coordinates. The idea is that a solution to the principal's problem for $\mathbf{v} \in V$ such that $||\mathbf{v}|| = 1$ (ie, those $\mathbf{v}$ which lie on the unit circle) will, by homogeneity, deliver a solution over the entire domain. In turn, the hope is that, because the value function is additively separable in the angular and radial components of $\mathbf{v}$, it might be possible to solve for each of these components in a separate subproblem. We show below that this does not appear to be possible. One intuition is that, while the principal's objective function is *additively separable* in the radial and angular components, her constraint set is *multiplicatively separable*, and thus optimization across these two components is inherently interconnected. The details are as follows.

For $\mathbf{v} \in V$ such that $\|\mathbf{v}\|_2 = 1$, let $\phi = \tan^{-1}(v_2/v_1)$. Now define $Q(\phi, s) := P(\mathbf{v}, s)$ for all $\mathbf{v} \in V$ with $\|\mathbf{v}\|_2 = 1$. For general $\mathbf{v} \in V$, part (d) of Theorem 2 says that under CARA preferences, the principal's value function takes the form

$$P(\mathbf{v}, s) = Q(\phi, s) - \frac{\log(r)}{1-\alpha}$$

where we've transformed to polar coordinates so that $\mathbf{v} = (r\cos(\phi), r\sin(\phi))$. Similarly, we transform $\mathbf{x} = (x_1, x_2) = (R\cos(\psi), R\sin(\psi))$. By part (f) of Theorem 2, the policy functions are homogenous of degree 1, so we want to write them in polar coordinates in a scale-free way. This will allow us to solve the principal's problem only when $r = 1$ and then recover the entire policy function by homogeneity. Namely, let $\overline{\rho} := R(r = 1)$ and $\rho_i := r_i(r = 1)$, i.e., the radial components of the control variables when the current $\mathbf{v}$ has $r = 1$.

Now, using the fact that $C(x) = -\log(x)$, we may write the Lagrangian (from Appendix H.1) as

$$\begin{aligned}
\mathscr{L}(r, \phi, s) = \; & f_{s1}\left[-\log(r\overline{\rho}\cos(\psi)) + \alpha Q(\phi_1, 1) - \frac{\alpha}{1-\alpha}\log(r\rho_1)\right] \\
& + f_{s2}\left[-\log(r\overline{\rho}\sin(\psi)) + \alpha Q(\phi_2, 2) - \frac{\alpha}{1-\alpha}\log(r\rho_2)\right] \\
& + \lambda_1\left[r\cos(\phi) - r\overline{\rho}\theta_1\cos(\psi) - \alpha r\rho_1\mathbf{f}_1 \cdot (\cos(\phi_1), \sin(\phi_1))\right] \\
& + \lambda_2\left[r\sin(\phi) - r\overline{\rho}\theta_2\sin(\psi) - \alpha r\rho_2\mathbf{f}_2 \cdot (\cos(\phi_2), \sin(\phi_2))\right] \\
& - \mu\left[r\overline{\rho}\theta_2\sin(\psi) + \alpha r\rho_2\mathbf{f}_2 \cdot (\cos(\phi_2), \sin(\phi_2)) - r\overline{\rho}\theta_2\cos(\psi) - \alpha r\rho_1\mathbf{f}_2 \cdot (\cos(\phi_1), \sin(\phi_1))\right]
\end{aligned}$$

Note that, the first two lines together simplify to

$$\begin{aligned}
& -\log(r) \cdot \frac{1}{1-\alpha} - \log(\overline{\rho}) + f_{s1}\left[-\log(\cos(\psi)) + \alpha Q(\phi_1, 1) - \frac{\alpha}{1-\alpha}\log(\rho_1)\right] \\
& + f_{s2}\left[-\log(\sin(\psi)) + \alpha Q(\phi_2, 2) - \frac{\alpha}{1-\alpha}\log(\rho_2)\right]
\end{aligned}$$

Define $\tilde{\lambda_1} = \lambda_1 r$ and similarly for the other multipliers. By Theorem 2(e), we know that the multipliers are homogenous of degree -1, so that $\tilde{\lambda_1}$ is just the PK1 multiplier when $r = 1$, etc. We then have the envelope conditions

[**PK**r] $$P_r(r,\phi,s) = -\frac{1}{1-\alpha}\cdot\frac{1}{r}$$

[**PK**ϕ] $$Q'(\phi,s) = -\tilde{\lambda}_1\sin(\phi) + \tilde{\lambda}_2\cos(\phi)$$

the FOCs for instantaneous utilities

[**FOC**$\overline{\rho}$] $$\frac{1}{\overline{\rho}} = -\tilde{\lambda}_1\theta_1\cos(\psi) - \tilde{\lambda}_2\theta_2\sin(\psi) - \tilde{\mu}\cdot[\theta_2\sin(\psi) - \theta_2\cos(\psi)]$$

[**FOC**ψ]
$$0 = f_{s1}\tan(\psi) - f_{s2}\text{cotan}(\psi) + \tilde{\lambda}_1\overline{\rho}\theta_1\sin(\psi) - \tilde{\lambda}_2\overline{\rho}\theta_2\cos(\psi) - \tilde{\mu}\overline{\rho}\theta_2\cos(\psi) - \tilde{\mu}\overline{\rho}\theta_2\sin(\psi)$$

and four similar FOCs for the continuation utility variables $(\phi_1, \phi_2, r_1, r_2)$.

We then have 11 equations: the 8 optimality conditions, the two promise-keeping constraints, and one (binding) incentive constraint. A *solution* consists of 6 functions:

- Three functions $\rho_1(\cdot)$, $\rho_2(\cdot)$, and $\overline{\rho}(\cdot)$, which determine the optimal radial components given the current angular component ϕ.
- Three functions $\phi_1(\cdot)$, $\phi_2(\cdot)$, and $\psi(\cdot)$, which determine the optimal angular components given the current angular component ϕ.

To obtain such a solution, we would need to "solve for" 4 other functions: the "angular value function" $Q(\cdot)$ and the three multiplier functions $\tilde{\lambda}_1(\cdot)$, $\tilde{\lambda}_2(\cdot)$, and $\tilde{\mu}(\cdot)$. The optimality conditions are a non-linear, coupled system of differential equations for these functions. We do not know of any methods for de-coupling them across the angular and radial dimensions.

K. Some Pathwise Properties of Markov Chains

In this section, we collect some miscellaneous facts about paths of Markov chains. Let E be the (countable) state space for a Markov process with transition probabilities $P(x, B)$ denoting the probability of transitioning from x to B. Let $\mathbf{P}$ denote the induced probability measure on the path space E^∞.

Lemma K.1. Let (X_n) be an E-valued Markov process with transitions given by the kernel P, and suppose $x \in E$ is *recurrent*. Then,

$$\mathbf{P}(X_n = x \text{ for infinitely many } n \mid X_0 = x) = 1$$

An elementary proof can be found on p. 577 of Shiryaev (1995).

Proposition K.2. The event $\{(s_{t-1}, s_t) = (i, j) \text{ for infinitely many } t\}$ occurs almost surely for all $i, j \in S$. Therefore, the event $\{s_t = i \text{ for infinitely many } t\}$ occurs almost surely for all $i \in S$.

(Here, 'almost surely' is with respect to the probability measure $\mathbf{P}$ induced on S^∞ by the Markov kernel.) Intuitively, the proposition says that low or high income shocks occur infinitely often with probability one. Moreover, consecutive low or high income shocks as well as low-high and high-low income shocks also occur infinitely often with probability one.

Proof of Proposition K.2. Recall that $S = \{1, 2\}$ with transition probabilities $P(i, j)$ represented by the transition matrix

$$\begin{array}{cc} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix} \end{array}$$

As the statement of the lemma notes, it suffices to prove that pairs of shocks of the form (i, j) occur infinitely along almost every sample path.

Towards the proof of this claim, it is useful to consider the *bivariate* Markov chain with states $E = \{11, 12, 21, 22\}$ and transition probabilities Q given by

$$\begin{array}{cc} & \begin{array}{cccc} 11 & 12 & 21 & 22 \end{array} \\ \begin{array}{c} 11 \\ 12 \\ 21 \\ 22 \end{array} & \begin{pmatrix} f_{11} & f_{12} & 0 & 0 \\ 0 & 0 & f_{21} & f_{22} \\ f_{11} & f_{12} & 0 & 0 \\ 0 & 0 & f_{21} & f_{22} \end{pmatrix} \end{array}$$

The transition matrix reflects the fact that one-step transitions have a simple form, namely

$$Q\big((i, j), (k, \ell)\big) = \mathbf{1}\{j = k\}P(k, \ell)$$

where $\mathbf{1}\{j = k\} = 1$ if $j = k$ and 0 otherwise. Then, the two-step transition probabilities are given by

$$Q^{(2)}\big((i, j), (k, \ell)\big) = P(j, k)P(k, \ell) > 0$$

Therefore, all states communicate with each other, which implies that the Markov chain is indecomposable. But because the state space E is finite, by Theorem 1 (and the subsequent discussion) on p. 580 of Shiryaev (1995), at least one of the states must be recurrent. The indecomposability of the process then implies that all states are recurrent. An application of Lemma K.1 completes the proof. □

References

Abreu, Dilip, David Pearce and Ennio Stacchetti (1990). 'Toward a Theory of Discounted Repeated Games with Imperfect Monitoring'. In: *Econometrica* 58.5, pp. 1041–1063 (cit. on p. 30).

Atkeson, Andrew and Robert E. Lucas Jr. (1992). 'On Efficient Distribution with Private Information'. In: *Review of Economic Studies* 59.3, pp. 427–453 (cit. on pp. 1, 2, 4, 13, 24–26, 29, 54).

Atkeson, Andrew and Robert E Lucas Jr (1995). 'Efficiency and Equality in a Simple Model of Efficient Unemployment Insurance'. In: *Journal of Economic Theory* 66, pp. 64–88 (cit. on p. 4).

Battaglini, Marco (2005). 'Long-Term Contracting with Markovian Consumers'. In: *American Economic Review* 95.3, pp. 637–658 (cit. on pp. 5, 6, 22, 23).

Battaglini, Marco and Rohit Lamba (2014). *Optimal Dynamic Contracting: The First-Order Approach and Beyond*. Tech. rep. (cit. on pp. 5, 21).

Border, Kim C. (2013). *Alternative Linear Inequalities*. Tech. rep. Caltech (cit. on p. 31).

Broer, Tobias, Marek Kapička and Paul Klein (2017). 'Consumption risk sharing with private information and limited enforcement'. In: *Review of Economic Dynamics* 23, pp. 170–190 (cit. on p. 4).

Di Tella, Sebastian and Yuliy Sannikov (2015). *Optimal Asset Managment Contracts with Hidden Savings*. Tech. rep. Stanford University GSB (cit. on p. 7).

Doepke, Matthias and Robert M Townsend (2006). 'Dynamic mechanism design with hidden income and hidden actions'. In: *Journal of Economic Theory* 126.1, pp. 235–285 (cit. on p. 4).

Esö, Péter and Bal'asz Szentes (2017). 'Dynamic contracting: An irrelevance theorem'. In: *Theoretical Economics* 12, pp. 109–139 (cit. on p. 23).

Farhi, Emmanuel and Iván Werning (2007). 'Inequality and Social Discounting'. In: *Journal of Political Economy* 115.3, pp. 365–402 (cit. on pp. 4, 26).

– (2013). 'Insurance and Taxation over the Lifecycle'. In: *Review of Economic Studies* 80.2, pp. 596–635 (cit. on p. 5).

Farinha Luz, Vitor (2015). *Dynamic Competitive Insurance*. Tech. rep. Yale University (cit. on p. 6).

Feldman, Naomi E and Joel Slemrod (2007). 'Estimating tax noncompliance with evidence from unaudited tax returns'. In: *The Economic Journal* 117.518, pp. 327–352 (cit. on p. 7).

Fernandes, Ana and Christopher Phelan (2000). 'A Recursive Formulation for Repeated Agency with History Dependence'. In: *Journal of Economic Theory* 91, pp. 223–247 (cit. on pp. 4, 13).

Fu, Shiming and R. Vijay Krishna (2014). *Dynamic Financial Contracting with Persistence*. Tech. rep. (cit. on pp. 5, 14, 22, 23).

Golosov, Mikhail and Aleh Tsyvinski (2006). 'Designing optimal disability insurance: A case for asset testing'. In: *Journal of Political Economy* 114.2, pp. 257–279 (cit. on p. 4).

Golosov, Mikhail, Aleh Tsyvinski and Nicolas Werquin (2016). 'Recursive Contracts and Endogenously Incomplete Markets'. In: *Handbook of Macroeconomics, Volume 2* (cit. on p. 4).

Green, Edward J. (1987). 'Lending and the Smoothing of Uninsurable Income'. In: *Contractual Agreements for Intertemporal Trade*. Ed. by Edward C. Prescott and Neil Wallace. University of Minnesota Press (cit. on pp. 3, 7, 13, 21).

Guo, Yingni and Johannes Hörner (2014). *Dynamic Mechanisms without Money*. Tech. rep. (cit. on pp. 6, 16, 22, 23).

Halac, Marina and Pierre Yared (2014). 'Fiscal rules and discretion under persistent shocks'. In: *Econometrica* 82.5, pp. 1557–1614 (cit. on p. 4).

Hatcher, Allen (2001). *Algebraic Topology*. Cambridge University Press (cit. on p. 52).

Kamihigashi, Takashi (2014). 'Elementary results on solutions to the Bellman equation of dynamic programming: existence, uniqueness, and convergence'. In: *Economic Theory* 56.2, pp. 251–273 (cit. on p. 41).

Kapička, Marek (2013). 'Efficient allocations in dynamic private information economies with persistent shocks: A first-order approach'. In: *The Review of Economic Studies*, rds045 (cit. on p. 5).

Kocherlakota, Narayana R (2010). *The New Dynamic Public Finance*. Princeton, NJ: Princeton University Press (cit. on p. 4).

Kreps, David M. (2012). *Microeconomic Foundations I: Choice and Competitive Markets*. Princeton, NJ: Princeton University Press (cit. on pp. 42, 43).

Lucas Jr., Robert E. (1992). 'On Efficiency and Distribution'. In: *Economic Journal* 102.411, pp. 233–247 (cit. on p. 2).

Meghir, Costas and Luigi Pistaferri (2004). 'Income variance dynamics and heterogeneity'. In: *Econometrica* 72.1, pp. 1–32 (cit. on p. 2).

Pavan, Alessandro (2016). 'Dynamic mechanism design: Robustness and endogenous types'. In: *2015 World Congress of the Econometric Society* (cit. on p. 5).

Pavan, Alessandro, Ilya Segal and Juuso Toikka (2014). 'Dynamic Mechanism Design: A Myersonian Approach'. In: *Econometrica* 82.2, pp. 601–653 (cit. on pp. 5, 8, 21, 22).

Phelan, Christopher (1998). 'On the Long Run Implications of Repeated Moral Hazard'. In: *Journal of Economic Theory* 79, pp. 174–191 (cit. on pp. 6, 21, 24, 25, 54, 65).

– (2006). 'Opportunity and social mobility'. In: *Review of Economic Studies* 73.2, pp. 487–504 (cit. on p. 4).

Rochet, Jean-Charles and Lars A Stole (2003). 'The Economics of Multidimensional Screening'. In: *Econometric Society Monographs* 35, pp. 150–197 (cit. on p. 5).

Shiryaev, Albert N (1995). *Probability*. Springer (cit. on pp. 53, 73, 74).

Stokey, Nancy L., Robert E. Lucas Jr and Edward C. Prescott (1989). *Recursive Methods in Economic Dynamics*. Cambridge, Ma: Harvard University Press (cit. on pp. 43, 45).

Storesletten, Kjetil, Christopher I. Telmer and Amir Yaron (2004a). 'Consumption and risk sharing over the life cycle'. In: *Journal of Monetary Economics* 51.3, pp. 609–633 (cit. on p. 2).

– (2004b). 'Cyclical dynamics in idiosyncratic labor market risk'. In: *Journal of Political Economy* 112.3, pp. 695–717 (cit. on p. 2).

Thomas, Jonathan P. and Tim Worrall (1990). 'Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-Agent Problem'. In: *Journal of Economic Theory* 51, pp. 367–390 (cit. on pp. 2–4, 6, 8, 13, 14, 16, 18, 20, 22, 23, 65).

Van Tiel, Jan (1984). *Convex Analysis*. Wiley (cit. on p. 52).

Vohra, Rakesh (2004). *Advanced Mathematical Economics*. Routledge (cit. on p. 31).

Williams, Noah (2011). 'Persistent Private Information'. In: *Econometrica* 79.4, pp. 1233–1275 (cit. on pp. 5, 6, 15, 22).

Williams, Noah (2015). ‘A Solvable Continuous Time Dynamic Principal-Agent Model’. In: *Journal of Economic Theory* (cit. on p. 7).
Zhang, Yuzhe (2009). ‘Dynamic Contracting with Persistent Shocks’. In: *Journal of Economic Theory* 144, pp. 635–675 (cit. on pp. 4, 5, 16).