Persistent Private Information Revisited*

Alexander W. Bloedel† R. Vijay Krishna‡ Bruno Strulovici§

November 3, 2020

Abstract

This paper revisits Williams' (2011) continuous-time model of optimal dynamic insurance with persistent private information and corrects several errors in that paper's analysis. We introduce and study the class of *self-insurance contracts* that are implementable as consumption-saving problems for the agent with constant taxes on savings chosen by the principal. We show that the contract asserted to be optimal in Williams (2011) is the special self-insurance contract with zero taxes. When the agent's private endowment is mean-reverting, that contract is strictly dominated by the optimal self-insurance contract, which imposes a strictly positive tax, induces immiseration when the rate of mean-reversion is high, and sends the agent to bliss when the rate of mean-reversion is low. When the agent's endowment is not mean-reverting, the contract derived in that paper is, in fact, optimal among all incentive compatible contracts; we provide a new explanation for its properties in terms of the agent's indifference among all reporting strategies. These results extend to the natural discrete-time analogue of the model. Separately, Williams' (2011) first-order approach to incentive compatibility relies on an erroneous and unjustified assumption on the space of feasible reporting strategies; our analysis does not.

(*) Subsections 3.3 and 4.2 of this paper build on Strulovici (2011) (the working paper version of Strulovici 2020) and subsume its analysis of the long-run properties of state-consistent contracts. Subsection 5.2 of this paper expands on and subsumes Section 6.2 of Bloedel, Krishna, and Leukhina (2018) (the working paper version of Bloedel, Krishna, and Leukhina 2020). We thank Arash Fahim and Ilya Segal for several helpful conversations.

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

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

(§) Northwestern University. Email: b-strulovici@northwestern.edu

1. Introduction

In an important and influential paper, Williams (2011) (henceforth W11) introduces continuous-time methods to study optimal dynamic contracts in settings with persistent private information, methods that have since been fruitfully applied to shed light on a range of difficult problems in contract theory.[1] These methods are applied in W11 to solve for the optimal contract in the canonical setting of a risk-neutral principal providing insurance to a risk-averse agent who faces a stochastic, privately-observed endowment stream (Green 1987; Thomas and Worrall 1990). In contrast to the classic finding that optimal insurance contracts lead to *immiseration*, wherein the agent's consumption and utility almost surely decrease without bound, the optimal contract in W11 sends the agent to *bliss*: with probability one, the agent's consumption and utility converge to their upper bounds. It is argued in W11 that these findings differ from those in the classic literature due to two novel features of its environment: (i) private information is persistent, rather than iid, and (ii) the model is cast in continuous, rather than discrete, time.

Unfortunately, the analysis in W11 contains several errors that implicate its main conclusions. Most importantly, except in the non-generic case where shocks to the agent's endowment are permanent, the contract asserted to be optimal in W11 based on a numerical observation (defined as Contract W in Subsection 2.2 below) is strictly suboptimal. Moreover, even in the special case of permanent shocks, the qualitative conclusions about the role of continuous-time modeling and persistent private information derived in W11 from Contract W are misleading. This is important, in part, because the results and intuitions stated in W11 appear to undermine the generality of the classic immiseration results, which are central to economists' understanding of dynamic insurance and have attracted substantial interest in the literature.[2] Separately, the first-order approach to incentive compatibility taken in W11 hinges on an erroneous restriction on the agent's reporting strategies, calling into question that paper's characterization of implementable contracts (see Remark 1 in Subsection 2.1 below).

This paper address these issues in the context of the main example from W11, in which the agent has exponential utility and his private endowment follows an Ornstein-Uhlenbeck process, which reduces to a Brownian motion when mean-reversion is zero.[3]

(1) See, for instance, Sannikov (2014), Prat and Jovanovic (2014), Williams (2015), DeMarzo and Sannikov (2016), He et al. (2017), Cisternas (2017), and Chen (2017).

(2) For instance, see Green (1987), Thomas and Worrall (1990), Atkeson and Lucas (1992), Atkeson and Lucas (1995), Phelan (1998), Phelan (2006), Farhi and Werning (2007), and Zhang (2009). Bloedel, Krishna, and Leukhina (2020) contains a more comprehensive literature review.

(3) This leading hidden endowment example is studied in Sections 6-7 of W11. Importantly, our observa-

We begin by observing that Contract W can be implemented as a standard consumption-saving problem for the agent in which he saves at the market rate. We then ask, and answer, the following three main questions (which are detailed further below):

1. *Is Contract W optimal among all contracts that can be implemented as consumption-saving problems?* We show that, generically, it is not: when the endowment is mean-reverting, Contract W can be improved upon by imposing a constant tax on the agent's savings. We also characterize the full class of contracts that can be implemented in this manner and solve for the optimal one.
2. *Are there conditions under which Contract W would, in fact, be optimal within the full class of incentive compatible contracts?* We provide two conditions when this is true: (a) the agent can covertly save and borrow outside of the contract, or (b) the agent's endowment has zero mean-reversion. In the latter case, we provide new explanations for the optimality of Contract W that differ from those provided in W11.
3. *Do these results hinge on the model being cast in continuous time, as suggested in W11?* We show that they do not: all of the aforementioned results hold in the discrete-time analogue of the model, indicating that the basic incentive problem operates similarly in both discrete and continuous time.

Self-Insurance Contracts. Our main analysis studies a specific class of *Self-Insurance Contracts*, which are defined in terms of the following indirect implementation: The principal provides the agent with some initial wealth and then allows the agent to self-insure at a risk-free interest rate — which may differ from the market rate at which the principal discounts — determined by a constant marginal tax imposed on the agent's capital gains. The principal aims to minimize the cost of her initial wealth transfer net of her lifetime tax revenue. Self-Insurance Contracts are incentive compatible when implemented as direct-revelation mechanisms, essentially by construction; under-reporting the endowment in the direct mechanism corresponds to under-saving (and over-consuming) in the indirect implementation.[4]

tions concerning the restrictions imposed by W11 on the agent's feasible reporting strategies (Remark 1 and Appendix K) implicate the general framework studied in Sections 2-5 of W11, and our results for and discussions concerning the permanent-shock case (Subsections 5.2 and Appendix K) apply equally well to the taste-shock example studied in Section 8 of W11.

(4) Consistent with Strulovici's (2020) analysis of incentive compatibility for state-consistent contracts (described below), we show that Self-Insurance Contracts remain incentive compatible (as direct mechanisms) even when the agent's feasible set of reporting strategies is enlarged to allow for deviations involving zero-probability events, such as jumps in the reported endowment, that are ruled out in our baseline formulation of incentive compatibility (which largely follows that of W11) because the principal could detect them (and punish the agent) in real-time. This robustness arises from the

Our first observation is that Contract W can be implemented as the special Self-Insurance Contract in which the principal does not tax the agent, who simply self-insures at the market rate. An immediate implication is that Contract W is, in fact, incentive compatible. This is important because the first-order approach to incentive compatibility taken in W11 relies on an unintended and erroneous restriction on the agent's feasible set of reporting strategies that prevents the agent from correcting for past misreports, thereby ruling out temporary deviations from truthful reporting. When this restriction is dropped, we show that (i) the verification given in W11 that Contract W is incentive compatible is incorrect as stated (except when mean-reversion is exactly zero), and (ii) the proof of that paper's general verification theorem for the first-order approach (Theorem 1 in W11) is incomplete.

Our reformulation of Contract W as a Self-Insurance Contract fills the former gap, while also facilitating a novel interpretation of the contract's key properties. For instance, the agent's (promised) utility defines a martingale under Contract W because his optimal consumption-smoothing induces the standard Euler equation, while the agent's consumption has positive drift and almost surely increases without bound due to his *precautionary saving* motive (which also leads him to acquire unbounded assets in the long run). These long-run properties are consistent with the literature on incomplete-market self-insurance problems (e.g., Chamberlain and Wilson 2000; Ljunqvist and Sargent 2000, Ch. 17).

Building on the above observation, we show that Contract W is strictly dominated by the optimal Self-Insurance Contract whenever the agent's endowment has non-zero mean-reversion (the "transient shock case"). We characterize the optimal Self-Insurance Contract in this case: it imposes a strictly positive tax on the agent's savings, which induces a negative drift in his (promised) utility but facilitates better risk-sharing. While the optimal tax rate is non-monotone in the persistence of the agent's endowment, the agent's risk exposure and precautionary savings rate are increasing in persistence. This latter feature causes the optimal Self-Insurance Contract to induce immiseration when persistence is below a cutoff (as in the classic literature) and send the agent to bliss when persistence is above the cutoff (as under Contract W). These long-run properties contrast starkly with those of fully optimal contracts (which generally do not admit self-insurance implementations) in the natural discrete-time and -type analogue of the model, which Bloedel, Krishna, and Leukhina (2020) have shown induce immiseration for general

fact that the agent's strategy space in the self-insurance problem, which allows for consumption processes with jumps, is naturally larger than that in the direct mechanism, where the aforementioned detectability of deviations imposes additional constraints on the agent.

Markovian type processes with arbitrary (imperfect) persistence.

On the other hand, we show that Contract W coincides with the optimal Self-Insurance Contract when the agent's endowment has exactly zero mean-reversion (the "permanent shock" case). This occurs because the principal cannot decrease the agent's risk exposure when shocks are permanent, so that imposing a non-zero tax rate simply distorts the drift of the agent's consumption, which is inefficient.

Are Self-Insurance Contracts fully optimal? Under technical conditions, it can be shown that the answer is generically negative: the optimal Self-Insurance Contract is strictly dominated whenever shocks are transient (see Remark 4). (As described further below, the answer is positive when shocks are permanent.)

Nonetheless, Self-Insurance Contracts have severable desirable properties that motivate our study of them, independently of their relation to Contract W. First, they are highly tractable because the the agent's optimal consumption-saving decision arises as the solution to a standard self-insurance problem, which we solve in closed form.[5] Second, they are "simple" in the sense that they do not require the highly history-dependent and non-linear tax schemes that are generally needed to implement full-commitment optima in related settings (cf. Farhi and Werning 2013). Third, we show that, perhaps surprisingly, the class of Self-Insurance Contracts is equivalent to two other natural classes of "simple" contracts with seemingly unrelated foundations: (i) the class of *Stationary Contracts* that we introduce, in which the agent's "information rent" is proportional to his promised utility (a near-defining property of Contract W in W11's analysis), and (ii) the class of "state-consistent" renegotiation-proof contracts introduced by Strulovici (2020), which are defined axiomatically in terms of the principal's limited commitment.

A corollary of our analysis is that, generically, the optimal full-commitment contract in the W11 model does not exhibit these forms of tractability or simplicity. While a full characterization of that contract (under transient shocks) is beyond the scope of this paper, we view our analysis as an important first step in that direction and provide several conjectures about its properties (see Sections 5.2.1 and 6).

When is Contract W fully optimal? Building on the above analysis, we identify two cases in which Contract W can, in fact, be derived as an optimal full-commitment contract. For technical reasons, we focus on contracts that are incentive compatible and, in addition,

(5) The solution we obtain is the continuous-time limit of the self-insurance solutions obtained by Caballero (1990) and Wang (2003) in discrete-time settings where the agent's endowment follows a Gaussian AR(1) process (and generalizations thereof).

satisfy a "first-order" incentive compatibility condition that we call *FO-IC* (cf. Remark 3).[6]

The first case is that in which the agent can covertly borrow and save directly at the market rate, in addition to possibly misreporting his endowment to the principal (cf. Allen 1985; Cole and Kocherlakota 2001). In this setting, any FO-IC contract that deters hidden borrowing and saving (which is without loss by a standard Revelation Principle argument) must satisfy the agent's Euler equation, as the agent could always circumvent punishments imposed by the contract by self-insuring in the ambient market. Thus, these additional incentive constraints endogenously and uniquely pin down Contract W.

The second case is that in which shocks to the agent's endowment are permanent (i.e., have zero mean-reversion) but, as in the baseline model of W11, the agent does *not* have access to hidden savings. In this case, we show that Contract W is, in fact, the optimal FO-IC contract as asserted in W11. This illustrates that permanent shocks overturn two pieces of received wisdom in the literature: (i) that hidden savings and renegotiation-proofness are binding constraints on the principal, and (ii) that optimal full-commitment contracts induce immiseration. We argue that both properties derive from the fact that, as we show, the agent is necessarily indifferent among (essentially) all reporting strategies given any FO-IC contract when shocks are permanent. This feature prevents the principal from manipulating the agent's risk exposure and eliminates the standard channel underlying immiseration, wherein high-powered incentives for the agent are optimally backloaded so as to facilitate better risk-sharing in early periods. More formally, it causes the principal's marginal cost process to violate the martingale property that Thomas and Worrall (1990) and Bloedel, Krishna, and Leukhina (2020) have shown is intimately related to the immiseration results. Building on our (partial) characterization of FO-IC contracts and Bloedel, Krishna, and Leukhina's (2020) martingale convergence arguments, we argue that the failure of immiseration is special to the permanent shock case and that immiseration is likely to be restored when either one of the assumptions of permanent shocks and exponential utility is relaxed.

The relation between continuous and discrete time. Finally, our analysis sheds light on a central claim in W11: that a primary reason for the differences between Contract W and optimal full-commitment contracts in related discrete-time models is that the agent's

(6) While it is intuitive that incentive compatibility implies first-order incentive compatibility, to formally establish this we must show that the agent's value function is suitably differentiable. We conjecture that this can be done but, as described in Remark 3, the present model does not satisfy the sufficient conditions for such differentiability typically invoked in the literature. Therefore, we restrict attention to contracts that are sufficiently well-behaved for this implication to hold.

incentive constraints are qualitatively different in continuous time. As noted in Section 6, the class of Self-Insurance Contracts has a precise analogue in the discrete-time version of the model, in which the optimal Self-Insurance Contract has similar properties to those described above and, critically, still strictly dominates the discrete-time analogue of Contract W. Our analysis of the optimal FO-IC contract under permanent shocks also naturally extends to the discrete-time model. Based on these observations, we conclude that the analysis of W11 does not have any direct implications for the relationship between discrete- and continuous-time dynamic contracting models with private information. Online Appendix K expands further on this point.

Outline. The rest of the paper is organized as follows. Section 2 presents the model and Contract W. Section 3 presents our main analysis of Self-Insurance Contracts. Section 4 presents the FOA to incentive compatibility and equivalent characterizations of the class of Self-Insurance Contracts. Section 5 presents our analysis of optimal FO-IC contracts. Section 6 concludes. Most proofs and some auxiliary results are contained in Appendices B–C and Online Appendices D–J. A detailed accounting of the errors in W11 is deferred to Online Appendix K. A few additional proofs and results are in the Supplementary Appendix (Bloedel, Krishna, and Strulovici 2020).

2. Preliminaries

Subsection 2.1 introduces the special case of the model in W11 in which the agent has exponential utility and his private information concerns his endowment, which is the setting of the main solved example in W11. Subsection 2.2 then introduces the contract that is asserted to be optimal in W11. We mostly follow the presentation in W11, but directly formulate the model over an infinite time horizon and correct some errors in its development of reporting strategies.[7] (We defer introducing a treatment of first-order incentive compatibility to Section 4.)

2.1. Environment

Time is continuous and runs over an infinite horizon. At $t = 0$, a risk-neutral principal (she) offers an insurance contract to a risk-averse agent (he), whose stochastic endowment stream is his private information. The principal aims to minimize her expected costs

(7) Sections 2-4 of W11 introduce the model and conditions for incentive compatibility over a finite time horizon $[0, T]$. Sections 5-6 and 8 of W11 then heuristically take the $T \to \infty$ limit to study applications.

subject to delivering a pre-specified lifetime utility promise to the agent. Once the contract is signed, neither party may renege at a later date. By the Revelation Principle, we may focus on truthful direct-revelation contracts.

We describe the agent's reporting problem and the principal's contracting problem below. Additional details are in Online Appendix E.

Type Process. At each time t, the agent receives a random endowment, $b_t \in \mathbb{R}$, of a consumption good. The agent's *endowment process* $b = (b_t)_{t \geqslant 0}$ evolves according to the SDE

$$\mathrm{d}b_t = (\mu - \lambda b_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t \tag{2.1}$$

where $\sigma > 0$. It is common knowledge that the initial endowment is b_0. The parameter $\lambda \geqslant 0$ specifies the rate of mean reversion. When $\lambda = 0$, b defines a Brownian motion with constant drift. We refer to this as the *permanent shock* case because the time-t shock (the Brownian increment $\mathrm{d}W_t$) has a non-vanishing additive effect on all b_T with $T > t$. When $\lambda > 0$, b defines an Ornstein-Uhlenbeck (OU) process. We refer to this as the *transient shock* case because the time-t shock has a vanishing effect on b_T as $T - t \to \infty$. Note that *smaller* values of λ correspond to *greater* persistence.[8]

Reporting Strategies. At each time t, the agent reports that his endowment is $y_t \in \mathbb{R}$. Call $y = (y_t)_{t \geqslant 0}$ the agent's *reporting process (or strategy)* and define the *misreporting process (or strategy)* $m = (m_t)_{t \geqslant 0}$ by $m_t := y_t - b_t$; both m and y are b-adapted. Thus, m_t is the (negative of the) amount of consumption good that the agent "diverts" at time t.

Each misreporting strategy m induces a probability measure P^m over paths of y. Let $\mathsf{P}^* := \mathsf{P}^{m^*}$ denote the measure over paths of y that arises under the *truthful strategy* $m^* = 0$, which corresponds to $y^* = b$. The agent is restricted to choosing from among misreporting strategies m for which $\mathsf{P}^m \ll \mathsf{P}^*$. As in W11, we also assume that the agent's misreports have absolutely continuous paths, i.e., there exists a process $\Delta = (\Delta_t)_{t \geqslant 0}$ such that $m_t \equiv \int_0^t \Delta_\tau \,\mathrm{d}\tau$ (where "$\equiv$" denotes a.s. equality for all $t \geqslant 0$). Thus, the agent's report evolves as $\mathrm{d}y_t = \mathrm{d}b_t + \Delta_t \mathrm{d}t$, where the drift adjustment Δ_t corresponds to misreporting the *increment* $\mathrm{d}b_t$. Let $\mathscr{M}$ denote the set of *admissible* misreporting strategies that satisfy the above conditions. It is easy to see that $m^* \in \mathscr{M}$.

(8) In particular, because b is Markov, for any given realization of b_t and $T > t$, we may write the solution to [2.1] as $b_T = \mu/\lambda + (b_t - \mu/\lambda)e^{-\lambda(T-t)} + e^{-\lambda(T-t)} \int_t^T \sigma e^{\lambda\tau}\,\mathrm{d}W_\tau$. Thus, we see that $\frac{\mathrm{d}b_T}{\mathrm{d}b_t} = e^{-\lambda(T-t)}$. In the language of Pavan, Segal, and Toikka (2014), $e^{-\lambda(T-t)}$ is the "impulse response" of b_T to b_t. Thus, smaller λ corresponds to larger impulse responses; when $\lambda = 0$, the impulse responses are identically 1.

Agent's Preferences. The agent discounts utility from consumption exponentially at rate $\rho > 0$. The agent's Bernoulli utility over consumption $u : \mathbb{R} \to \mathbb{R}_{--}$ is exponential, $u(c) := e^{-\theta c}$, where $\theta > 0$ is the coefficient of absolute risk aversion.[9] Note that $u'(c) = -\theta u(c)$.

Contracts. A *contract* is a y-adapted process $s = (s_t)_{t \geqslant 0}$ that specifies the transfer from principal to agent. Note that, because s is y-adapted, it may depend arbitrarily on the agent's past and current reports, but not his true endowment history. A contract is *incentive compatible (IC)* if it satisfies[10]

$$m^* \in \arg\max_{m \in \mathcal{M}} \mathbf{E}_0^m \left[\int_0^\infty e^{-\rho t} u(s_t + y_t - m_t) \, \mathrm{d}t \right] \tag{IC}$$

Define the agent's *(recommended) consumption process* $c = (c_t)_{t \geqslant 0}$ by $c_t := s_t + y_t$ and his *(recommended) flow utility process* $u = (u_t)_{t \geqslant 0}$ by $u_t := u(c_t)$. These are the consumption and flow utility processes induced under truthful reporting; the agent's actual consumption is $c_t - m_t$.[11] As is standard, it is also useful to define his *promised utility* $q = (q_t)_{t \geqslant 0}$ by

$$q_t := \mathbf{E}_t^* \left[\int_t^\infty e^{-\rho(\tau - t)} u_\tau \, \mathrm{d}\tau \right] \tag{2.2}$$

which is his lifetime continuation utility under truthful reporting.

Principal's Problem. The principal also discounts exponentially at rate $\rho > 0$. A common interpretation, which will be important in the subsequent analysis, is that $\rho > 0$ represents the interest rate at which the principal can borrow and save on a risk-free bond market. She chooses a contract so as to minimize the expected lifetime cost of transfers to the agent (assuming truthful reporting)

$$\mathbf{E}_0^* \left[\int_0^\infty e^{-\rho t} \underbrace{(c_t - y_t)}_{= s_t \text{ under } m^*} \mathrm{d}t \right] \tag{2.3}$$

(9) We let $\mathbb{R}_{--} := \{x \in \mathbb{R} \mid x < 0\}$ and $\mathbb{R}_{++} := \{x \in \mathbb{R} \mid x > 0\}$.
(10) Throughout, the expectation operator $\mathbf{E}_t^m$ denotes the expectation with respect to the agent's time-t information under misreporting strategy $m \in \mathcal{M}$.
(11) Note that the agent's actual time-t consumption $c_t - m_t$ is independent of Δ_t. However, Δ_t affects his future consumption path through two channels: directly through the evolution of $(m_\tau)_{\tau > t}$, and indirectly through the distribution over future transfers $(s_\tau)_{\tau > t}$.

subject to (i) [IC] and (ii) the *promise keeping* constraint $q_0 \leqslant \mathbf{E}_0^* \left[\int_0^\infty e^{-\rho t} u_t \, \mathrm{d}t\right]$, where the initial promised utility $q_0 < 0$ is a given parameter. An *optimal (full-commitment) contract* is any contract that minimizes the principal's costs subject to these two constraints. [12]

Remark 1. The requirement that $m \leqslant 0$ is imposed in W11, meaning that the agent cannot over-report his endowment. This can be motivated by (i) the agent's inability to borrow outside of the contract and (ii) the principal's ability to request that the agent deposit (some fraction of) his endowment before providing transfers at each instant (p. 1239). However, it is then incorrectly asserted in W11 that $m \leqslant 0$ implies the stronger restriction that $\Delta \leqslant 0$ (p. 1240), meaning that the agent cannot correct for past under-reports, which rules out temporary deviations from truthtelling. [13] The proof of the main verification theorem for incentive compatibility in W11 (Theorem 1, p. 1247) relies on this latter restriction. [14] Separately, the verification in W11 (on pp. 1271–72) that the purportedly optimal contract (Contract W) is IC also relies on this restriction, and is incorrect as stated without it (see Online Appendix K for details). We do not impose any restrictions (beyond admissibility) on either m or Δ in this paper, and our main conclusions would remain valid even if such restrictions were imposed (again see Online Appendix K).

2.2. The Purportedly Optimal Contract

For future reference, we now record a description of the contract derived in W11, and asserted to be optimal. Define the y-adapted process $W^y = (W_t^y)_{t \geqslant 0}$ by $\sigma W_t^y := y_t - \int_0^t (\mu - \lambda y_\tau) \, \mathrm{d}\tau$. This is the shock process that the principal would infer the agent faced if (a) the principal were to assume that the agent is truthfully reporting and (b) the agent is actually following strategy y. It coincides with the standard Brownian motion W when the agent follows the truthful strategy y^* and is a Brownian motion with drift more generally (see Online Appendix E).

(12) As usual, we implicitly restrict to the class of contracts for which the integral defining the principal's lifetime cost is well-defined.

(13) Bloedel, Krishna, and Leukhina (2020) discuss the "No Hidden Borrowing" assumption that $m \leqslant 0$ in detail. Several authors have noted W11's use of this assumption (Kapička 2013; Cisternas 2017; Battaglini and Lamba 2019), but none have made explicit note of its stronger requirement that $\Delta \leqslant 0$. To see that $m \leqslant 0$ does not imply $\Delta \leqslant 0$, note that there exist many non-positive functions with locally strictly positive derivatives.

(14) Several steps of the proof of Theorem 1 in W11 (Appendix A.2, pp. 1265-69) require that $\Delta \leqslant 0$ in order to bound various terms from above. We do not know if the bounds stated in W11's Theorem 1 itself (displays 16-18, p. 1247) remain sufficient when either Δ or m are unconstrained, but conjecture that they do not.

Definition 2.1 (Contract W)**.** The contract asserted to be optimal in W11 (henceforth Contract W) is the unique contract under which promised utility follows the geometric Brownian motion

$$q_t = q_0 \exp\left(-\frac{(k_\circ^* \sigma)^2}{2} t - k_\circ^* \sigma W_t^y\right) \tag{2.4}$$

and consumption follows the arithmetic Brownian motion

$$c_t = \bar{c}(q_0, \rho) + \frac{(k_\circ^* \sigma)^2}{2\theta} t + \frac{k_\circ^* \sigma}{\theta} W_t^y \tag{2.5}$$

where $k_\circ^*(\lambda) := \rho\theta/(\rho + \lambda)$ and $\bar{c}(q, r) := -\log(-rq)/\theta$.[15]

Contract W has several striking properties (under truthful reporting) emphasized in W11 (pp. 1253–54), and which will be important for what follows:

(i) Promised utility is a *martingale*: applying Itô's lemma to [2.4] yields $\mathrm{d}q_t = -\sigma k_\circ^* q_t \,\mathrm{d}W_t$.

(ii) The contract results in *bliss*, not immiseration: As $t \to \infty$, we have $q_t, u_t \to 0$ and $c_t \to +\infty$, and so the agent receives maximal utility and consumption in the long-run.[16]

(iii) Constant *utility delivery rate* of ρ: Formally, $u_t \equiv \rho q_t$, meaning the principal delivers utility at a constant rate ρ (and independently of all other model parameters).

(iv) Constant *geometric sensitivity*: Promised utility follows a geometric Brownian motion, and the volatility of $\log(-q_t)$ is constant.

The error in the derivation of Contract W as the optimal full-commitment contract in W11 (pp. 1252–54 and 1269–71) is based on the incorrect assertion, obtained by a numerical observation, that the optimal contract must satisfy property (iv) above. A more precise explanation of this error requires definitions not introduced until Section 4 — where we show that Contract W is suboptimal even within the class of contracts with constant geometric sensitivity — and so is deferred to Online Appendix K. Also, as noted in Remark 1, when $\lambda > 0$, the verification in W11 that Contract W is IC is incomplete without the restriction that $\Delta \leqslant 0$.

The next proposition will follow from Theorems 1, 2, and 5 below:

Proposition 2.2. Contract W is IC for all $\lambda \geqslant 0$. If $\lambda > 0$, then Contract W is strictly dominated by other IC contracts. If $\lambda = 0$, then Contract W is optimal among contracts satisfying the first-order IC condition [**FO-IC**], introduced in Section 4.

(15) In what follows, we will suppress the dependence of $k_\circ^*$ on λ when there is no cause for confusion.

(16) Intuitively, $q_t \to 0$ because $-(k_\circ^* \sigma)^2 t/2 - k_\circ^* \sigma W_t \to -\infty$ almost surely, as it is Brownian motion with negative drift, $c_t \to \infty$ because it is Brownian motion with positive drift, and $u_t \to 0$ because $u_t = -\exp(-\theta c_t)$.

3. Self-Insurance

This section contains our main analysis of self-insurance contracts. Subsection 3.1 introduces and solves a standard incomplete-market self-insurance problem for the agent, which is then used in Subsections 3.2 and 3.3 to, respectively, re-interpret Contract W and establish its strict suboptimality. Subsection 3.3 also characterizes the optimal self-insurance contract.

3.1. Agent's Self-Insurance Problem

Consider the *self-insurance problem* faced by the agent when he receives only (i) his endowment stream b and (ii) some initial asset holdings $A_0 \in \mathbb{R}$, and must self-insure by borrowing and saving in a risk-free bond market at given interest rate $r > 0$. Formally, the agent solves [17]

$$V(A_0, b_0) := \sup_{\hat{c} \in \mathscr{A}(A_0, b_0)} \mathbf{E}_0 \left[\int_0^\infty e^{-\rho t} u(\hat{c}_t) \, \mathrm{d}t \right] \tag{3.1}$$

where $\mathscr{A}(A_0, b_0)$ is the set of (A_0, b_0)-*feasible* consumption processes $\hat{c} = (\hat{c}_t)_{t \geqslant 0}$, which are those that are b-adapted and induce an *asset process* $A^{\hat{c}} = \left(A_t^{\hat{c}}\right)_{t \geqslant 0}$ that solves

$$\mathrm{d}A_t^{\hat{c}} = (r A_t^{\hat{c}} + b_t - \hat{c}_t) \, \mathrm{d}t \tag{3.2}$$

and satisfies the *no-Ponzi* condition

$$\lim_{t \to \infty} e^{-rt} A_t^{\hat{c}} \geqslant 0 \tag{3.3}$$

$\mathbf{P}$-almost surely. A consumption process $\hat{c}$ is said to *solve the agent's self-insurance problem* if it attains the supremum in [3.1].

Proposition 3.1. The process $\hat{c}^*$ defined by

$$\hat{c}_t^* := \hat{C}(A_0, b_0) + \left(\frac{r - \rho + \sigma^2 f(r; \lambda)^2/2}{\theta} \right) t + \frac{\sigma f(r; \lambda)}{\theta} W_t \tag{3.4}$$

solves the agent's self-insurance problem, where $f(r; \lambda) := r\theta/(r + \lambda)$ and

$$\hat{C}(A, b) := rA + \frac{r}{r + \lambda} b + \frac{\mu}{\lambda + r} - \left[\frac{r - \rho}{r\theta} + \frac{1}{2} \frac{\sigma^2 f(r; \lambda)^2}{r\theta} \right] \tag{3.5}$$

(17) When analyzing the agent's self-insurance problem and the class of self-insurance contracts, we let $\mathbf{P}$ denote the probability measure over paths of b induced by the SDE [2.1] and let $\mathbf{E}_t$ denote the associated conditional expectation operators. We use this notation — instead of the truthful reporting measure $\mathbf{P}^*$ from Section 2.1 — to emphasize that there is no communication in the self-insurance setting.

The proof of Proposition 3.1 is in Online Appendix D. The consumption process [3.4] is the natural continuous-time limit of Caballero's (1990) and Wang's (2003) solution to the analogous discrete-time self-insurance problem, though to our knowledge the continuous-time derivation is new.[18] Notice that the solution described in [3.4] and [3.5] satisfies $\hat{c}_t^* \equiv \hat{C}\left(A_t^*, b_t\right)$ where $A_t^* := A_t^{\hat{c}^*}$ (see Lemma D.3 in Online Appendix D). Thus, at time t the agent consumes a multiple r of his *permanent income*

[3.6] $$A_t^* + \mathbf{E}_t\left[\int_t^\infty e^{-r(\tau-t)} b_\tau \, \mathrm{d}\tau\right] = A_t^* + \frac{1}{r}\left[\frac{r}{r+\lambda} b_t + \frac{\mu}{\lambda+r}\right]$$

adjusted by two constant terms. To interpret those terms, consider the agent's *savings* at time t, which are given by

[3.7] $$\frac{\mathrm{d}A_t^*}{\mathrm{d}t} = \underbrace{\frac{\lambda b_t - \mu}{r+\lambda}}_{\text{“rainy day”}} + \underbrace{\frac{r-\rho}{r\theta}}_{\text{impatience}} + \underbrace{\frac{\sigma^2 f(r;\lambda)^2/2}{r\theta}}_{\text{precautionary}}$$

The agent's savings behavior is determined by three channels: (i) his savings "for a rainy day" (the first term in [3.7]) whereby he saves if b_t exceeds its long-run average μ/λ and dis-saves otherwise, (ii) his impatience relative to the return on saving (the second term), and (iii) his *precautionary savings* (the third term). Thus, the constants shifting the agent's consumption process relative to the fraction r of his permanent income correspond to his (dis)savings due to impatience and his precautionary savings (cf. Wang 2003, p. 1464).

To understand the agent's *risk exposure* — i.e., the sensitivity $f(r;\lambda)/\theta = r/(r+\lambda)$ of consumption to endowment shocks — observe that the derivative of the agent's permanent income [3.6] with respect to b_t is $1/(r+\lambda)$. Due to the absence of wealth effects under exponential utility, the agent optimally responds to a marginally higher endowment by permanently shifting up his future consumption, i.e., increasing each c_τ for $\tau > t$ by the same constant. Setting this constant to $r/(r+\lambda)$ increases the expected present value of consumption by precisely $1/(r+\lambda)$, exactly matching the increase in permanent income. It is useful to note that $f(\cdot;\lambda)$ is strictly increasing when $\lambda > 0$ and constant when $\lambda = 0$, so that the sensitivity of consumption is (at least weakly) increasing in the shadow rate. Moreover, the agent's risk exposure is an increasing transformation of his precautionary savings.

(18) Caballero (1990) considers a general class of discrete-time ARMA endowment processes. Wang (2003) specializes to a discrete-time AR(1) endowment process which, when shocks are Gaussian, is precisely the discrete-time analogue of the OU process [2.1].

Going forward, it will also be useful to observe that the solution in Proposition 3.1 is the unique feasible consumption process satisfying the agent's familiar *Euler equation*

$$e^{(r-\rho)t}u'(\hat{c}_t^*) = \mathbf{E}_t\left[e^{(r-\rho)\tau}u'(\hat{c}_\tau^*)\right] \qquad \text{for all } \tau > t \tag{3.8}$$

which specifies that the agent's discounted marginal utility defines a martingale. Under exponential utility, then, discounted flow utility $e^{(r-\rho)t}u(\hat{c}_t^*)$ is a martingale (Lemma D.7). In addition, at the optimum the agent's continuation value process $V = (V_t)_{t\geqslant 0}$ defined by $V_t := V(A_t^*, b_t)$ satisfies

$$\frac{u(\hat{c}_t^*)}{r} = V_t \tag{3.9}$$

$$= V_0 \exp\left[-\left(r - \rho + \frac{\sigma^2 f(r;\lambda)^2}{2}\right)t - f(r;\lambda)\sigma W_t\right] \tag{3.10}$$

By [3.9] (Corollary D.6), the solution to the agent's self-insurance problem induces *constant utility delivery rate* of r, and by [3.10] (from [3.9] and [3.4]), his continuation value exhibits *constant geometric sensitivity* (in the sense of properties (iii)-(iv) from Subsection 2.2). See Online Appendix D for derivations of [3.8]–[3.9].

3.2. Implementing Contract W

By comparing [3.4] to [2.5] and [3.10] to [2.4], we see that Contract W induces the same consumption and continuation value process for the agent as the solution to his self-insurance problem (with suitably chosen A_0) at $r = \rho$: simply identify q_t with V_t and observe that $f(\rho;\lambda) = k_\circ^*$. Consequently, Contract W is outcome-equivalent to allowing the agent to self-insure at rate ρ. This should not be surprising, for properties (i) and (iii) of Contract W, together with the assumption of exponential utility, immediately imply that the agent's Euler equation [3.8] (with $r = \rho$) is satisfied under Contract W.[19]

This immediately suggests the following indirect implementation of Contract W. Recall that the principal has access to a bond market where he can borrow and lend at the rate ρ. On the other hand, the agent can only transact with the principal. Thus, let the principal act as the agent's bank: she gives him initial assets A_0 and commits to offering him the market rate ρ on his savings (by trading in the bond market on his behalf). The agent is free to choose any consumption-savings strategy he wishes. Since b_0 is common

(19) It is not observed in W11 that the agent's Euler equation holds under Contract W. However, it is observed (on p. 1262) that the Euler equation is satisfied under the optimal contract in the taste shock example from that paper's Section 8, which has a similar structure to that of Contract W in the $\lambda = 0$ case.

knowledge, the agent need not communicate any information to the principal. It follows from the above observations that this implementation induces the same consumption process as Contract W. By the Revelation Principle, this also provides a complete proof that Contract W is IC (cf. Remark 1).

It remains to specify initial conditions and compute the principal's cost. Let b_0 be given and suppose that the principal must deliver q_0 lifetime utiles to the agent. From Proposition 3.1 and [3.9]–[3.10], it is straightforward to verify that she must provide $A_0(b_0, q_0, \rho)$ initial assets to the agent, where

$$A_0(b, q, r) := \frac{\overline{c}(q, r)}{r} - \frac{\mu}{r(r+\lambda)} - \frac{b}{r+\lambda} + \frac{r-\rho}{\theta r^2} + \frac{\sigma^2}{2}\frac{f(r;\lambda)^2}{\theta r^2} \tag{3.11}$$

Note that the first three terms in [3.11] constitute the initial wealth needed for the agent to achieve lifetime utility q in his self-insurance problem if his endowment were not subject to shocks (i.e., if $\sigma = 0$ in [2.1]). The penultimate term, which is proportional to the (dis)savings due to impatience term in [3.7] (and which vanishes under Contract W) compensates the agent for his impatience. The final term, which is proportional to the precautionary savings in [3.7], is the additional wealth needed to compensate the agent for bearing the uninsurable risk in his endowment while still achieving the same lifetime utility. On the other hand, the principal's cost (in the sense of [2.3]) of Contract W in W11 may be written as a function of the agent's report and promised utility as

$$J^W(q, y) := \underbrace{\left[\frac{\overline{c}(q, \rho)}{\rho} - \frac{\mu}{\rho(\rho+\lambda)} - \frac{y}{\rho+\lambda}\right]}_{=:J^*(q,y)} + \underbrace{\frac{\sigma^2\theta}{2(\rho+\lambda)^2}}_{=\sigma^2 f(\rho;\lambda)^2/(2\theta\rho^2)} \tag{3.12}$$

where $J^*(q, y)$ is the first-best value function that arises from complete risk-sharing under symmetric information.[20] W11 interprets the last term as the additional cost of information rents needed to induce [IC] when the agent's endowment is private information. Clearly, $J^W(q_0, b_0) = A(b_0, q_0, \rho)$, so that the indirect implementation also generates the same lifetime cost for the principal as the direct mechanism, with precautionary savings costs in the former corresponding to information rent costs in the latter.

Implications for Long-Run Properties. This implementation helps to explain why Contract W leads to long-run bliss, rather than immiseration as in the classic literature. As

(20) The first-best allocation perfectly stabilizes the agent's consumption at the level $\overline{c}(q_0, \rho)$ which delivers promised utility q_0. Thus, under symmetric information, the optimal contract sets the drift and volatility of both promised utility and consumption to zero. (Zero volatility corresponds to perfect risk-sharing. Zero drift follows from the fact that the principal's and agent's discount rates are identical by assumption.) See Lemma A.1 in Appendix A.1 for a derivation of $J^*(q, y)$.

discussed further in Subsection 5.2.1 below, immiseration arises from the principal's optimal cost-smoothing under the optimal full-commitment contract, which formally means that her marginal cost of providing promised utility defines a martingale. However, under Contract W this marginal cost process $J_q^W(b_t, q_t)$ defines a positive *sub*martingale that is unbounded above, so that the Martingale Convergence Theorem has no implications for its long-run properties. Instead, the dynamics of Contract W are driven by a different martingale, namely, the agent's discounted marginal utility (which defines a martingale by the Euler equation [**3.8**]). In particular, it is well-known that when the agent self-insures at rate $r \geqslant \rho$, convergence properties of this discounted marginal utility martingale imply quite generally that $A_t^*, \hat{c}_t^* \to +\infty$ as $t \to \infty$, which intuitively arises from the agent's precautionary savings (Sotomayor 1984; Chamberlain and Wilson 2000; Ljunqvist and Sargent 2000, Ch. 17). Notably, the same asymptotic behavior also emerges in the discrete-time version of the W11 model, in which the analogue to Contract W exactly coincides with the discrete-time version of our self-insurance solution (Proposition 3.1) obtained by Caballero (1990) and Wang (2003).

3.3. Self-Insurance Contracts

In the implementation of Contract W described above, the principal acts as the agent's bank and allows the agent to borrow and save at the market rate ρ. Of course, it is equally feasible for the principal to offer the agent an effective rate of $r \neq \rho$ by taxing ($r < \rho$) or subsidizing ($r > \rho$) his capital gains.

Definition 3.2. A *Self-Insurance Contract* (b_0, q_0, r) that delivers lifetime utility $q_0 < 0$ with interest rate $r > 0$, and where time-0 income is b_0, consists of the following steps:

(1) The principal gives the agent initial assets $A_0(b_0, q_0, r)$, as defined in [**3.11**], and
(2) The principal, acting as the agent's bank, allows the agent to freely borrow and save at risk-free *shadow rate* r by imposing a constant marginal tax (or subsidy) $\tau(r) := 1 - r/\rho$ on his capital gains ρA_t, thereby collecting tax revenue $\tau(r)\rho A_t$, at each instant t.

The principal's expected lifetime cost of this implementation is $\Pi(b_0, q_0, r) := A_0 - \mathbf{E}_0\left[\int_0^\infty e^{-\rho t}\rho\tau(r)A_t \mathrm{d}t\right]$.

Clearly, Contract W corresponds to the particular Self-Insurance Contract in which the principal imposes a zero tax. Moreover, because Proposition 3.1 characterizes the agent's self-insurance solution for *any* $r > 0$, the class of Self-Insurance Contracts inherit some of Contract W's properties, such as constant utility delivery rate and constant geometric sensitivity (recall [3.10]–[3.9]). In particular, an application of Itô's lemma to [3.10] yields the SDE

$$\mathrm{d}V_t = (\rho - r)V_t\mathrm{d}t - V_t f(r;\lambda)\sigma \mathrm{d}W_t \tag{3.13}$$

for the agent's continuation value. Identifying V_t with q_t, we see that when $r = \rho$, [3.13] coincides with the SDE for promised utility [2.4] under Contract W, as must be the case. The principal's cost of any Self-Insurance Contract (b_0, q_0, r) also admits a simple closed-form expression:

Lemma 3.3. The principal's cost $\Pi(b_0, q_0, r)$ satisfies

$$\begin{aligned}\Pi(b_0, q_0, r) &= \frac{\log(\rho/r)}{\rho\theta} + \frac{r-\rho}{\theta\rho^2} + J^*(q_0, b_0) + \underbrace{\frac{\sigma^2\theta r^2}{2\rho^2(r+\lambda)^2}}_{=\sigma^2 f(r;\lambda)^2/(2\theta\rho^2)} \\ &= \mathbf{E}_0\left[\int_0^\infty e^{-\rho t}\left(\hat{c}^*_t - b_t\right)\,\mathrm{d}t\right]\end{aligned} \tag{3.14}$$

where $\hat{c}^*$ is as in [3.4] in Proposition 3.1.

The proof of Lemma 3.3 is in Appendix A.2. Notice that [3.14] shows that the principal's cost of a Self-Insurance Contract, defined in terms of a wealth transfer and tax revenue, is actually equivalent to the expected resource cost of the contract, as defined in [2.3] in the context of the principal's full-commitment problem.

Remark 2. The expression in [3.14] is the principal's lifetime cost at time $t = 0$, but his continuation cost of a Self-Insurance Contract at times $t > 0$ will generally take a different form. This is because the self-insurance implementation features a lump-sum transfer at $t = 0$, rather than flow transfers as described in Subsection 2.1. While this is arguably the economically simplest indirect implementation, there are outcome-equivalent implementations (without communication) in which the lump-sum transfer of A_0 is replaced with a deterministic flow transfer process $\hat{s}$ with $\mathrm{d}\hat{s}_t := (\alpha - \lambda\hat{s}_t)\,\mathrm{d}t$ and $\hat{s}_0, \alpha$ chosen appropriately. Because this transfer process is Markovian, the principal's continuation cost at any time $t \geqslant 0$ will be given by $\Pi(V_t, b_t, r)$ under this alternative implementation.

Direct-Revelation Formulation. Any Self-Insurance Contract (b_0, q_0, r) can be reformulated as a direct-revelation contract as follows: The principal constructs a y-adapted *virtual asset* process $A^v = (A^v_t)_{t\geqslant 0}$ with initial condition $A^v_0 := A_0(b_0, q_0, r)$ and law of motion

$$\text{[3.15]} \qquad \mathrm{d}A^v_t = \left[\frac{r - \rho + \sigma^2 f(r;\lambda)^2/2}{r\theta} + \frac{\lambda y_t - \mu}{r + \lambda}\right] \mathrm{d}t$$

and recommends that the agent consume according to $c_t := \hat{C}(A^v_t, y_t)$; the agent's actual consumption is $c_t - m_t$. In other words, the principal "saves" on the agent's behalf and recommends that he consume as would be optimal in his self-insurance problem *assuming that y were his true endowment process*. The virtual assets are not a physical object, but simply a state variable that the principal uses to track what the agent would have done in his self-insurance problem. Call the above mechanisms *Direct-Revelation Self-Insurance Contracts*. (See Online Appendix F for a more formal definition of these mechanisms and proofs of statements below.)

By comparing [3.15] to [3.7], we see that the agent *under-reporting* in the direct mechanism corresponds to him *under-saving* (and thus *over-consuming*) relative to his optimal strategy in the indirect mechanism (as described in Proposition 3.1). This immediately suggests the following result, which is effectively a consequence of the Revelation Principle:

Theorem 1. *Every Direct-Revelation Self-Insurance Contract is IC.*

The formal proof of Theorem 1 requires showing that every reporting strategy feasible for the agent in the direct mechanism induces a virtual asset process A^v and (actual) consumption process $c - m$ that would have been feasible for the agent in the indirect mechanism (i.e., A^v must satisfy the no-Ponzi condition [3.3] under any feasible reporting strategy). We show that this property is guaranteed by admissibility of the agent's reports. However, it can be shown that the agent's feasible set of consumption-savings strategies is actually *strictly* larger in the indirect mechanism. For instance, the agent is permitted to choose consumption processes $\hat{c}$ with jumps in the self-insurance problem, which correspond to misreporting strategies m with jumps in the direct mechanism, and these are ruled out by admissibility. Consequently, Direct-Revelation Self-Insurance Contracts remain IC even when the agent's space of feasible reporting strategies is enlarged.

Optimal Self-Insurance Contract. We now describe the *optimal* Self-Insurance Contract. We immediately see from [3.14] that the optimal shadow rate r^* satisfies the first-order condition

$$\text{[3.16]} \qquad \frac{\mathrm{d}}{\mathrm{d}r}\left[-\log(r^*) + \frac{r^*}{\rho}\right] + \underbrace{\frac{\mathrm{d}}{\mathrm{d}r}\left[\frac{\sigma^2 f^2(r^*;\lambda)}{2\rho}\right]}_{>0 \text{ if } \lambda>0,\ =0 \text{ if } \lambda=0} = 0$$

and is therefore independent of q_0, b_0, μ but dependent on all other parameters. When $\lambda > 0$, the optimal Self-Insurance Contract entails $r^* < \rho$ and thus strictly dominates Contract W, establishing the first part of Proposition 2.2. Intuitively, there are two ways to "stabilize" the agent's continuation utility: (a) induce zero drift (i.e., eliminate his (dis)savings due to impatience) and (b) reduce its sensitivity to endowment shocks (i.e., reduce his risk exposure). By [**3.13**], the principal necessarily faces a tradeoff: increasing $r \in (0, \rho)$ increases both (a) the drift $V_t(\rho - r) < 0$ and (b) the sensitivity $-V_t f(r; \lambda)\sigma > 0$. The first term in [**3.16**] represents the principal's marginal *value* of (a), while the second term represents her marginal *cost* of (b). At the optimum, she equates these marginal effects by setting a strictly positive tax, thereby inducing the strictly positive "intertemporal wedge" familiar from optimal dynamic taxation (e.g., Golosov, Kocherlakota, and Tsyvinski 2003). It is optimal to set $r^* \leqslant \rho$ because, when $r \geqslant \rho$ the drift is already positive, so that increasing r further destabilizes the agent's continuation utility in both dimensions.

By contrast, Contract W resolves this tradeoff entirely in favor of channel (a), which is generally suboptimal. Indeed, Contract W is the optimal Self-Insurance Contract only when $\lambda = 0$. In that case, the agent cannot effectively self-insure, no matter the shadow rate, as can be seen by noting that his savings [**3.7**] are deterministic when shocks are permanent. The principal consequently faces no tradeoff between channels (a) and (b), and so simply sets the drift in [**3.13**] to zero.

The following theorem formally records these observations and highlights some other important features of the solution.

Theorem 2. *Given any b_0, q_0, there exists an optimal Self-Insurance Contract. Any such contract is characterized by $r^* \in (0, \rho]$, where r^* is a minimizer of $\Pi(b_0, q_0, \cdot)$ (as defined in [**3.14**]).* [21] *It is independent of (b_0, q_0, μ) but depends on $(\lambda, \sigma, \rho, \theta)$. As a function of (λ, σ):*

(i) $r^*(\lambda, \cdot) < \rho$ *for all* $\lambda > 0$ *and* $r^*(0, \cdot) \equiv \rho$.

(ii) For each $\sigma > 0$, $\lambda \mapsto k^*(\lambda, \sigma) := f(r^*(\lambda; \sigma); \lambda)$ *is strictly decreasing, while* $\lambda \mapsto r^*(\lambda; \sigma)$ *is non-monotone:* $\lim_{\lambda \to 0} r^*(\lambda, \sigma) = \lim_{\lambda \to \infty} r^*(\lambda, \sigma) = \rho$.

(iii) For each $\lambda > 0$, *both* $\sigma \mapsto k^*(\lambda, \sigma)$ *and* $\sigma \mapsto r^*(\lambda; \sigma)$ *are strictly decreasing, with* $\lim_{\sigma \to 0} k^*(\lambda, \sigma) = k^*_\circ(\lambda)$ *and* $\lim_{\sigma \to \infty} k^*(\lambda, \sigma) = 0$.

(21) Going forward, we slightly abuse terminology by referring to *the* optimal Self-Insurance Contract. The optimum is indeed unique in many cases. When $\lambda = 0$, it can be shown that $\Pi(b_0, q_0, \cdot)$ is strictly convex, so that the optimal Self-Insurance Contract is unique. When $\lambda > 0$, [**3.16**] has at most two local minima when $3\lambda < \rho < \lambda + \sigma^2\theta^2/3$, and a unique local minimum in all other cases. We state without proof that even in the case of two local minima, there is a unique global minimizer for generic values of the parameters $(\lambda, \sigma, \theta, \rho)$.

The proof of Theorem 2 is in Online Appendix G. The intuition for point (i) was given above. To understand the monotonicity of $k^*(\cdot;\sigma)$ in point (ii), note that $f(r;\cdot)$ is strictly decreasing for any fixed r; whenever $r^*(\cdot;\sigma)$ is locally increasing, it does so slowly enough that the sensitivity decreases. The non-monotonicity of $r^*(\cdot,\sigma)$ corresponds to the principal setting (asymptotically) zero taxes when the endowment is either nearly permanent or perfectly transient, and setting a maximal tax rate when persistence is intermediate. As $\lambda \to \infty$, the agent's utility responds little to endowment shocks no matter the shadow rate, so the principal is primarily concerned with setting the drift of utility to zero; clearly, first-best insurance is achieved in this limit. The intuition at $\lambda = 0$ is similar: although the principal would find it valuable to reduce the sensitivity $f(r;\lambda)$, she cannot do so for reasons described above. Finally, point (iii) follows from the fact that reducing the sensitivity of continuation utility becomes relatively more valuable as the size of shocks increases.

Long-Run Properties. We now describe the long-run properties of the optimal Self-Insurance Contract. Recall from Subsection 2.2 that a contract leads to *bliss* if $u_t, q_t \to 0$ and $\hat{c}_t^* \to +\infty$ $\mathbf{P}^*$-almost surely. Analogously, say that a contract leads *immiseration* (or that the agent converges to *misery*) if $u_t, q_t, \hat{c}_t^* \to -\infty$ $\mathbf{P}^*$-almost surely.[22]

Theorem 3. *Under the optimal Self-Insurance Contract:*

(i) The agent's continuation value has a strictly negative drift if $\lambda > 0$, and zero drift if $\lambda = 0$.

(ii) For each fixed $\sigma > 0$, there exists $\overline{\lambda}(\sigma) > 0$ such that the agent converges to misery if $\lambda > \overline{\lambda}(\sigma)$ and to bliss if $\lambda \in \left[0, \overline{\lambda}(\sigma)\right)$.

(iii) For each fixed $\lambda > 0$, there exists $\overline{\sigma}(\lambda) \geqslant 0$ such that the agent converges to misery if $\sigma > \overline{\sigma}(\lambda)$ and to bliss if $\sigma \in (0, \overline{\sigma}(\lambda))$. Moreover, $\overline{\sigma}(\lambda) = 0$ if and only if $\lambda \geqslant \rho$.

The proof of Theorem 3 is in Appendix B. Intuitively, each point of Theorem 3 follows from the corresponding point of Theorem 2. In short, the agent converges to misery when the principal finds it optimal to set a large tax, in which case the agent's dissavings due to impatience outweigh his precautionary motive, leading to a front-loaded consumption profile. By contrast, the agent converges to bliss when the tax is sufficiently small, in which case the agent's precautionary savings motive dominates, leading to back-loaded consumption as described in Subsection 3.2.

Interestingly, the optimal long-run outcome depends sensitively on the properties of the endowment process, leading to bliss when the Brownian shocks have a large

(22) Lemma B.2 in Appendix B shows that, when $\lambda > 0$, the agent's assets have the same asymptotic properties as his consumption.

instantaneous ($\sigma \to \infty$) or very long-lived ($\lambda \to 0$) effect on the agent's endowment, and leading to misery when these effects are small. Point (iii) reveals that, even when the endowment is nearly deterministic ($\sigma \to 0$), the optimal long-run outcome depends on the persistence of shocks. This sensitivity contrasts with both Contract W, which always leads to bliss, and optimal full-commitment contracts in approximating discrete-time models, which lead to misery for essentially arbitrary endowment processes (see Subsection 5.2.1). Note that although the negative drift in point (i) is reminiscent of immiseration, it is not sufficient: because continuation utility follows the *geometric* Brownian [**3.13**], its drift must be *sufficiently negative* for $V_t \to -\infty$ $\mathbf{P}^*$-almost surely. Indeed, immiseration arises only if the drift of the agent's consumption (which follows the *arithmetic* Brownian motion [**3.4**]) is strictly negative, while bliss occurs only if this drift is positive. Economically, as described above, the drift of consumption is determined by the relative size of the agent's dissavings due to impatience and his precautionary savings.

3.4. The Road Ahead

The above analysis raises two further questions, which we answer in the remainder of the paper:

(i) *For $\lambda > 0$, is there a sense in which Contract W is in fact optimal?* Building on its implementation described in Subsection 3.2, we show in Subsection 5.1 that it is optimal in an alternative model wherein the agent can covertly save and borrow at the same rate as the principal.

(ii) *When $\lambda = 0$, what explains the peculiar properties of Contract W, which is both the optimal Self-Insurance Contract and the optimal full-commitment contract (per Proposition 2.2)?* In Subsection 5.2, we show that its properties derive from two closely related facts: (a) for incentive reasons, the principal is unable to insure the agent and therefore cannot manipulate his risk exposure, and (b) the agent is necessarily indifferent among all reporting strategies.

To formally develop these ideas, we must first introduce a richer set of direct-revelation contracts. We turn to this task in the next section.

4. Direct-Revelation Contracts

Subsections 4.1 introduces the first-order approach from W11 (pp. 1243-46 and 1248-49), though our presentation is somewhat different from that paper's. Subsection 4.2 uses the first-order approach to provide equivalent characterizations of, and foundations for, the class of Self-Insurance Contract.

4.1. First-Order Incentive Compatibility

It is well known that the agent's incentive constraint [**IC**] is difficult to analyze directly. The *first-order approach (FOA)* in W11 considers only the agent's incentives to make "small" deviations starting from histories where he had been truthful in the past, so that setting $\Delta_t = 0$ is optimal at histories where $m_\tau = 0$ for all $\tau \leqslant t$.

The FOA requires tracking two statistics of the contract. First, as is well understood, we must keep track of the agent's promised utility q (recall [**2.2**]). Second, because the agent's information is persistent, the approach in W11 proposes tracking (the negative of) the agent's *marginal promised utility* $p = (p_t)_{t \geqslant 0}$, defined as

[**4.1**] $$p_t := -\mathbf{E}_t^* \left[\int_t^\infty e^{-(\rho+\lambda)(\tau - t)} \underbrace{u'(c_\tau)}_{= -\theta u_\tau} \, \mathrm{d}\tau \right]$$

The quantity $-p_t$ represents the marginal change in q_t from a marginal increase in b_t, assuming the agent truthfully reports going forward. Thus, $-p_t$ is the agent's *information rent* from a small, instantaneous misreport. As is intuitive, the agent's information rents are larger when his endowment is more persistent (ie, when λ is smaller), and hence his informational advantage after a deviation is longer-lived.

Under any contract delivering lifetime utility q_0 and truthful reporting by the agent, by a Martingale Representation Theorem these processes satisfy the SDEs

[**4.2**] $$\mathrm{d}q_t = [\rho q_t - u_t] \, \mathrm{d}t + \gamma_t \sigma \, \mathrm{d}W_t^y$$

[**4.3**] $$\mathrm{d}p_t = [\rho p_t - \lambda \gamma_t - \theta u_t] \, \mathrm{d}t + Q_t \sigma \, \mathrm{d}W_t^y$$

where W^y is as defined in Subsection 2.2, the initial condition $q_0 < 0$ is given, and q and p satisfy the terminal conditions $\lim_{t\to\infty} e^{-\rho t} q_t = \lim_{t \to \infty} e^{-\rho t} p_t = 0$ $\mathbf{P}^*$-almost surely and the transversality conditions $\lim_{t\to\infty} \mathbf{E}^* \left[e^{-\rho t} q_t \right] = \lim_{t\to\infty} \mathbf{E}^* \left[e^{-\rho t} p_t \right] = 0$. The principal's choice of contract determines the initial $p_0 < 0$ and the y-adapted *sensitivity processes* $\gamma = (\gamma_t)_{t \geqslant 0}$ and $Q = (Q_t)_{t \geqslant 0}$.

Incentive compatibility imposes restrictions on the sensitivity processes. Following the same steps as in W11 (pp. 1243–44), we may heuristically derive the *first-order IC* condition

[**FO-IC**] $$\gamma_t + p_t = 0,$$

which is a first-order condition for the agent to find setting $\Delta_t = 0$ optimal at truthful histories where $m_t = 0$.[23] Note that [**FO-IC**] does not rule out global deviations in which the agent sets $\Delta_t \neq 0$ also at non-truthful histories where $m_t \neq 0$. In our analysis of direct revelation contracts, we restrict attention to contracts satisfying [**FO-IC**].

Definition 4.1. Any contract that satisfies [**FO-IC**] is said to be *first-order IC* (FO-IC).

Remark 3. We do not assert that all IC contracts are necessarily FO-IC; though we conjecture that this is true, without additional argument there may in principle exist IC contracts that strictly dominate all FO-IC contracts. While by definition [**IC**] rules out the profitability of "local" deviations from truthtelling, the agent's value function must be suitably differentiable in order for this to imply the first-order (or envelope) condition [**FO-IC**]. It is claimed in W11 that [**FO-IC**] is implied by [**IC**] by appealing to the stochastic maximum principle (pp. 1243-44, 1264), but we are not aware of versions of the maximum principle that directly apply to agent's reporting problem in the present model because the agent's utility function is unbounded below and does not satisfy growth conditions found in the literature.[24] By focusing on contracts that are both FO-IC and IC, we are implicitly restricting attention to those contracts that are sufficiently regular for [**IC**] to imply [**FO-IC**].

4.2. Self-Insurance Contracts: Equivalent Formulations

In this subsection, we present equivalent ways of representing Self-Insurance Contracts as FO-IC contracts. We will show that this class is equivalent to (a) the class of Stationary Contracts, which are uniquely tractable and closely related to the analysis in W11, and (b) the class of Proportional-Utility Contracts, which Strulovici (2020) derives from an axiomatic notion of renegotiation-proofness.

Definition 4.2. An FO-IC contract is a:

(i) *Stationary Contract* if its *geometric sensitivity* $k = (k_t)$ defined by $k_t := p_t/q_t$ is a constant process.

(23) W11 presents [**FO-IC**] as an inequality (display 10, p. 1244), but this is due to that paper's assumption that $\Delta \leqslant 0$ (recall Remark 1). As presented here, [**FO-IC**] is precisely the continuous-time analogue of Pavan, Segal, and Toikka's (2014) "IC-FOC" envelope condition (see also Kapička 2013; Bergemann and Strack 2015).

(24) Likewise, sufficient conditions found in the literature for the applicability of variational arguments (cf. Cvitanić and Zhang 2012; Sannikov 2014; Prat and Jovanovic 2014; He et al. 2017) or the envelope theorem (cf. Kapička 2013; Pavan, Segal, and Toikka 2014; Bergemann and Strack 2015) rely on boundedness or growth assumptions on the agent's utility function that are not satisfied here. We emphasize that this has little to do with the present continuous-time formulation, and that similar technical issues would arise even in the discrete-time version of the model.

(ii) *Proportional-Utility Contract* if its *utility delivery rate* $\beta = (\beta_t)_{t \geqslant 0}$ defined by $\beta_t := u_t/q_t$ is a constant process.

For a Stationary Contract, the constant value of geometric sensitivity will be denoted by $k_\circ$. For a Proportional Utility Contract, the constant value of utility delivery rate will be denoted $\beta_\circ$. Notice that under a Proportional Utility Contract, the agent's recommended consumption may be written as $c_t = \overline{c}(q_t, \beta_\circ) = -\log(-\beta_\circ q_t)/\theta$.

Proposition 4.3. If $\lambda > 0$, then the following direct-revelation contracts are outcome-equivalent:[25]

(i) Direct-Revelation Self-Insurance Contract with shadow rate r.

(ii) Proportional Utility Contract with $\beta_\circ = r$.

(iii) Stationary Contract with $k_\circ = f(r; \lambda)$.

If $\lambda = 0$, then (i) and (ii) remain outcome-equivalent but are strictly more restrictive than (iii). Consequently, given any $\lambda \geqslant 0$, every Direct-Revelation Self-Insurance Contract is FO-IC and every Proportional Utility Contract is IC; given $\lambda > 0$, every Stationary Contract is IC.

The proof of Proposition 4.3 is in Appendix C. An important implication of Proposition 4.3 is that the analysis of Section 3 could have been equivalently carried out in terms of Stationary (or Proportional-Utility) Contracts. In particular, when $\lambda > 0$ Stationary Contracts constitute the largest class of FO-IC contracts that satisfy two defining properties of Contract W: constant utility delivery rate and constant geometric sensitivity (recall properties (iii)-(iv) in Subsection 2.2). Thus, it is possible to derive the strict suboptimality of Contract W by optimizing over Stationary Contracts. This would have the advantage of hewing more closely to the method of analysis in W11. However, our self-insurance approach is useful because it generates a richer set of intuitions and a simple proof that Stationary Contracts are IC.

It is also worth noting that this characterization of Stationary Contracts indicates that they are uniquely tractable among all FO-IC contracts, as they allow us to (a) dispense with p as a separate state variable and (b) reduce the principal's problem to a one-dimensional minimization problem (over $\beta_\circ$ or $k_\circ$). Indeed, the necessity of tracking both q and p is

(25) As described in Remark 2, our definition of a Self-Insurance Contract requires the principal makes a lump-sum transfer at $t = 0$ instead of making flow transfers at all $t \geqslant 0$ as in Section 2 and Subsection 4.1. Thus, per Remark 2, the "outcome equivalence" here is in terms of the agent's induced consumption process and the principal's cost at $t = 0$. To ensure that the principal's continuation cost coincides for all $t \geqslant 0$, one may simply revise the definition of Self-Insurance Contract as described in that remark.

known to be a major obstacle to analyzing full-commitment optima in contracting models with persistent states. This has led many authors to restrict attention to environments in which p can be dispensed with. [26]

Proposition 4.3 also reveals a connection to the class of renegotiation-proof contracts introduced by Strulovici (2020) (henceforth S20). Intuitively, a contract is renegotiation-proof if, at each report history $y^t = (y_\tau)_{\tau \in [0,t]}$, the principal cannot replace the continuation of that contract with a "challenger" contract that Pareto dominates it. To formalize this idea, one must define which contracts are valid challengers. In settings with Markovian private information, the state (y_t, q_t) captures all payoff-relevant aspects of the history. Based on this observation, S20 proposes that a renegotiation-proof contract should be *state-consistent*: the valid challengers to the contract's y^t-continuation are "suitable transformations" of that same contract's $\tilde{y}^{\tilde{t}}$-continuations, where the transformation is determined by the states (y_t, q_t) and $(\tilde{y}_{\tilde{t}}, \tilde{q}_{\tilde{t}})$ that those histories induce.

In the context of the present model, S20 shows that state-consistent contracts are not only Markovian in the state (y, q), but in fact are equivalent to our Proportional Utility Contracts. [27] In conjunction with Proposition 4.3, this establishes a three-way equivalence between Self-Insurance, Proportional-Utility, and state-consistent contracts (and, when $\lambda > 0$, a fourth equivalence to Stationary Contracts). This equivalence provides an alternative justification for our analysis based on the principal's limited commitment; conversely, it connects the axiomatic definition of state-consistency to a simple class of self-insurance problems with history-independent taxes. From a methodological standpoint, S20 establishes that state-consistent contracts are IC by solving a relaxation of the agent's reporting problem in which the agent is allowed to report jumps in his endowment (which violates admissibility). [28] As described in Subsection 3.3 above (and elaborated in

(26) For instance, Prat and Jovanovic (2014) and Williams (2015) study moral hazard models with hidden states and restrict to CARA utility, under which all FO-IC contracts have constant geometric volatility (see also He et al. (2017) for a similar reduction). In the present setting, this stronger property only holds when $\lambda = 0$, while stationary contracts preserve this structure for all $\lambda \geqslant 0$. Chen (2017) studies a hidden-information model in which the agent has state-independent preferences, which implies that $p \equiv 0$ under any FO-IC contract. Thus, that paper's principal conditions only on y and q, the latter of which has zero volatility.

(27) That paper derives the cost function [**3.14**] and the constancy of the utility delivery rate directly from the axiomatic definition of state-consistency, which has no parallel here. Conversely, our approach based on self-insurance, which can be viewed as providing an indirect implementation for state-consistent contracts in the pure hidden information case, has no parallel in S20, which also does not analyze the long-run properties of the optimal state-consistent contract or study optimal full-commitment contracts (as we do in Section 5 below).

(28) The agent's reporting problem is linear in Δ, meaning that he is either indifferent between all Δ (as is

Online Appendix F), our self-insurance approach reveals that report jumps in the direct mechanism correspond to consumption jumps in the agent's self-insurance problem.

Remark 4. Following calculations in W11 (Appendix A.3.1), under suitable technical conditions it can be shown that, if the optimal FO-IC contract is stationary, then it must be Contract W (see Online Appendix K.2 for further details). But when shocks are transient, Contract W is strictly dominated by the optimal Self-Insurance Contract (per Theorem 2), so it follows that the optimal FO-IC contract must be non-stationary. If the FOA is valid, in the sense that (i) every IC contract is FO-IC and (ii) the optimal FO-IC is IC, then it follows that the optimal full-commitment contract must also be non-stationary when shocks are transient. Details of these calculations are in the Supplementary Appendix (Bloedel, Krishna, and Strulovici 2020).

5. Optimal Contracts

In this section, we use the formulation of FO-IC contracts from Section 4 to answer, in turn, the two questions raised in Subsection 3.4.

5.1. Hidden Savings

As described in Subsection 3.2, Contract W is a special Self-Insurance Contract in which the principal does not tax the agent. This observation suggests that, while Contract W is generically strictly suboptimal within this class of contracts (Theorem 2), it would be optimal within this class in an alternative setting in which the agent could circumvent any imposed tax by directly self-insuring at the market rate ρ without the principal's knowledge. We show that Contract W is indeed optimal in such a setting, even within the strictly larger class of FO-IC contracts.

Consider the *hidden savings* variant of the model developed in Subsection 2.1, in which (a) the agent can directly self-insure via the bond market at rate ρ and (b) both the agent's endowment and trading activity are his private information (see Online Appendix I for a formal description). This model is essentially the continuous-time version of Allen (1985) and Cole and Kocherlakota (2001).Following Cole and Kocherlakota (2001, Proposition 1), it is without loss (in terms of implementable consumption processes) to restrict attention to direct revelation contracts under which the agent finds it optimal to (i)

the case on-path) or does not have a best response (as may be the case off-path). As alluded to in W11 (pp. 1230, fn. 10), this makes it difficult to directly analyze the agent's reporting problem. Allowing for jump reports, which correspond to $\Delta = \pm\infty$, effectively compactifies the agent's feasible set and restores existence.

truthfully report his endowment and (ii) engage in zero trade outside of the contract. We call such contracts *no-savings incentive compatible (NS-IC)*. When requirement (i) of NS-IC is replaced by the requirement that [**FO-IC**] holds, we say the contract is *first-order NS-IC*.

Theorem 4. *Given any $\lambda \geqslant 0$, Contract W is (i) NS-IC and (ii) the unique (hence optimal) first-order NS-IC contract.*

The proof of Theorem 4 is in Online Appendix I; a simple calculation shows the agent's Euler equation [**3.8**] with $r = \rho$ substituted into [**4.2**] and [**4.3**] uniquely pins down the geometric volatility at $k^*_\circ$. This result mirrors the main findings of Allen (1985) and Cole and Kocherlakota (2001), which show that the agent's ability to covertly self-insure severely restricts the principal's ability to provide additional insurance, and therefore leads to qualitatively different contracts than arise in the full-commitment model (see also Thomas and Worrall 1990, p. 371; Ljunqvist and Sargent 2000, Ch. 20). Intuitively, in the hidden savings model, the agent only cares about the expected present value of the principal's transfers because he can self-insure against fluctuations in the realized transfers. Thus, the principal must provide the same present value to the agent regardless of his report, for otherwise the agent would lie to obtain a higher present value. This causes a complete unravelling of insurance provision.

5.2. The Curious Case of Permanent Shocks

In the generic case of transient shocks, the preceding results imply that the optimal [**FO-IC**] (or renegotiation-proof) contract strictly dominates Contract W, which is the optimal first-order NS-IC contract. Moreover, it can be shown that, under suitable conditions, the optimal [**FO-IC**] is strictly dominated by the optimal full-commitment contract. This strict ranking of contracts is consistent with the received wisdom that hidden savings generally imposes a binding constraint (cf. Cole and Kocherlakota 2001) and that renegotiation-proofness is a binding constraint when types are persistent (cf. Fernandes and Phelan 2000). Strikingly, this wisdom is overturned when shocks are permanent. Recall that, when $\lambda = 0$, Contract W is the optimal stationary contract and optimal first-order NS-IC contract.

Theorem 5. *If $\lambda = 0$, then Contract W is the unique optimal FO-IC contract.*

The proof of Theorem 5 is in the Supplementary Appendix (Bloedel, Krishna, and Strulovici 2020). This result implies the "if" direction of Proposition 2.2 and, modulo the discrepancy between [**IC**] and [**FO-IC**] (see Remark 3), confirms the full-commitment optimality of Contract W asserted in W11 (pp. 1252-56). The intuition is similar to that given in Subsection 3.3: because the principal cannot manipulate the agent's risk exposure when $\lambda = 0$ (by Lemma H.1(ii)), she faces no tradeoff between insurance and incentive provision, and so simply sets $\beta_0 = \rho$ to eliminate the drift in the agent's continuation utility. What is potentially surprising about Theorem 5 is that this intuition remains valid even when the principal can commit to (essentially) any history-dependent contract.

Why does the principal not benefit from her additional commitment power? We argue that it is because, when $\lambda = 0$, FO-IC *requires* that the agent be indifferent among essentially *all* reporting strategies. To formalize this idea, for technical reasons we restrict to *well-behaved* reporting strategies that do not cause m or q to "explode." (See Online Appendix J.0.1 for a formal definition, and Lemma J.2 therein for demonstration that a broad class of strategies, including truthtelling, are well-behaved.)

Theorem 6. *If $\lambda = 0$, then the agent is indifferent among all well-behaved reporting strategies given any FO-IC contract.*

The proof of Theorem 6 is in Online Appendix J.0.2. In effect, Theorem 6 says that *all* of the agent's *global* incentive constraints hold with equality under *any* FO-IC contract. This sheds light on the coincidence of the optimal [**FO-IC**], stationary, and first-order NS-IC contracts under permanent shocks: even when the principal has full commitment power, she cannot insure the agent because of the large multiplicity of constraints. For comparison, recall that in discrete-time models with iid types, only the agent's "local downward" incentive constraints hold with equality under the optimal contract: at any history, the agent is indifferent between truthfully reporting and marginally under-reporting his endowment for an instant, but strictly prefers truthfully reporting to over-reporting his endowment or under-reporting it by more than an infinitesimal amount (cf. Thomas and Worrall 1990).[29] By contrast, when $\lambda = 0$, our agent is indifferent between truthfully reporting and making misreports of (essentially) arbitrary size, including those that overstate his endowment. While it is known that persistence may cause *some* global incentive constraints to hold with equality under the *optimal* contract (cf. Battaglini and Lamba 2019), to our knowledge the present setting is the first in which (essentially) *all* incentive constraints hold with equality under *all* FO-IC contracts.

It is observed in W11 (on p. 1256) that the agent is indifferent among all reporting strategies under Contract W because it specifies a transfer process s that is deterministic,

(29) We emphasize that the contrast here is between permanent and iid (perfectly transient) shocks, not between continuous- and discrete-time models. Indeed, as noted in Section 6, it can be shown that an analogue to Theorem 6 emerges in the discrete-time version of the present model.

i.e., that depends on time but not the agent's reports.[30] This argument extends to any FO-IC contract with a deterministic transfers, which can be shown (under $\lambda = 0$) to be equivalent to having a deterministic β process (thus encompassing all Proportional-Utility Contracts). By contrast, Theorem 6 applies to *all* FO-IC contracts, allowing for any y-adapted β process. This requires different mathematical arguments and highlights that the agent's indifference under permanent shocks is a robust consequence of FO-IC, rather than the restriction to a particular subclass of contracts.[31]

5.2.1. *Implications for Long-Run Properties*

Given the importance of immiseration to the literature on dynamic insurance, it is worth understanding why it fails under the optimal full-commitment (or at least optimal FO-IC) contract when shocks are permanent.[32, 33] This failure is especially notable given Bloedel, Krishna, and Leukhina's (2020) (henceforth BKL20's) finding that immiseration emerges in a broad class of discrete-time insurance settings in which the agent's private type follows an arbitrary finite-state, fully-connected Markov process. Because such processes can approximate *any* ergodic Markov process, BKL20's result suggests that immiseration is to be expected in a wide range of environments — including arbitrarily good discrete-time approximations of the present model with $\lambda > 0$ (though establishing that the optimal discrete-time contract converges to the optimal continuous-time contract requires additional argument).

Importantly, even though Contract W is optimal when shocks are permanent, the

(30) To see this, recall that when $\lambda = 0$ consumption under Contract W evolves as $\mathrm{d}c_t = (\theta\sigma^2/2)\,\mathrm{d}t + \sigma\,[\mathrm{d}W_t + (\Delta_t/\sigma)\,\mathrm{d}t]$ (see [**2.5**]), while the agent's report evolves as $\mathrm{d}y_t = (\mu + \Delta_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$. Thus, the transfer $s_t = c_t - y_t$ evolves as $\mathrm{d}s_t = (\theta\sigma^2/2 - \mu)\,\mathrm{d}t$, implying that the process s is deterministic and, moreover, absolutely continuous as a function of time.

(31) It is easily shown that there exist many FO-IC contracts with non-deterministic β. In principle, allowing for such history-dependence may be required for optimality even when, as is the case here, geometric sensitivity k is necessarily constant. For instance, in a related setting, Chen (2017) shows that the optimal contract involves history-dependent β, even though all FO-IC contracts satisfy $k_t \equiv 0$. That paper also obtains an indifference result similar to Theorem 6, but for a different reason: in that model, the agent's preferences are independent of his private type (which determines the principal's preferences).

(32) This discussion extends almost verbatim to the taste-shock example studied in Section 8 of W11, which focuses exclusively on the case without mean-reversion.

(33) We emphasize that even though Theorem 3 shows that the optimal Self-Insurance Contract results in bliss when λ is sufficiently small, because that result pertains to a restricted class of contracts it does not necessarily have any implications for the optimal full-commitment contract. Indeed, we argue below that the full-commitment optimum is likely to induce immiseration even as $\lambda \to 0$, and yield bliss only exactly at $\lambda = 0$.

explanations given in W11 for its long-run properties are misleading in important ways and seem to have caused some confusion in the literature. These issues are discussed further in Online Appendix K. Below, we instead use Theorem 6, together with the results of BKL20, to argue that the long-run properties of Contract W derive from special, non-generic features of the present model with permanent shocks.

Role of Agent Indifference. Thomas and Worrall (1990) and BKL20 present the basic intuition for the optimality of immiseration as follows. Because the agent is risk-averse, the principal finds it costly to provide incentives (i.e., make q_t vary with the report y_t). While the strength of the agent's incentives is determined by the sensitivity of his *utility*, the principal's cost of providing incentives is determined by the sensitivity of the agent's *consumption* (because the principal is risk neutral). Thus, all else equal, the principal can provide *higher-powered incentives* at the same (instantaneous) cost when the level of promised utility q_t is *lower* because the agent's marginal utility of consumption is *higher*. By sending $q_t \to -\infty$ as $t \to \infty$, the principal can use the specter of arbitrarily high-powered (but cheap) incentives in later periods to reduce the variability of consumption needed to maintain incentive compatibility in earlier periods. It is optimal to backload incentives in this manner because the principal wants to smooth her costs over time (again, due to the agent's risk aversion).

Put slightly differently, immiseration arises from the principal's desire to relax (i.e., reduce the shadow cost of) the agent's incentive constraints over time. Theorem 6 describes a sense in which this is impossible when shocks are permanent: because all (global) incentive constraints always hold with equality, there is no scope for relaxing them. Thus, the mechanism underlying immiseration is shut off by fiat in the present model with $\lambda = 0$.

Failure of Martingale Property. The connection between immiseration and the Martingale Convergence Theorem — as articulated by Thomas and Worrall (1990) and generalized by BKL20 — provides a more formal perspective on the above intuition. An important insight from BKL20 is that the appropriate notion of the principal's "marginal cost" is the directional derivative of her value function corresponding to the marginal cost of increasing the agent's promised utility *while holding the schedule of information rents fixed*. Applying the FOA to the present model, the analogous object is the partial derivative $J_q(y, q, p)$, which is the marginal cost of increasing the level of promised utility while holding its slope (i.e., information rents) fixed. When this partial derivative exists, the principal's cost-smoothing under the optimal contract renders the process $J_q(y_t, q_t, p_t)$

a strictly positive martingale, which must converge by the Martingale Convergence Theorem. The essence of BKL20's argument is that $J_q(y_t, q_t, p_t) \to 0$ as $t \to 0$ and that this implies that $q_t, p_t \to -\infty$ almost surely.

It is easy to see that this line of reasoning fails when $\lambda = 0$. Lemma H.1(ii) shows that FO-IC *requires* the proportionality $p = \theta q$ in this case. Because q and p cannot be varied separately, $J_q(y, q, p)$ is *not* a well-defined object. While the *total* derivative $J_q^W(y, q) = -1/(\rho\theta q) > 0$ exists (where J^W is the cost of Contract W given in [**3.12**]), it evolves as

$$\mathrm{d}J_q^W(y_t, q_t) = -\frac{\sigma^2\theta}{\rho q_t}\,\mathrm{d}t + \frac{\sigma}{\rho q_t}\,\mathrm{d}W_t, \tag{5.1}$$

and therefore has strictly positive drift and is unbounded above.[34] Thus, this process is a non-negative, unbounded *sub*martingale, for which the Martingale Convergence Theorem has no implications. Instead, as we have seen in Subsection 3.2, the long-run properties of Contract W are governed by a different martingale, namely, the agent's Euler equation. Thus, under Contract W we have $q_t, p_t \to 0$ and $J_q^W(y_t, q_t) \to +\infty$, which is the polar opposite of BKL20's result.

Importantly, this failure of the martingale property is special to the $\lambda = 0$ case. When $\lambda > 0$, Lemma H.1(i) shows that the domain D of implementable (q, p) pairs has full dimension, so that $J_q(y, q, p)$ is well-defined and, by standard envelope theorem logic, will define a martingale at the optimum (cf. Farhi and Werning 2013, Section 5). And even when $\lambda = 0$, Lemma H.1(ii) relies sensitively on the assumption of exponential utility; it is easy to see that D can have full dimension under alternative functional form assumptions. It is natural to conjecture that (a) BKL20's martingale convergence arguments could be extended to establish immiseration under the optimal full-commitment contract whenever D has full dimension, and (b) D has full dimension for "generic" endowment processes and (smooth, concave) utility functions.

6. Concluding Remarks

Discrete-Time Model. Most of our results hold in the natural discrete-time analogue of the model in which the agent's endowment follows a discrete-time AR(1) process with Gaussian shocks (and many also extend to more general discrete-time endowment

(34) We refer to this as the total derivative because $J^W(y, \cdot)$ is a projection of the "full" value function $J(y, \cdot, \cdot)$ from D onto $\mathbb{R}_{--}$, where D is the set of pairs (q, p) that are incentive compatible; see Online Appendix H. The SDE [**5.1**] follows from an application of Itô's lemma to the SDE for q given in property (i) in Subsection 2.2.

processes). The material from Section 3 goes through effectively verbatim, with the discrete-time self-insurance solution obtained by Caballero (1990) and Wang (2003) replacing our continuous-time solution in Proposition 3.1, and the discrete-time analogue of Contract W defined to be the self-insurance solution that arises when the agent discounts at the market rate. (The long-run properties analogous to those in Theorem 3 can be shown to hold in discrete-time when the agent is sufficiently patient, and we conjecture that they hold independently of the agent's discrete-time discount factor.) All material from Sections 4–5 can be shown to go through effectively verbatim by applying the discrete-time FOA due to Pavan, Segal, and Toikka (2014). The discrete-time analogue to Theorem 6 states that the agent is indifferent among all reporting strategies that involve deviations in finitely-many periods whenever the endowment follows a Markov process (with full-support, absolutely-continuous transition probabilities) that exhibits "permanent shocks," in the sense that the endowment's "impulse response functions" (in the language of Pavan, Segal, and Toikka 2014) are constant and equal to 1. See Online Appendix K for further discussion of the relation between the discrete- and continuous-time versions of the model.

An Open Question. We conclude with an open question: *What is the optimal full-commitment contract in the present model?* A first step would be to verify that every IC contract is necessarily FO-IC (cf. Remark 3). By Theorem 5, this would prove that Contract W is optimal when $\lambda = 0$. For $\lambda > 0$, we conjecture that the full-commitment optimum is non-stationary, assuming the FOA is valid (which would need to be verified, cf. Remark 3). We conjecture that the optimal contract is in fact non-Markovian with respect to the variables (y_t, q_t). This would make it challenging to obtain a full solution, as the extra state variable p_t would render the principal's HJB equation a second-order PDE. Nonetheless, we conjecture that the full-commitment optimum exhibits worse risk-sharing as persistence increases (as shown in Theorem 2 for the optimal stationary contract).[35] Finally, as noted in Subsection 5.2.1 above, we conjecture that BKL20's martingale convergence arguments could be adapted to establish that the full-commitment optimum induces immiseration.

(35) This conjecture is borne out in numerical comparative statics exercises from BKL20, Zhang (2009), and Kapička (2013).

Appendices

A. Computing the Cost of Self-Insurance Contracts

A.1. First-Best Contract

Lemma A.1. The first-best (full-information) optimal contract given initial condition (b_0, q_0) consists of the principal giving the agent the constant consumption stream $\bar{c}(q_0, \rho)$ in exchange for the agent's income. The principal's first-best cost function $J^*(q, b)$ is given by

$$J^*(q,b) = \frac{\bar{c}(q,\rho)}{\rho} - \frac{\mu}{\rho(\rho+\lambda)} - \frac{b}{\rho+\lambda} \tag{A.1}$$

Proof. By standard arguments, the principal gives the agent a constant flow utility process; to achieve lifetime utility q_0, this requires the agent consume the constant amount $\bar{c}(q_0, \rho)$ at each time. The principal's lifetime cost from this promise of consumption is $\bar{c}(q_0, \rho)/\rho$. Lemma 2.4 in Bloedel, Krishna, and Strulovici (2020) shows that $\mathbf{E}[\int_0^\infty e^{-\rho t} b_t \,\mathrm{d}t] = b_0/(\rho+\lambda) + \mu/\rho(\rho+\lambda) + \frac{\sigma}{\rho+\lambda}\mathbf{E}[\int_0^\infty e^{-\rho t}\,\mathrm{d}W_t]$, and Lemma 2.5 (also in Bloedel, Krishna, and Strulovici 2020) shows (when $\alpha = \rho$) that $\mathbf{E}[\int_0^\infty e^{-\rho t}\,\mathrm{d}W_t] = 0$. Display [A.1] follows immediately. □

A.2. Proof of Lemma 3.3

We begin with a preliminary lemma. Recall the value function $\Pi(b_0, q_0, r)$ defined in [3.14] above.

Lemma A.2. Let A_t^* satisfy [D.6] with $A_0 = A_0(b_0, q_0, r)$ as in [3.11]. Then:

$$\begin{aligned} A_0 - (\rho - r)\int_0^\infty e^{-\rho t} A_t^* \,\mathrm{d}t \\ &= \underbrace{\frac{\bar{c}(q_0,r)}{\rho} - \frac{b_0}{\rho+\lambda} - \frac{\mu}{\rho(\rho+\lambda)} + \frac{(r-\rho)+\sigma^2 f^2(r;\lambda)/2}{\theta\rho^2}}_{=\Pi(b_0,q_0,r)} \\ &\quad + \frac{(r-\rho)\lambda\sigma}{\rho(r+\lambda)(\rho+\lambda)}\int_0^\infty e^{-\rho t}\,\mathrm{d}W_t \end{aligned} \tag{A.2}$$

Proof. We first compute $\int_0^\infty e^{-\rho t} A_t^* \,\mathrm{d}t$. Notice that $\mathrm{d}(e^{-\rho t}A_t^*) = -\rho e^{-\rho t} A_t^*\,\mathrm{d}t + e^{-\rho t}\,\mathrm{d}A_t^*$, so that

$$\int_0^T e^{-\rho t} A_t^* \,\mathrm{d}t = \frac{A_0 - e^{-\rho T} A_T^*}{\rho} + \frac{1}{\rho}\int_0^T e^{-\rho t}\,\mathrm{d}A_t^*$$

By Lemma D.4 (with $\alpha = \rho$), we obtain $\lim_{T\to\infty} e^{-\rho T} A_T^* = 0$. Thus, we have

$$\int_0^\infty e^{-\rho t} A_t^* \,\mathrm{d}t = \frac{A_0}{\rho} + \frac{1}{\rho}\int_0^\infty e^{-\rho t}\,\mathrm{d}A_t^*$$

Plugging in the expression for $\mathrm{d}A_t^*$ from [D.6], we obtain

$$\begin{aligned}\frac{1}{\rho}\int_0^\infty e^{-\rho t}\,\mathrm{d}A_t^* &= \int_0^\infty e^{-\rho t}\left(-\frac{\rho - r}{\theta r} - \frac{\mu}{\lambda + r} + \frac{1}{2}\frac{\sigma^2 f^2(r;\lambda)}{\theta r} + \frac{\lambda}{\lambda + r} b_t\right)\mathrm{d}t \\ &= \frac{1}{\rho^2}\left(-\frac{\rho - r}{\theta r} - \frac{\mu}{\lambda + r} + \frac{1}{2}\frac{\sigma^2 f^2(r;\lambda)}{\theta r}\right) + \frac{1}{\rho}\frac{\lambda}{\lambda + r}\int_0^\infty e^{-\rho t} b_t\,\mathrm{d}t\end{aligned}$$

It follows from Lemma 2.4 (with $\alpha = \rho$) in Bloedel, Krishna, and Strulovici (2020) that

$$\int_0^\infty e^{-\rho t} b_t\,\mathrm{d}t = \frac{b_0}{\rho + \lambda} + \frac{\mu}{\rho(\rho + \lambda)} + \frac{\sigma}{\rho + \lambda}\int_0^\infty e^{-\rho t}\,\mathrm{d}W_t \tag{A.3}$$

Putting all these elements together, we find that

$$\begin{aligned} & A_0 - (\rho - r)\int_0^\infty e^{-\rho t} A_t^*\,\mathrm{d}t \\ &= \frac{r}{\rho}A_0 - \frac{(\rho - r)}{\rho}\int_0^\infty e^{-\rho t}\,\mathrm{d}A_t^* \\ &= \frac{r}{\rho}\underbrace{\left[\frac{\bar{c}(q_0, r)}{r} - \frac{b_0}{r + \lambda} - \frac{\mu}{r(r + \lambda)} + \frac{r - \rho + \sigma^2 f(r;\lambda)^2/2}{r^2\theta}\right]}_{=A_0(q_0, b_0, r)\text{ from [3.11]}} \\ &\quad - \frac{\rho - r}{\rho^2}\left(-\frac{\rho - r}{\theta r} - \frac{\mu}{\lambda + r} + \frac{1}{2}\frac{\sigma^2 f^2(r;\lambda)}{\theta r}\right) \\ &\quad - \frac{\rho - r}{\rho}\frac{\lambda}{\lambda + r}\underbrace{\left[\frac{b_0}{\rho + \lambda} + \frac{\mu}{\rho(\rho + \lambda)} + \frac{\sigma}{\rho + \lambda}\int_0^\infty e^{-\rho t}\,\mathrm{d}W_t\right]}_{\text{from [A.3]}}\end{aligned} \tag{A.4}$$

We now collect the coefficients from [A.4] and simplify. The coefficient of $\bar{c}(q_0; r)$ is $1/\rho$. The coefficient of b_0 is

$$\begin{aligned}-\frac{r}{\rho(r + \lambda)} - \frac{(\rho - r)}{\rho}\frac{\lambda}{\lambda + r}\frac{1}{\rho + \lambda} &= \frac{-1}{\rho(r + \lambda)}\left[r + \frac{\lambda(\rho - r)}{\lambda + \rho}\right] \\ &= -\frac{1}{\lambda + \rho}\end{aligned} \tag{A.5}$$

The coefficient of μ is

$$\begin{aligned} & \frac{-1}{\rho(r + \lambda)} + \frac{\rho - r}{\rho^2(\lambda + r)} - \frac{(\rho - r)}{\rho}\frac{\lambda}{\lambda + r}\frac{1}{\rho(\rho + \lambda)} \\ &= \frac{1}{\rho(r + \lambda)}\left[-1 + \frac{\rho - r}{\rho} - \frac{\rho - r}{\rho}\frac{\lambda}{\lambda + \rho}\right] \\ &= -\frac{1}{\rho(\rho + \lambda)}\end{aligned} \tag{A.6}$$

The coefficient of $\sigma^2 f^2(r;\lambda)/2$ is

[**A.7**] $$\frac{1}{r\rho\theta} - \frac{\rho - r}{r\rho^2\theta} = \frac{1}{r\rho\theta}\left[1 - \frac{\rho - r}{\rho}\right] = \frac{1}{\theta\rho^2}$$

The coefficient of $\int_0^\infty e^{-\rho t}\,\mathrm{d}W_t$ is

[**A.8**] $$\frac{(r-\rho)\lambda\sigma}{\rho(r+\lambda)(\rho+\lambda)}$$

Finally, the remaining terms in [**A.4**] (which are unaccounted for in [**A.5**]–[**A.8**]) are

[**A.9**] $$\frac{r-\rho}{r\rho\theta} + \frac{(r-\rho)^2}{\rho^2\theta r} = \frac{r-\rho}{\rho^2\theta}$$

Combining [**A.5**], [**A.6**], [**A.7**], [**A.8**], and [**A.9**], we find that [**A.4**] reduces to [**A.2**], as claimed. □

We are now in a position to prove Lemma 3.3.

Proof of Lemma 3.3. To provide the agent with q_0 utiles from a Self-Insurance Contract (b_0, q_0, r) where the agent borrows and saves at an effective rate r, the principal needs to give the agent an initial asset level A_0 so that $V(A_0, b_0) = q_0$, where V is the agent's value function in [**3.1**]. It follows from Lemmas D.2 and D.3 that A_0 should be chosen as in [**3.11**], i.e.,

$$A_0(b_0, q_0, r) = \frac{\bar{c}(q_0, r)}{r} - \frac{b_0}{r+\lambda} - \frac{\mu}{r(r+\lambda)} + \frac{r - \rho + \sigma^2 f(r;\lambda)^2/2}{r^2\theta}$$

where $\bar{c}(q_0, r) = -\theta^{-1}\log(-rq_0)$. Thus, the principal's one-time, upfront cost is A_0. His flow payments, via taxes, is $\mathbf{E}[\int_0^\infty e^{-\rho t}(\rho - r)A_t^*\,\mathrm{d}t]$. The process A_t^* is described in [**D.6**]. Lemma A.2 above shows that

$$A_0 - \mathbf{E}\left[\int_0^\infty e^{-\rho t}(\rho - r)A_t^*\,\mathrm{d}t\right] = \Pi(b_0, q_0, r) - \frac{(\rho - r)\lambda\sigma}{\rho(r+\lambda)(\rho+\lambda)}\mathbf{E}\left[\int_0^\infty e^{-\rho t}\,\mathrm{d}W_t\right]$$

That $\mathbf{E}[\int_0^\infty e^{-\rho t}\,\mathrm{d}W_t] = 0$ is established in Lemma 2.5 in Bloedel, Krishna, and Strulovici (2020).

Moreover, we have that $\mathrm{d}(e^{-\rho t}A_t^*) = -\rho e^{-\rho t}A_t^*\,\mathrm{d}t + e^{-\rho t}\,\mathrm{d}A_t^*$. Because A^* is defined by the law of motion [**3.2**], we obtain

$$\int_0^T e^{-\rho t}(\hat{c}_t^* - b_t)\,\mathrm{d}t = A_0 - (\rho - r)\int_0^T e^{-\rho t}A_t^*\,\mathrm{d}t - e^{-\rho T}A_T^*$$

Lemma D.4 (with $\alpha = \rho$) delivers that $\lim_{T\to\infty} e^{-\rho T}A_T^* = 0$. Therefore, sending $T \to \infty$ in the above display and then taking expectations yields

$$\mathbf{E}\left[\int_0^\infty e^{-\rho t}(\hat{c}_t^* - b_t)\,\mathrm{d}t\right] = A_0 - \mathbf{E}\left[\int_0^\infty e^{-\rho t}(\rho - r)A_t^*\,\mathrm{d}t\right] = \Pi(q_0, b_0, r)$$

which proves the lemma. □

B. Proof of Theorem 3

Lemma 3.3 shows that the principal's problem of choosing an optimal Self-Insurance Contract may be written as

$$\text{[B.1]} \qquad \inf_{r>0} \left[-\frac{\log(r)}{\theta\rho} + \frac{r-\rho+\sigma^2 f(r;\lambda)^2/2}{\theta\rho^2} - \frac{\log(-q_0)}{\theta\rho} - \frac{b_0}{\rho+\lambda} - \frac{\mu}{\rho(\rho+\lambda)} \right]$$

Let $(\lambda,\sigma) \mapsto r^*(\lambda,\sigma)$ denote a selection from the argmin correspondence of [B.1]. Let $k^*(\lambda,\sigma) := f(r^*(\lambda,\sigma),\lambda)$.

We begin with a result about the long-run properties of consumption, flow utility, and promised utility.

Lemma B.1. Let $D^* : \mathbb{R}_+ \times \mathbb{R}_{++} \to \mathbb{R}$ be defined by

$$\text{[B.2]} \qquad D^*(\lambda,\sigma) := r^*(\lambda,\sigma) - \rho + \frac{\sigma^2}{2} f\left(r^*(\lambda,\sigma);\lambda\right)^2$$

Then, the following hold:

(a) If $D^*(\lambda,\sigma) > 0$, then $\hat{c}^*_t \to \infty$ and $V_t, u(\hat{c}^*_t) \to 0$ as $t \to \infty$, a.s.
(b) If $D^*(\lambda,\sigma) < 0$, then $\hat{c}^*_t \to -\infty$ and $V_t, u(\hat{c}^*_t) \to -\infty$ as $t \to \infty$, a.s.
(c) If $D^*(\lambda,\sigma) = 0$, then $\limsup_{t\to\infty} \hat{c}^*_t = +\infty$, $\limsup_{t\to\infty} V_t, u(\hat{c}^*_t) = 0$, and $\liminf_{t\to\infty} \hat{c}^*_t = \liminf_{t\to\infty} V_t, u(\hat{c}^*_t) = -\infty$ a.s.

Proof. Recall that $\hat{c}$ follows the arithmetic Brownian motion [3.4]. The long-run properties of $\hat{c}^*$ in point (i) and (ii) then follow from Lemma 2.1 in Bloedel, Krishna, and Strulovici (2020); the long-run properties of $\hat{c}^*$ in point (iii) follow from standard properties of Brownian motion (e.g., the Law of the Iterated Logarithm stated in [2.4]–[2.5] in Bloedel, Krishna, and Strulovici 2020). The long-run properties of $u(\hat{c}^*_t)$ and V_t then follow from the long-run properties of $\hat{c}^*$ and the Continuous Mapping Theorem (or, equivalently, by long-run results for geometric Brownian motion as in Oksendal 2010, p. 65). □

We may now prove the theorem itself.

Proof of Theorem 3. We consider each point in turn.
Point (i): From [3.13], we see that the drift of V_t is $V_t(\rho - r)$. Because $V_t < 0$, it follows from Theorem 2(i) that $V_t(\rho - r) < 0$ if $\lambda > 0$ and $V_t(\rho - r) = 0$ if $\lambda = 0$.
Preliminaries for points (ii) and (iii): Multiplying the first-order condition [3.16] through by $\theta\rho r^*(\lambda,\sigma) > 0$ yields the equivalent optimality condition

$$\text{[B.3]} \qquad -\rho + r^*(\lambda,\sigma) + \sigma^2 r^*(\lambda,\sigma) f\left(r^*(\lambda,\sigma);\lambda\right) \frac{\partial}{\partial r} f\left(r^*(\lambda,\sigma);\lambda\right) = 0.$$

Plugging [B.3] into [B.2] yields

$$D^*(\lambda,\sigma) = \underbrace{\sigma^2 f\left(r^*(\lambda,\sigma);\lambda\right)}_{>0} \cdot \underbrace{\left[\frac{1}{2} f\left(r^*(\lambda,\sigma);\lambda\right) - r^*(\lambda,\sigma)\frac{\partial}{\partial r} f\left(r^*(\lambda,\sigma);\lambda\right)\right]}_{=:\,\hat{D}^*(\lambda,\sigma)}$$

so it suffices to study the sign of $\hat{D}^*(\lambda,\sigma)$. Short calculations yield

$$\hat{D}^*(\lambda,\sigma) > 0 \iff r^*(\lambda,\sigma) > \lambda \iff k^*(\lambda,\sigma) > \theta/2, \tag{B.4}$$

$$\hat{D}^*(\lambda,\sigma) < 0 \iff r^*(\lambda,\sigma) < \lambda \iff k^*(\lambda,\sigma) < \theta/2. \tag{B.5}$$

Point (ii): Let $\sigma > 0$ be given. By Theorem 2(ii), there exists a unique $\overline{\lambda}(\sigma) > 0$ such that $k^*(\lambda,\sigma) > \theta/2$ if and only if $\lambda \in \left[0, \overline{\lambda}(\sigma)\right)$ and $k^*(\lambda,\sigma) < \theta/2$ if and only if $\lambda > \overline{\lambda}(\sigma)$. (Notice that $\lim_{\lambda\to 0} r^*(\lambda,\sigma) = \rho$ implies that $\lim_{\lambda\to 0} k^*(\lambda,\sigma) = \theta$.) The result then follows from [B.4]–[B.5] and Lemma B.1.

Point (iii): Let $\lambda > 0$ be given. By Theorem 2(iii), there exists a unique $\overline{\sigma}(\lambda) \geqslant 0$ such that $k^*(\lambda,\sigma) > \theta/2$ if and only if $\sigma \in (0, \overline{\sigma}(\lambda))$ and $k^*(\lambda,\sigma) < \theta/2$ if and only if $\sigma > \overline{\sigma}(\lambda)$. Theorem 2(iii) also shows that $\lim_{\sigma\to 0} k^*(\lambda,\sigma) = k^*_\circ(\lambda)$. Notice that $k^*_\circ(\lambda) > \theta/2$ if and only if $\rho > \lambda$. Thus, $\overline{\sigma}(\lambda) > 0$ if and only if $\rho > \lambda$, as desired. The result then follows from [B.4]–[B.5] and Lemma B.1. □

We conclude with the lemma, mentioned in footnote 22 in the main text, that characterizes the long-run properties of the agent's asset holdings.

Lemma B.2. Under the optimal Self-Insurance Contract, the agent's optimal asset holdings A^* satisfy the following properties:

(i) If $\lambda > 0$, then (a) $\hat{c}^*_t \to +\infty$ implies that $A^*_t \to +\infty$ and (b) $\hat{c}^*_t \to -\infty$ implies that $A^*_t \to -\infty$.

(ii) If $\lambda = 0$, then (a) $A^*_t \to +\infty$ when $D^*(\lambda,\sigma) > \frac{\mu}{r^*(\lambda,\sigma)}$ and (b) $A^*_t \to -\infty$ when $D^*(\lambda,\sigma) < \frac{\mu}{r^*(\lambda,\sigma)}$.

Proof. We consider each point in turn.

Point (i): Let $\lambda > 0$. Rewrite [D.7] as

$$A^*_t = A_0 + \left[\frac{r-\rho}{\theta r} + \frac{1}{2}\frac{\sigma^2 f(r;\lambda)^2}{\theta r}\right] t + \frac{\lambda}{\lambda+r}\left[\left[\int_0^t b_\tau \,\mathrm{d}\tau - \frac{\mu}{\lambda}\cdot t\right]\right.$$

$$= A_0 + \left[\frac{r-\rho}{\theta r} + \frac{1}{2}\frac{\sigma^2 f(r;\lambda)^2}{\theta r}\right] t + \frac{\lambda}{\lambda+r}\left[\left[b_0 - \frac{\mu}{\lambda}\right]\left[\frac{1-e^{-\lambda t}}{\lambda}\right] + \frac{\sigma W_t - X_t}{\lambda}\right] \tag{B.6}$$

where the second line follows from substituting in the closed-form expression for $\int_0^t b_\tau \,\mathrm{d}\tau$ given in [2.2] and X_t is as defined in [2.2] (both displays in Bloedel, Krishna, and

Strulovici 2020). Now, Lemma 2.1 in Bloedel, Krishna, and Strulovici (2020) shows that $\lim_{t\to\infty} W_t/t = 0$ almost surely and (the proof of) Lemma 2.3(i) in Bloedel, Krishna, and Strulovici (2020) shows that $\lim_{t\to\infty} X_t/t = 0$ almost surely. It then follows from [B.6] that

$$\lim_{t\to\infty} \frac{A_t^*}{t} = \frac{r-\rho}{\theta r} + \frac{1}{2}\frac{\sigma^2 f(r;\lambda)^2}{\theta r}$$

almost surely. The claim the follows from Lemma B.1 and standard arguments.

Point (ii): Let $\lambda = 0$. Then [D.7] simply reduces to

$$A_t^* = A_0 + \left(\frac{r-\rho}{\theta r} + \frac{1}{2}\frac{\sigma^2 f(r;\lambda)^2}{\theta r} - \frac{\mu}{r}\right) t$$

from which the asserted long-run properties are immediate. □

C. Proof of Proposition 4.3

It suffices to verify the stationary and proportional utility properties of the q and u processes; equality of the principal's cost then follows from straightforward accounting.

(i) $\implies$ (ii) and (iii): This is immediate from the proportional utility and stationary properties of Self-Insurance Contracts, stated in [3.9] and [3.10], respectively. (See Lemmas D.6 and D.7 for proofs.)

(ii) $\implies$ (i) and (iii): Consider a Proportional-Utility Contract with utility delivery rate $\beta_\circ > 0$. Plugging $u_t \equiv \beta_\circ q_t$ into [4.1] yields

[C.1] $$p_t = \beta_\circ \theta \int_t^\infty e^{-(\rho+\lambda)(\tau-t)}\, \mathbf{E}_t^*\left[q_\tau\right] \mathrm{d}\tau$$

where we have applied Tonelli's Theorem to interchange the order of integration. By Itô's lemma and [4.2] (with $u_t \equiv \beta_\circ q_t$), integrating $\mathrm{d}\left[e^{(\beta_\circ-\rho)v} q_v\right]$ over $v \in [t,\tau]$ for some $t < \tau$ yields

[C.2] $$e^{(\beta_\circ-\rho)(\tau-t)} q_\tau = q_t + \sigma \int_t^\tau e^{(\beta_\circ-\rho)(v-t)} \gamma_v \,\mathrm{d}W_v$$

Because the Martingale Representation Theorem used to derive the representations [4.2]–[4.3] establishes that the process $(M_s)_{s\in[t,\tau]}$ defined by $M_s := \int_t^s e^{(\beta_\circ-\rho)(v-t)}\gamma_v \,\mathrm{d}W_v$ is a (true, rather than local) martingale and $\max_{v\in[t,\tau]} e^{(\beta_\circ-\rho)(v-t)} < \infty$, it can be shown that the stochastic integral in [C.2] has mean zero. Thus, taking expectations in [C.2] yields

$$\mathbf{E}_t^*\left[e^{(\beta_\circ-\rho)(\tau-t)} q_\tau\right] = q_t$$

Multiplying both sides of this expression by $\exp\left(-(\lambda+\beta_\circ)(\tau-t)\right)$ yields

[C.3] $$e^{-(\lambda+\rho)(\tau-t)}\, \mathbf{E}_t^*\left[q_\tau\right] = e^{-(\lambda+\beta_\circ)(\tau-t)} q_t$$

Plugging [C.3] into [C.1] then yields

$$p_t = \beta_\circ \theta q_t \int_t^\infty e^{-(\lambda+\beta_\circ)(\tau - t)}\, \mathrm{d}\tau = f(\beta_\circ;\lambda) q_t$$

as desired.

(iii) $\implies$ (i) and (ii) when $\lambda > 0$: Let $\lambda > 0$. Consider a Stationary Contract with $k_\circ \in (0, \theta)$. Under this contract, the drift (respectively, volatility) of $(p_t)_{t \geqslant 0}$ must equal the drift (respectively, volatility) of $(k_\circ q_t)_{t \geqslant 0}$ almost everywhere by the unique decomposition property for Itô processes. Equating the drift terms in [4.2] (scaled by $k_\circ$) and [4.3] yields $k_\circ (\rho q_t - u_t) = \rho p_t - \lambda \gamma_t - \theta u_t$. Similarly, equating the volatility terms yields $k_\circ \gamma_t = Q_t$. Combining these equalities with the hypothesized identify $k_\circ q_t = p_t$ and [FO-IC], we see that this contract satisfies $k_\circ \equiv f(\beta_t;\lambda)$, $\gamma_t \equiv -k_\circ q_t$, and $Q_t \equiv -k_\circ^2 q_t$. Because $f(\cdot;\lambda)$ is strictly increasing for $\lambda > 0$, this implies that β is a constant process. It follows that this contract is Proportional-Utility and a Self-Insurance Contract.

(iii) does not imply (i) and (ii) when $\lambda = 0$: By Lemma H.1, every FO-IC contract is stationary with $k_\circ = \theta$ when $\lambda = 0$. However, it is easy to verify that there exist contracts that are both FO-IC and IC, yet violate the proportional utility property. For instance, it can be shown that any contract that provides a deterministic but time-varying transfer suffices.

References

Allen, Franklin. 1985. "Repeated Principal-Agent Relationships with Hidden Lending and Borrowing". *Economics Letters* 17 (1-2): 27–31. (Cit. on pp. 6, 26, 27).

Atkeson, Andrew, and Robert E Lucas Jr. 1995. "Efficiency and Equality in a Simple Model of Efficient Unemployment Insurance". *Journal of Economic Theory* 66:64–88. (Cit. on p. 2).

Atkeson, Andrew, and Robert E. Lucas Jr. 1992. "On Efficient Distribution with Private Information". *Review of Economic Studies* 59 (3): 427–453. (Cit. on pp. 2, 67).

Battaglini, Marco, and Rohit Lamba. 2019. "Optimal Dynamic Contracting: The First-Order Approach and Beyond". *Theoretical Economics* 14 (4): 1435–1482. (Cit. on pp. 10, 28).

Bergemann, Dirk, and Philipp Strack. 2015. "Dynamic Revenue Maximization: A Continuous Time Approach". *Journal of Economic Theory* 159:819–853. (Cit. on p. 23).

Bloedel, Alexander W., R. Vijay Krishna, and Oksana Leukhina. 2018. *Insurance and Inequality with Persistent Private Information*. Tech. rep. Stanford University and Federal Reserve Bank of St. Louis. (Cit. on p. 1).

– . 2020. “Insurance and Inequality with Persistent Private Information”. Working paper, Stanford University and Federal Reserve Bank of St. Louis. (Cit. on pp. 1, 2, 4, 6, 10, 29).

Bloedel, Alexander W., R. Vijay Krishna, and Bruno Strulovici. 2020. “Supplementary Appendix to Persistent Private Information Revisited”. (Cit. on pp. 7, 26, 27, 33–38, 45, 51, 63).

Caballero, Ricardo J. 1990. “Consumption Puzzles and Precautionary Savings”. *Journal of Monetary Economics* 25 (1): 113–136. (Cit. on pp. 5, 13, 16, 32).

Chamberlain, Gary, and Charles A. Wilson. 2000. “Optimal Intertemporal Consumption under Uncertainty”. *Review of Economic Dynamics* 3 (3): 365–395. (Cit. on pp. 4, 16).

Chen, Yi. 2017. *Dynamic Communication with Commitment*. Tech. rep. Yale University. (Cit. on pp. 2, 25, 29).

Cisternas, Gonzalo. 2017. “Two-sided learning and the ratchet principle”. *Review of Economic Studies* 85 (1): 307–351. (Cit. on pp. 2, 10).

Cole, Harold L., and Narayana R. Kocherlakota. 2001. “Efficient Allocations with Hidden Income and Hidden Storage”. *Review of Economic Studies* 68 (3): 523–542. (Cit. on pp. 6, 26, 27).

Cvitanić, Jakša, and Jianfeng Zhang. 2012. *Contract Theory in Continuous-Time Models*. Springer Science & Business Media. (Cit. on p. 23).

DeMarzo, Peter M., and Yuliy Sannikov. 2016. “Learning, termination, and payout policy in dynamic incentive contracts”. *Review of Economic Studies* 84 (1): 182–236. (Cit. on p. 2).

Edlin, Aaron S., and Chris Shannon. 1998. “Strict monotonicity in comparative statics”. *Journal of Economic Theory* 81 (1): 201–219. (Cit. on pp. 54, 55).

Farhi, Emmanuel, and Iván Werning. 2007. “Inequality and Social Discounting”. *Journal of Political Economy* 115 (3): 365–402. (Cit. on p. 2).

– . 2013. “Insurance and Taxation over the Lifecycle”. *Review of Economic Studies* 80 (2): 596–635. (Cit. on pp. 5, 31, 67).

Fernandes, Ana, and Christopher Phelan. 2000. “A Recursive Formulation for Repeated Agency with History Dependence”. *Journal of Economic Theory* 91:223–247. (Cit. on p. 27).

Golosov, Mikhail, Narayana R. Kocherlakota, and Aleh Tsyvinski. 2003. “Optimal Indirect and Capital Taxation”. *Review of Economic Studies* 70 (3): 569–587. (Cit. on pp. 19, 66, 67).

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski. 2016. “Redistribution and Social Insurance”. *American Economic Review* 106 (2): 359–386. (Cit. on p. 67).

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 p. 2).

He, Zhiguo, Bin Wei, Jianfeng Yu, and Feng Gao. 2017. “Optimal Long-Term Contracting with Learning”. *Review of Financial Studies* 30 (6): 2006–2065. (Cit. on pp. 2, 23, 25).

Kapička, Marek. 2013. “Efficient allocations in dynamic private information economies with persistent shocks: A first-order approach”. *The Review of Economic Studies*: rds045. (Cit. on pp. 10, 23, 32, 65).

Karatzas, Ioannis, and Steven E. Shreve. 1998. *Brownian Motion and Stochastic Calculus*. Second Edition. New York, NY: Springer. (Cit. on p. 48).

Ljunqvist, Lars, and Thomas J. Sargent. 2000. *Recursive Macroeconomic Theory*. 1st ed. Cambridge, Ma: MIT Press. (Cit. on pp. 4, 16, 27).

Oksendal, Bernt. 2010. *Stochastic Differential Equations*. Sixth. Springer-Verlag. (Cit. on p. 36).

Pavan, Alessandro, Ilya Segal, and Juuso Toikka. 2014. “Dynamic Mechanism Design: A Myersonian Approach”. *Econometrica* 82 (2): 601–653. (Cit. on pp. 8, 23, 32, 67).

Phelan, Christopher. 1998. “On the Long Run Implications of Repeated Moral Hazard”. *Journal of Economic Theory* 79:174–191. (Cit. on p. 2).

– . 2006. “Opportunity and social mobility”. *Review of Economic Studies* 73 (2): 487–504. (Cit. on p. 2).

Prat, Julien, and Boyan Jovanovic. 2014. “Dynamic contracts when the agent’s quality is unknown”. *Theoretical Economics* 9 (3): 865 –914. (Cit. on pp. 2, 23, 25, 67).

Sadzik, Tomasz, and Ennio Stacchetti. 2015. “Agency Models with Frequent Actions”. *Econometrica* 83 (1): 193–237. (Cit. on p. 66).

Sannikov, Yuliy. 2008. “A Continuous-Time Version of the Principal-Agent Problem”. *Review of Economic Studies* 75 (3): 957–984. (Cit. on p. 59).

– . 2014. *Moral hazard and long-run incentives*. Tech. rep. Princeton University. (Cit. on pp. 2, 23).

Sotomayor, Marilda. 1984. “On Income Fluctuations and Capital Gains”. *Journal of Economic Theory* 32 (1): 14–35. (Cit. on p. 16).

Strulovici, Bruno. 2020. *Renegotiation-Proof Contracts with Persistent Private Information*. Tech. rep. Northwestern University. (Cit. on pp. 1, 3, 5, 23, 25, 64).

– . 2011. *Renegotiation-proof Contracts with Moral Hazard and Persistent Private Information*. Tech. rep. Center for Mathematical Studies in Economics and Management Science. (Cit. on p. 1).

Thomas, Jonathan P., and Tim Worrall. 1990. "Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-Agent Problem". *Journal of Economic Theory* 51:367–390. (Cit. on pp. 2, 6, 27, 28, 30, 65, 67).

Touzi, Nizar. 2018. "Stochastic Control and Application to Finance". Manuscript, Ecole Polytechnique Paris. (Cit. on p. 44).

Wang, Neng. 2003. "Caballero meets Bewley: The Permanent-Income Hypothesis in General Equilibrium". *American Economic Review* 93 (3): 927–936. (Cit. on pp. 5, 13, 16, 32).

Williams, Noah. 2015. "A Solvable Continuous Time Dynamic Principal-Agent Model". *Journal of Economic Theory* 159:989–1015. (Cit. on pp. 2, 25).

– . 2011. "Persistent Private Information". *Econometrica* 79 (4): 1233–1275. (Cit. on pp. 1, 2, 65).

Yong, Jiongmin, and Xun Yu Zhou. 1999. *Stochastic Controls: Hamiltonian Systems and HJB Equations*. Vol. 43. Springer Science & Business Media. (Cit. on p. 44).

Zhang, Yuzhe. 2009. "Dynamic Contracting with Persistent Shocks". *Journal of Economic Theory* 144:635–675. (Cit. on pp. 2, 32, 65, 67).

Online Appendices

D. Proofs for the Agent's Self-Insurance Problem

In this appendix, we formally define and solve the agent's self-insurance problem from Subsection 3.1. In the sequel, it will be useful to recall the function $f(r;\lambda)$ defined in the statement of Proposition 3.1 as $f(r;\lambda) := \frac{r\theta}{r+\lambda}$.

D.1. Preliminaries

We begin with a formal definition of the agent's feasible set in the self-insurance problem.

Definition D.1. The consumption process $\hat{c}$ is *feasible at* (A,b), denoted $\hat{c} \in \mathscr{A}(A,b)$, if:

(i) $\hat{c}$ is b-adapted, and

(ii) $A^{\hat{c}}$ is a (strong) solution to [3.2] and satisfies the no-Ponzi condition [3.3].

Fixing an initial condition b_0 for the endowment process b, it is standard to show that a b-adapted consumption process $\hat{c}$ satisfies the sequential budget constraints [3.2]–[3.3] if and only if it almost surely satisfies the the intertemporal budget constraint

[D.1] $$\int_0^\infty e^{-rt}\hat{c}_t \,\mathrm{d}t \leqslant \int_0^\infty e^{-rt} b_t \,\mathrm{d}t + A_0$$

Given any $\xi \in \mathbb{R}$, it is then straightforward to verify via [D.1] that (i) $\hat{c} \in \mathscr{A}(A,b)$ if and only if $\hat{c} + r\xi \in \mathscr{A}(A+\xi, b)$, and (ii) $\mathscr{A}(A, b+\xi) = \mathscr{A}(A+\xi/(r+\lambda), b)$. Property (i) is immediate and property (ii) follows from inserting the solution for b_t given in footnote 8 into [D.1].

We next present a characterization of the agent's value function up to a scaling parameter that will be pinned down later. First, note that $V(A,b) \leqslant 0$ is well-defined because $u(\cdot) < 0$, and also satisfies $V(A,b) < 0$ because [D.1] renders infeasible those consumption processes that approximate $\hat{c}_t \equiv +\infty$ (which is necessary for $V(A,b) = 0$). Second, $V(A,b) > -\infty$ because, as shown below in the course of proving Proposition 3.1, the consumption process [3.4] is feasible and delivers finite lifetime utility (independently of its asserted optimality).

Lemma D.2. Let $\xi \in \mathbb{R}$. The value function $V(A,b)$ satisfies:

(i) $V(A+\xi, b) = e^{-\theta r \xi} V(A,b)$,

(ii) $V(A, b+\xi) = e^{-f(r;\lambda)\xi} V(A,b)$, and

(iii) $V(A,b) = -\exp\big(-\theta r(A + b/(r+\lambda) + \gamma)\big)$, where $-e^{-\theta r\gamma} := V(0,0)$.

Proof. We consider each point in turn.

Point (i): This follows from observing that, given any consumption strategy $\hat{c}$ that is feasible at (A, b), the strategy given by $\hat{c}_t + r\xi$ for all $t \geqslant 0$ is feasible at $(A + \xi, b)$, for any $\xi \in \mathbb{R}$.

Point (ii): Recall our observation above that $\mathscr{A}(A, b + \xi) = \mathscr{A}(A + \xi/(r + \lambda), b)$. Now, from point (i), it follows that $V(A, b + \xi) = V(A + \xi/(r + \lambda), b) = e^{-r\theta\xi/(r+\lambda)}V(A, b) = e^{-f(r;\lambda)\xi}V(A, b)$, as claimed.

Point (iii): Immediate from points (i) and (ii). □

The function $V(A, b)$ is completely determined once we specify γ. We do this next.

D.2. Towards the Proof of Proposition 3.1

Because the value function $V(A, b) \in \mathbf{C}^\infty(\mathbb{R}^2)$ by Lemma D.2(iii), one can show that it satisfies the HJB equation

[**D.2**] $$\rho V(A, b) = \sup_{c\in\mathbb{R}} \left[(rA + b - c)V_A(A, b) + (\mu - \lambda b)V_b(A, b) + \tfrac{1}{2}\sigma^2 V_{bb}(A, b) + u(c)\right]$$

by adapting standard arguments to the present setting (for instance, see Yong and Zhou 1999, Theorem 3.3 and Touzi 2018, Propositions 2.4-2.5). This allows us to calculate $V(0, 0)$ (equivalently, γ) and the optimal *Markovian* consumption policy $\hat{C}$.

Lemma D.3. The value function $V(A, b)$ and the optimal policy $\hat{c} = (\hat{c}_t)$ are such that:

(i) The parameter γ, where $-e^{-r\theta\gamma} := V(0, 0)$, is given by

[**D.3**] $$\gamma = \frac{\mu}{r(\lambda + r)} + \frac{\log(r)}{r\theta} - \left[\frac{r - \rho}{\theta r^2} + \frac{1}{2}\frac{(f(r;\lambda)\sigma)^2}{\theta r^2}\right]$$

(ii) The supremum in [**D.2**] at state (A, b) is uniquely attained by

[**D.4**] $$\hat{C}(A, b) := rA + \frac{r}{\lambda + r}b + \frac{\mu}{\lambda + r} - \left[\frac{r - \rho}{\theta r} + \frac{1}{2}\frac{(f(r;\lambda)\sigma)^2}{\theta r}\right]$$

(iii) We have $u(\hat{C}(A, b)) = rV(A, b)$.

Proof. When optimizing over c in [**D.2**], notice that the necessary and sufficient condition for optimality in state (A, b) is $u'(c) = V_A(A, b)$. Using the exponential form of u and the form of V derived in Lemma D.2(iii), we conclude that $\theta u(c) = r\theta V(A, b)$, which reduces to

[**D.5**] $$c = rA + \frac{r}{\lambda + r}b + r\gamma - \frac{\log r}{\theta}$$

upon taking logs. Substitution this back into [**D.2**], using the fact (established above) that $\theta u(c) = r\theta V(A, b)$, and solving for γ yields [**D.3**] in point (i) of the lemma. Substituting [**D.3**] into [**D.5**] then yields [**D.4**] in point (ii) of the lemma. For point (iii), we recall that $u(c) = rV(A, b)$ at the optimum, and substituting $\hat{C}$ as in [**D.4**]completes the proof. □

Substituting the consumption policy in [**D.5**] into the ODE for A_t, we obtain the process A^* that satisfies [**D.6**] below.

Lemma D.4. Let A^* be defined to satisfy

[**D.6**] $$\mathrm{d}A_t^* = \left(\frac{r-\rho}{\theta r} + \frac{1}{2}\frac{\sigma^2 f(r;\lambda)^2}{\theta r} + \frac{\lambda b_t - \mu}{\lambda + r} \right) \mathrm{d}t$$

For any $\alpha > 0$, we have $\lim_{t\to\infty} e^{-\alpha t} A_t^* = 0$.

Proof. Writing [**D.6**] in integrated form yields

[**D.7**] $$A_t^* = A_0 + \left(\frac{r-\rho}{\theta r} + \frac{1}{2}\frac{\sigma^2 f(r;\lambda)^2}{\theta r} - \frac{\mu}{\lambda + r} \right) t + \frac{\lambda}{\lambda + r} \int_0^t b_\tau \,\mathrm{d}\tau.$$

When $\lambda = 0$, A^* is deterministic and affine in t, so the claim follows immediately. When $\lambda > 0$, the deterministic terms of $e^{-\alpha t} A_t^*$ again vanish as $t \to \infty$. Thus, it suffices to show that $\lim_{t\to\infty} e^{-\alpha t} \int_0^t b_\tau \,\mathrm{d}\tau = 0$. This is established in Lemma 2.3(iii) in Bloedel, Krishna, and Strulovici (2020), which completes the proof. □

Lemma D.5. Let the function $\hat{C}(A, b)$ be as defined in [**D.4**] and the process A^* be as defined in [**D.6**]. The consumption process $\hat{c}^*$ defined by $\hat{c}_t^* := \hat{C}(A_t^*, b_t)$ satisfies the following properties:

(i) Its induced asset process $A^{\hat{c}^*}$ satisfies $A^{\hat{c}^*} = A^*$.

(ii) It evolves as

[**D.8**] $$\mathrm{d}\hat{c}_t^* = \left(\frac{r - \rho + \sigma^2 f(r;\lambda)^2/2}{\theta} \right) \mathrm{d}t + \frac{\sigma r}{r + \lambda} \mathrm{d}W_t$$

(iii) It is feasible.

Proof. We consider each point in turn.
Point (i): Substituting $\hat{c}_t^* = \hat{C}(A_t^*, b_t)$ in [**3.2**] gives [**D.6**].
Point (ii): From the functional form for $\hat{C}$, $\hat{c}^*$ satisfies the SDE $\mathrm{d}\hat{c}_t^* = r\,\mathrm{d}A_t^* + \frac{r}{(r+\lambda)}\,\mathrm{d}b_t$. Substituting the ODE for A^* given in [**D.6**] and the defining SDE for b given in [**2.1**] into this equation delivers [**D.8**].
Point (iii): Immediate from point (i) of the present lemma, Lemma D.4, and Definition D.1. □

Corollary D.6. Under the consumption policy $\hat{c}^*$ (as described in Lemma D.5), we have $u(\hat{c}_t^*) \equiv rV(A_t^{\hat{c}^*}, b_t)$.

Proof. Follows immediately from Lemmas D.3 and D.5. □

Lemma D.7. Under the consumption policy $\hat{c}^*$ (as described in Lemma D.5), the processes $\left(e^{(r-\rho)t}u'(\hat{c}_t^*)\right)$, $\left(e^{(r-\rho)t}u(\hat{c}_t^*)\right)$, and $\left(e^{(r-\rho)t}V(A_t^*, b_t)\right)$ define martingales.

Proof. Let $Z_t := e^{(r-\rho)t}u(\hat{c}_t^*)$ and recall that $u(c) = -e^{-\theta c}$. Then:

$$\begin{aligned} Z_t &= -e^{(r-\rho)t} \exp\left((\rho - r)t - \tfrac{1}{2}\sigma^2(f(r;\lambda))^2 - \sigma f(r;\lambda)W_t - \theta\hat{c}_0^*\right) \\ &= -e^{-\theta\hat{c}_0^*} \exp\left(-\tfrac{1}{2}\sigma^2(f(r;\lambda))^2 - \sigma f(r;\lambda)W_t\right) \end{aligned}$$

Noting that $\exp\left(-\frac{1}{2}\sigma^2(f(r;\lambda))^2 - \sigma f(r;\lambda)W_t\right)$ is an exponential Brownian martingale, a straightforward calculation yields $\mathbf{E}[Z_\tau \mid \mathscr{F}_t] = Z_t$ for all $\tau > t$, as claimed. The martingale property of $e^{(r-\rho)t}u'(\hat{c}_t^*)$ then follows from the fact that $u'(c) = -\theta u(c)$. The martingale property of $\left(e^{(r-\rho)t}V(A_t^*, b_t)\right)$ follows from Corollary D.6. □

D.3. Proof of Proposition 3.1

We present one final technical lemma before proceeding to the proof of the proposition.

Lemma D.8. Given the asset process A^* in [D.6], the process $M_t := \int_0^t e^{-\rho s}V_b(A_s^*, b_s)\sigma \, \mathrm{d}W_s$ defines a martingale.

Proof. It suffices to establish that $\mathbf{E}\left[\int_0^T \left(e^{-\rho t}V_b(A_t^*, b_t)\sigma\right)^2 \mathrm{d}t\right] < \infty$ for all $T > 0$. Using the functional form of V established in Lemma D.2(iii), we note that $V_b(A_t^*, b_t) \equiv \frac{(-r\theta)}{r+\lambda}V(A_t^*, b_t)$. By Lemma D.3(i), we know that $u(\hat{c}_t^*) = rV(A_t^*, b_t)$. An application of Fubini's Theorem indicates it suffices to show that $\int_0^T e^{-2\rho t}\,\mathbf{E}\left[u(\hat{c}_t^*)^2\right] \mathrm{d}t < \infty$.

By Lemma D.5(ii), $\hat{c}_t^* = \hat{c}_0^* + \left(\frac{r-\rho+\sigma^2 f(r;\lambda)^2/2}{\theta}\right)t + \frac{\sigma f(r;\lambda)}{\theta}W_t$. Therefore, $u(\hat{c}_t^*)^2 = \exp(-2\theta\hat{c}_t^*)$, which in turn is equal to

$$\exp(-2\theta\hat{c}_0^*)\exp(-2(r-\rho)t + \sigma^2 f^2(r;\lambda)t)\exp(-2\sigma^2 f^2(r;\lambda)t - 2\sigma f(r;\lambda)W_t)$$

But $\exp(-2\sigma^2 f^2(r;\lambda)t - 2\sigma f(r;\lambda)W_t)$ is an exponential martingale with $\mathbf{E}[\exp(-2\sigma^2 f^2(r;\lambda)t - 2\sigma f(r;\lambda)W_t)] = 1$. Therefore,

$$\begin{aligned} e^{-2\rho t}\,\mathbf{E}\left[u(\hat{c}_t^*)^2\right] &= e^{-2\rho t}\exp(-2\theta\hat{c}_0^*)\exp(-2(r-\rho)t + \sigma^2 f^2(r;\lambda)t) \\ &= \exp(-2\theta\hat{c}_0^*)\exp(-2rt + \sigma^2 f^2(r;\lambda)t) \end{aligned}$$

Straightforward integration then yields that $\int_0^T e^{-2\rho t}\,\mathbf{E}\left[u(\hat{c}_t^*)^2\right] \mathrm{d}t < \infty$, which establishes the claim. □

We now are in a position to prove the proposition.

Proof of Proposition 3.1. It suffices to verify that the consumption process $\hat{c}^*$ defined in Lemma D.5 attains lifetime utility $V(A_0, b_0)$ when starting from initial condition (A_0, b_0). Consider the Itô expansion of $e^{-\rho t}V(A_t^*, b_t)$, whereby we have

[**D.9**]
$$e^{-\rho T}V(A_t^*, b_t) = V(A_0, b_0) + \int_0^T e^{-\rho t}\big[\mathscr{L}^{\hat{c}^*}V(A_t^*, b_t) - \rho V(A_t^*, b_t)\big]\,\mathrm{d}t + \int_0^T e^{-\rho t}V_b(A_t^*, b_t)\sigma\,\mathrm{d}W_t$$

where $\mathscr{L}^{\hat{c}^*}$ is the differential operator defined as $\mathscr{L}^{\hat{c}^*}v(A, b) = (rA + b - \hat{c}^*)\partial_A v(A, b) + \lambda(\mu/\lambda - b)\partial_b v(A, b) + \frac{1}{2}\sigma^2\partial_{bb}v(A, b)$ for any function $v \in \mathbf{C}^2(\mathbb{R}^2)$. By Lemma D.3(ii), $\hat{c}^*$ achieves the supremum in [**D.2**] at each state, i.e., $\mathscr{L}^{\hat{c}^*}V(A_t^*, b_t) - \rho V(A_t^*, b_t) = u(\hat{c}_t^*)$. Substituting this into [**D.9**], taking expectations on both sides of [**D.9**], and applying Lemma D.8 to show that the stochastic integral in [**D.9**] is a true (not just local) martingale, we obtain

[**D.10**]
$$V(A_0, b_0) = \mathbf{E}\left[\int_0^T e^{-\rho t}u(\hat{c}_t^*)\,\mathrm{d}t\right] + \mathbf{E}\left[e^{-\rho T}V(A_t^*, b_t)\right]$$

Now, Lemma D.7 shows that $e^{(r-\rho)t}V(A_t^*, b_t)$ is a martingale. Therefore,

$$\mathbf{E}\left[e^{-\rho T}V(A_t^*, b_t)\right] = e^{-rT}\,\mathbf{E}\left[e^{(r-\rho)T}V(A_t^*, b_t)\right] = e^{-rT}V(A_0, b_0)$$

which implies that $\lim_{T\to\infty}\mathbf{E}\left[e^{-\rho T}V(A_t^*, b_t)\right] = 0$. Thus, letting $T \to \infty$ in [**D.10**] gives us

$$V(A_0, b_0) = \mathbf{E}\left[\int_0^\infty e^{-\rho t}u(\hat{c}_t)\,\mathrm{d}t\right]$$

meaning that $\hat{c}^*$ achieves the supremum in [**3.1**], as desired. □

Remark 5. By adapting the argument in the above proof, it can be shown that $\hat{c}^*$ (as defined in Lemma D.5) is *uniquely* optimal among all feasible consumption processes $\hat{c}$ for which (i) the process $M_t := \int_0^t e^{-\rho s}V_b(A_s^{\hat{c}}, b_s)\sigma\,\mathrm{d}W_s$ defines a martingale (cf. Lemma D.8) and (ii) $\lim_{T\to\infty}\mathbf{E}\left[e^{-\rho T}V(A_t^{\hat{c}}, b_t)\right] = 0$.

E. Details for the Agent's Reporting Problem

We work on the canonical filtered probability space for processes with continuous paths, denoted $\big(\mathbf{C}[0,\infty), \mathscr{F}, \mathbf{P}^W, \mathbb{F}\big)$. Here $\mathbf{C}[0,\infty)$ is the space of real-valued continuous functions $\omega : [0,\infty) \to \mathbb{R}$ endowed with the topology of compact convergence, $\mathbf{P}^W$ is the

Wiener measure under which the coordinate process $t \mapsto W_t(\omega) := \omega_t$ defines a standard Brownian motion, $\mathscr{F}$ is the P^W-augmentation of the Borel σ-algebra on $\mathbf{C}[0,\infty)$, and $\mathbb{F}$ is the (suitably augmented) filtration generated by $W = (W_t)_{t\geqslant 0}$. See Karatzas and Shreve 1998, Ch. 2.4 for details. Let P denote the measure over paths of the endowment process b defined by [2.1].

As in the main text, the agent chooses a b-adapted *misreporting strategy* m, which induces a b-adapted *reporting strategy* y defined by $y_t := b_t + m_t$. Under the truthful strategy $m^* \equiv 0$, the distribution over paths of y is denoted by P^*, which clearly coincides with P. (We distinguish between these measures to make clear whether the underlying process is y or b.)

Definition E.1. Misreporting strategy m is *admissible*, denoted $m \in \mathscr{M}$, if:

(i) It is b-adapted.

(ii) It admits the representation $m_t \equiv \int_0^t \Delta_\tau d\tau$ for some process Δ.

(iii) The induced *density process* Γ^m defined by

$$\Gamma_t^m := \exp\left[\int_0^t \left(\frac{\Delta_\tau + \lambda m_\tau}{\sigma}\right) \mathrm{d}W_\tau - \frac{1}{2}\int_0^t \left(\frac{\Delta_\tau + \lambda m_\tau}{\sigma}\right)^2 \mathrm{d}\tau\right] \tag{E.1}$$

is a well-defined[36] uniformly integrable (UI) martingale under P^*.

Remark 6. We present some concrete examples of admissible strategies:

- *Example 1:* The truthful strategy $m^* \equiv 0$ is admissible, for $\Gamma^{m^*} \equiv 1$ is clearly a UI martingale.
- *Example 2:* Fix $M > 0$ and $T > 0$. Any strategy m for which $|\Delta_t| \leqslant M$ for all $t \in [0,T]$ and $m_t = 0$ for all $t \geqslant T$ is admissible. To see this, note that, because $m_t = 0$ for all $t \geqslant T$, we have that Γ^m is a P^*-UI martingale if and only if $\mathsf{E}^*[\Gamma_T^m] = 1$. But it is easy to see that $(\Gamma_t^m)_{t\in[0,T]}$ satisfies Novikov's condition, so the latter property holds. (In discrete time, the analogous class of strategies would be those for which the agent lies only in a finite number of periods. The Lipschitz condition $|\Delta_t| \leqslant M$ is imposed here for technical reasons related to the continuous-time formulation.)
- *Example 3:* Let $\lambda = 0$. Fix $M > 0$ and $T > 0$. Any strategy m for which $|\Delta_t| \leqslant M$ for all $t \in [0,T]$ and $\Delta_t = 0$ for all $t > T$ is admissible. Note that m satisfies $m_t = m_T$ for all $t \geqslant T$ but, unlike in Example 2, m_t may be nonzero for $t > T$. But since m does not appear directly in the density process [E.1] when $\lambda = 0$, the proof of admissibility is analogous to that in Example 2. This class of strategies will be important for the proof of Theorem 6 in Online Appendix J.

(36) Namely, this means that the stochastic integral in [E.1] is well-defined.

It is a standard observation that Γ^m defined as in [E.1] is a local martingale under $\mathbf{P}^*$. Given points (i) and (ii) of the above definition, the assumption in point (iii) that it is a UI (true) martingale is necessary and sufficient for us to view the agent as choosing a probability measure over paths of y that is absolutely continuous with respect to $\mathbf{P}^*$. More formally, by Girsanov's Theorem, every $m \in \mathscr{M}$ induces a probability measure $\mathbf{P}^m$ on $(\mathbf{C}[0,\infty), \mathscr{F})$ defined by the Radon-Nikodym derivative $\frac{\mathrm{d}\,\mathbf{P}^m}{\mathrm{d}\,\mathbf{P}^*} = \Gamma^m_\infty$, where $\lim_{t\to\infty} \Gamma^m_t = \Gamma^m_\infty$ $\mathbf{P}^*$-almost surely. (The random variable Γ^m_∞ is well-defined by the assumption that Γ^m is a UI martingale.) In addition, under $\mathbf{P}^m$, the b-adapted process W^m defined by

[E.2] $$\mathrm{d}W^m_t := \mathrm{d}W_t - \frac{(\lambda m_t + \Delta_t)}{\sigma}\,\mathrm{d}t$$

is a standard Brownian motion, and the report process y evolves as

[E.3] $$\mathrm{d}y_t = (\mu - \lambda(y_t - m_t) + \Delta_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W^m_t$$

with the same initial condition $y_0 = b_0$.

F. Details for Direct-Revelation Self-Insurance Contracts

In this online appendix, we provide details supporting the discussion of Direct-Revelation DR-SICs from Subsection 3.3. Recall the definition of the consumption function $\hat{C}$ from [3.5].

Formal Definition of the Direct Mechanisms. The following definition formalizes the intuitive description given in Subsection 3.3.

Definition F.1. Given the Self-Insurance Contract (b_0, q_0, r), define the corresponding *Direct-Revelation Self-Insurance Contract*, denoted by DR-SIC (b_0, q_0, r), as the following direct revelation mechanism:

(i) The principal keeps track of the y-adapted *virtual asset* process A^v defined by

$$A^v_t := A_0(b_0, q_0, r) + \left[\frac{r - \rho + \sigma^2 f(r;\lambda)^2/2}{r\theta} - \frac{\mu}{r+\lambda}\right] t + \frac{\lambda}{r+\lambda}\int_0^t y_\tau\,\mathrm{d}\tau$$

(ii) The agent's *recommended consumption* is the y-adapted process c defined by

$$c_t := \hat{C}\left(A^v_t, y_t\right) = rA^v_t + \frac{r}{\lambda + r} y_t - \left[\frac{r - \rho + \sigma^2 f(r;\lambda)^2/2}{r\theta} - \frac{\mu}{r+\lambda}\right]$$

(iii) The agent is allowed to choose any admissible misreporting strategy (in the sense of Definition E.1), and consumes according to $c^m_t := c_t - m_t$.

Comparing Deviations in the Direct and Indirect Mechanisms. Let $\hat{c}^*$ denote the agent's optimal consumption strategy in the Self-Insurance Contract (b_0, q_0, r) and let A^* denote the induced asset process, so that $\hat{c}^* = \hat{C}(A^*, b_t)$ (see Proposition 3.1). Notice that the agent's "virtual savings" in the corresponding DR-SIC (b_0, q_0, r) satisfy

$$\frac{\mathrm{d}A_t^v}{\mathrm{d}t} = \frac{\mathrm{d}A_t^*}{\mathrm{d}t} + \frac{\lambda}{r+\lambda} m_t$$

so that under-reporting by setting $m_t < 0$ in the direct mechanism corresponds to under-saving by $\frac{\lambda}{r+\lambda} m_t < 0$ in the self-insurance problem. The following lemma characterizes how misreports in the direct mechanism correspond to consumption choices in the agent's self-insurance problem.

Lemma F.2. Given any b-adapted misreporting strategy m in the DR-SIC (b_0, q_0, r), the agent's actual consumption $c_t^m := c_t - m_t$ satisfies

$$c_t^m = \hat{C}(A_t^v, b_t) - \frac{\lambda}{r+\lambda} m_t \tag{F.1}$$

$$= \hat{c}_t^* + \frac{\lambda}{r+\lambda}\left(r\int_0^t m_\tau \,\mathrm{d}\tau - m_t\right) \tag{F.2}$$

Proof. By direct calculation, using $y_t = b_t + m_t$, we see that

$$\begin{aligned} c_t^m = \hat{C}\left(A_t^v, y_t\right) - m_t &= \hat{C}\left(A_t^v, b_t\right) + \frac{r}{r+\lambda} m_t - m_t \\ &= \hat{C}\left(A_t^v, b_t\right) - \frac{\lambda}{r+\lambda} m_t \end{aligned}$$

which yields [F.1]. Similarly, we have

$$\begin{aligned} c_t^m = \hat{C}\left(A_t^v, y_t\right) - m_t = &\left[\frac{r-\rho+\sigma^2 f(r;\lambda)^2/2}{\theta} - \frac{r\mu}{r+\lambda}\right] t + \frac{r\lambda}{r+\lambda}\int_0^t b_\tau \,\mathrm{d}\tau + \frac{r}{\lambda+r} b_t \\ &+ rA_0(b_0, q_0, r) + \frac{\mu}{\lambda+r} - \left[\frac{r-\rho}{\theta r} + \frac{1}{2}\frac{\left(f(r;\lambda)\sigma\right)^2}{\theta r}\right] \\ &+ \frac{r\lambda}{r+\lambda}\int_0^t m_\tau \,\mathrm{d}\tau + \frac{r}{\lambda+r} m_t - m_t \\ = &\,\hat{C}\left(A_t^*, b_t\right) + \frac{\lambda}{r+\lambda}\left(r\int_0^t m_\tau \,\mathrm{d}\tau - m_t\right) \end{aligned}$$

which establishes [F.2] because $\hat{c}_t^* = \hat{C}\big(\hat{A}_t^*, b_t\big)$. □

Lemma F.2 states that under-reporting by setting $m_t < 0$ in the direct mechanism corresponds to over-consuming by $-\frac{\lambda}{r+\lambda} m_t > 0$ in the self-insurance problem, *holding* *fixed the agent's assets at* A_t^v (display [F.1]). However, comparing the agent's consumption under such a deviation in the direct mechanism to $\hat{c}_t^*$ requires information about the entire history of misreports, and may be either positive or negative depending on that history (display [F.2]).

Corollary F.3. Fix a DR-SIC (b_0, q_0, r). Given any admissible misreporting strategy m, the mapping $t \mapsto c_t^m - \hat{c}_t^*$ is absolutely continuous.

Proof. Immediate from Lemma F.2 and point (i) in the definition of admissibility (Definition E.1). □

F.1. Proof of Theorem 1

Lemma F.4. Given any $r > 0$ and admissible m, we have
(i) $\mathbf{P}^m\left(\lim_{t\to\infty} e^{-rt} x_t = 0\right) = 1$.
(ii) $\mathbf{P}^m\left(\lim_{t\to\infty} e^{-rt} \int_0^t x_\tau \, \mathrm{d}\tau = 0\right) = 1$.
for all processes $x \in \{y, m\}$.

Proof. Note that y satisfies $\mathbf{P}^*\left(\lim_{t\to\infty} e^{-rt} y_t = 0\right) = \mathbf{P}^*\left(\lim_{t\to\infty} e^{-rt} \int_0^t y_\tau \, \mathrm{d}\tau = 0\right) = 1$ because it coincides with b and therefore solves [2.1] under truthful reporting, and b satisfies these properties by Lemma 2.3 in Bloedel, Krishna, and Strulovici (2020). Then y also satisfies these properties under $\mathbf{P}^m$ because $\mathbf{P}^m \ll \mathbf{P}^*$ by admissibility (by arguments in Online Appendix E). That m also satisfies the stated properties under $\mathbf{P}^m$ then follows from the fact that $y_t = b_t + m_t$ by definition. □

Lemma F.5. Fix a DR-SIC (b_0, q_0, r). Given any admissible m, the virtual asset process A^v has the following properties:
(i) It solves $\mathrm{d}A_t^v = (rA_t^v + b_t - c_t^m)\,\mathrm{d}t$ with $A_0^v = A_0(b_0, q_0, r)$.
(ii) It satisfies the no-Ponzi condition $\mathbf{P}^m\left(\lim_{t\to\infty} e^{-rt} A_t^v = 0\right) = 1$.

Proof. It is immediate from parts (ii)-(iii) of Definition F.1 that A^v solves $\mathrm{d}A_t^v = (rA_t^v + y_t - c_t)\,\mathrm{d}t$. Plugging in $y_t := b_t + m_t$ and $c_t^m := c_t - m_t$ then delivers point (i) of the lemma. Point (ii) is immediate from Lemma F.4(ii) and parts (i)-(ii) of Definition F.1. □

Proof of Theorem 1. Fix a DR-SIC (b_0, q_0, r) and an admissible misreporting strategy m. By Lemma F.5, the b-adapted consumption strategy c^m would be feasible in the corresponding Self-Insurance Contract (b_0, q_0, r), in which it is (weakly) dominated by the optimal consumption strategy $\hat{c}^*$.[37] But truthful reporting $m^* \equiv 0$ by construction induces $c^{m^*} = \hat{c}^*$, so truthful reporting (weakly) dominates m. It follows that m^* is optimal for the agent in the DR-SIC (b_0, q_0, r). □

F.2. Enlarged Space of Reporting Strategies

We now explain why the above analysis shows that a DR-SIC remains incentive compatible even when we allow the agent to choose some non-admissible reporting strategies, such as those including jumps. Fix a DR-SIC (b_0, q_0, r). Define the y-adapted process q by $q_t := V(A_t^v, y_t)$, where V is the agent's value function in the DR-SIC (b_0, q_0, r) characterized in Lemma D.2. This is simply the agent's promised utility in the direct mechanism, as in [2.2]. A simple calculation shows that $\hat{C}(A_t^v, y_t) = \overline{c}(q_t, r) := -\log(-rq_t)/\theta$. It follows that the DR-SIC can be written recursively using (y_t, q_t) as state variables, as in Subsection 4.1.

By the proof of Theorem 1, the agent can induce strictly more consumption processes in the Self-Insurance Contract (b_0, q_0, r) than in the DR-SIC (b_0, q_0, r), even though the agent's optimal consumption process is the same in both. Consider the following extension of the DR-SIC that closes this gap in feasible strategies.

Definition F.6. The *extended DR-SIC* (b_0, q_0, r) is defined by points (ii)-(iii) of Definition F.1, and with point (i) thereof replaced by:

(i') The agent is allowed to choose, and consume according to, any b-adapted misreporting strategy m such that c^m is feasible in the DR-SIC (b_0, q_0, r).

Fix (b_0, q_0, r). Because the agent's feasible set in the DR-SIC allows for consumption processes with jumps (among other features), it is easy to see that the agent is allowed to choose reporting strategies with jumps in the extended DR-SIC. Thus, the misreporting process m need not have absolutely continuous paths, so the process Δ may not be well-defined. However, the agent is still restricted to those m for which $\lim_{t\to\infty} e^{-rt}\int_0^t m_\tau \, d\tau \geqslant 0$ almost surely, for otherwise the induced virtual asset process would violate the no-Ponzi condition.

It is useful to characterize the agent's value function and optimal reporting strategy in the DR-SIC. We proceed somewhat informally. By Definition F.6, the value function must be isomorphic to the agent's value function in the corresponding Self-Insurance Contract (which is given in Online Appendix D). When the extended DR-SIC is in state

(37) More precisely, an immediate corollary of Lemma F.5(ii) is that $\mathbf{P}((\lim_{t\to\infty} e^{-rt}(A^v \circ y)_t = 0) = 1$, where $A^v \circ y$ is the b-adapted process achieved by composing the agent's b-adapted reports y with the y-adapted virtual asset process A^v.

(y_t, q_t), an agent who was been truthfully reporting on $[0, t)$ (i.e, for whom $b^\tau = y_\tau$ for $\tau \leqslant t$) is promised continuation utility $q_t = V(A_t^v, y_t)$. But if the agent's actual endowment is $b_t = y_t - m_t$, then by definition his optimal strategy in the Self-Insurance Contract yields him continuation value $V(A_t^v, y_t - m_t) = V(A_t^v, y_t - m_t) \exp(f(r; \lambda) m_t)$ therein. Therefore, it is apparent that the agent's value function in the *e*xtended DR-SIC can be written as a function of (q, m) as $V^e(q, m) := q \exp(f(r; \lambda) m)$. To derive the agent's best-response, assume for simplicity that the agent is restricted to y with RCLL paths. For an RCLL process x, let $x_{t-} := \lim_{\varepsilon \to 0} x_{t-\varepsilon}$ as usual. Fix a state (q_{t-}, m_{t-}) and prior report y_{t-}. Suppose the true endowment is b_t (which $= b_{t-}$ by continuity). The left-limit of the agent's continuation value in the extended DR-SIC is $q_{t-} = V^e(q_{t-}, y_{t-}) = V^e(q_{t-}, 0) \exp(f(r; \lambda) m_{t-})$. However, by definition, his optimal continuation value in the Self-Insurance Contract is $V(A_{t-}^v, b_t) = V^e(q_{t-}, 0)$; by definition, this must also be his optimal continuation value in the extended DR-SIC. To achieve this optimal value, the agent may report a jump of size $-m_{t-}$, which brings his report y_{t-} back to the truth b_t. This induces promised utility $q_t = q_{t-} + q_{t-} \cdot (\exp(-f(r; \lambda) m_{t-}) - 1) = V^e(q_{t-}, 0)$. In words, the agent instantly reverts to his optimal consumption plan in the extended DR-SIC. Without a jump in reports, the contract and his past lies specify that he should consume $\hat{C}(A_{t-}^v, y_{t-}) - m_{t-}$. But it is optimal to consume $\hat{C}(A_{t-}^v, y_{t-} - m_{t-})$, where $y_{t-} - m_{t-} = b_t$. So the jump in the report corresponds to a jump of size $\lambda m_{t-}/(\lambda + r)$ in his consumption, which instantly "re-initializes" his consumption path at the optimal level.

G. Proof of Theorem 2

Recall the principal's problem of finding the cost-minimizing shadow rate r, [B.1] in Appendix B. When $\lambda > 0$, in which case $f(\cdot; \lambda)$ is strictly increasing, it will be convenient to define the inverse function $r(k; \lambda) := f^{-1}(k; \lambda) = \frac{\lambda k}{\theta - k}$ and rewrite [B.1] as

$$\text{[G.1]} \quad \inf_{k \in (0, \theta)} \left[-\frac{\log(r(k; \lambda))}{\theta \rho} + \frac{r(k; \lambda) - \rho + \sigma^2 k^2/2}{\theta \rho^2} - \frac{\log(-q_0)}{\theta \rho} - \frac{b_0}{\rho + \lambda} - \frac{\mu}{\rho(\rho + \lambda)} \right]$$

Existence: Let (b_0, q_0) be given. It is straightforward to verify that $\lim_{r \to 0} \Pi(b_0, q_0, r) = \lim_{r \to \infty} \Pi(b_0, q_0, r) = +\infty$ and that $\Pi(b_0, q_0, \cdot)$ is continuous on $\mathbb{R}_{++}$. Thus, a solution r^* to problem [B.1] exists, and every solution r^* thereof satisfies $r^* > 0$ and the necessary first-order condition [3.16].

Point (i): Immediate from the previous paragraph and [3.16].

Point (ii): Let $\sigma > 0$ be given. Let $r^*(\cdot, \sigma)$ and $k^*(\cdot, \sigma)$ denote corresponding selections from the argmin correspondences of [B.1] and [G.1], respectively. We first show that $k^*(\cdot, \sigma)$ is strictly decreasing. It is easy to verify that $\frac{\partial^2}{\partial r \partial \lambda} r(k; \lambda) > 0$, so an application of of Edlin and Shannon (1998, Theorem 1) (adapted to minimization problems) to [G.1] implies that $k^*(\cdot;\sigma)$ is strictly decreasing.

Since point (i) of the theorem established that $r^*(\cdot,\sigma) < \rho$ for all $\lambda > 0$, the non-monotonicity of this function will follow from establishing its limiting behavior.

Let $(\lambda,\sigma) \mapsto \hat{r}(\lambda,\sigma)$ denote any selection of (non-negative real) solutions to the first-order condition [3.16]. By the same argument showing that $r^*(\lambda,\sigma) \leqslant \rho$, we have that $\hat{r}(\lambda,\sigma) \leqslant \rho$ for all $(\lambda,\sigma) \in \mathbb{R}_+ \times \mathbb{R}_{++}$. Because $\frac{\mathrm{d}}{\mathrm{d}r}\log(r) = 1/r \to \infty$ as $r \to 0$, it is also easy to see that $\hat{r}(\lambda,\sigma) > 0$ for all $(\lambda,\sigma) \in \mathbb{R}_+ \times \mathbb{R}_{++}$. These observations are useful in what follows.

To establish $\lim_{\lambda\to\infty} r^*(\lambda,\sigma) = \rho$, it suffices to show that $\lim_{\lambda\to\infty} \hat{r}(\lambda,\sigma) = \rho$. Multiplying the first-order condition [3.16] through by $\theta\rho\hat{r}(\lambda,\sigma) > 0$ and rearranging yields the equivalent condition

[G.2] $$\rho = \hat{r}(\lambda,\sigma) + \sigma^2 \cdot \frac{\lambda\theta^2\hat{r}(\lambda,\sigma)^2}{(\hat{r}(\lambda,\sigma)+\lambda)^3}$$

Because $\hat{r}(\cdot,\sigma)$ is bounded, we have

$$\lim_{\lambda\to\infty} \frac{\lambda\theta^2\hat{r}(\lambda,\sigma)^2}{(\hat{r}(\lambda,\sigma)+\lambda)^3} = \lim_{\lambda\to\infty} \frac{1}{\lambda^2} = 0$$

given which [G.2] implies that $\lim_{\lambda\to\infty} \hat{r}(\lambda,\sigma) = \rho$, as desired.

Finally, to establish $\lim_{\lambda\to 0} r^*(\lambda,\sigma) = \rho$, it suffices to show that $\underline{r} := \liminf_{\lambda\to 0} r^*(\lambda,\sigma) > 0$, given which we have

$$0 \leqslant \liminf_{\lambda\to 0} \frac{\lambda\theta^2 r^*(\lambda,\sigma)^2}{(r^*(\lambda,\sigma)+\lambda)^3} \leqslant \limsup_{\lambda\to 0} \frac{\lambda\theta^2 r^*(\lambda,\sigma)^2}{(r^*(\lambda,\sigma)+\lambda)^3} \leqslant \limsup_{\lambda\to 0} \frac{\lambda\theta^2 r^*(\lambda,\sigma)^2}{\underline{r}^3} = 0$$

because $r^*(\cdot,\sigma) \leqslant \rho$. This then implies via [G.2] (with $\hat{r} = r^*$) that $\lim_{\lambda\to 0} r^*(\lambda,\sigma) = \rho$, as desired. To complete the proof, note that $\underline{r} > 0$ follows from optimality. Suppose towards a contradiction that there exists a sequence $(\lambda_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty}\lambda_n = 0$ and $r_n := r^*(\lambda_n,\sigma)$ with $\lim_{n\to\infty} r_n = 0$. The objective in [B.1] satisfies

$$\begin{aligned}
&\liminf_{n\to\infty}\left[-\frac{\log(r_n)}{\theta\rho} + \frac{r_n + \sigma^2 f(r_n;\lambda_n)^2/2}{\theta\rho^2} - \frac{\log(-q_0)}{\theta\rho} - \frac{b_0}{\rho+\lambda_n} - \frac{\mu}{\rho(\rho+\lambda_n)}\right] \\
&= -\lim_{n\to\infty}\frac{\log(r_n)}{\theta\rho} + \underbrace{\liminf_{n\to\infty}\frac{r_n + \sigma^2 f(r_n;\lambda_n)^2/2}{\theta\rho^2}}_{\geqslant 0} - \frac{\log(-q_0)}{\theta\rho} - \frac{b_0}{\rho} - \frac{\mu}{\rho^2} \\
&= +\infty
\end{aligned}$$

Thus, for all n sufficiently large, r_n is strictly dominated by $r_n' = \rho$ (the cost of which remains bounded as $n \to \infty$). This establishes the desired contradiction.

Point (iii): Let $\lambda > 0$. Let $r^*(\lambda, \cdot)$ and $k^*(\lambda, \cdot)$ denote corresponding selections from the argmin correspondences of [B.1] and [G.1], respectively. We first show that the optimizers are decreasing in σ. It is easy to verify that $\frac{\partial^2}{\partial r \partial \sigma}[\sigma^2 f(r;\lambda)] > 0$. Consequently, an application of Edlin and Shannon (1998, Theorem 1) (adapted to minimization problems) to [B.1] implies that $r^*(\lambda, \cdot)$ is strictly decreasing, from which it follows that $k^*(\lambda, \cdot)$ is also strictly decreasing.

We now show that $\lim_{\sigma\to\infty} k^*(\lambda, \sigma) = \lim_{\sigma\to\infty} r^*(\lambda, \sigma) = 0$. Let $\hat{r}(\lambda, \cdot)$ be defined as in the proof of point (ii) above. It suffices to show that $\lim_{\sigma\to\infty} \hat{r}(\lambda, \sigma) = 0$. In order for [G.2] to hold for all $\sigma > 0$ it must be that

$$\lim_{\sigma\to\infty} \frac{\hat{r}(\lambda,\sigma)^2}{(\hat{r}(\lambda,\sigma)+\lambda)^3} = 0$$

because $\hat{r}(\lambda, \cdot) > 0$. Since $\hat{r}(\lambda, \cdot) < \rho$, this implies that $\lim_{\sigma\to\infty} \hat{r}(\lambda, \sigma) = 0$.

Finally, we show that $\lim_{\sigma\to 0} r^*(\lambda, \sigma) = \rho$ and thus $\lim_{\sigma\to 0} k^*(\lambda, \sigma) = k^*_\circ$. Again, it suffices to show that $\lim_{\sigma\to\infty} \hat{r}(\lambda, \sigma) = \rho$. Because $\hat{r}(\lambda, \cdot) \in (0, \rho)$, the term multiplying σ^2 in [G.2] is uniformly bounded as a function of σ. Thus, [G.2] immediately implies that $\lim_{\sigma\to\infty} \hat{r}(\lambda, \sigma) = \rho$, as desired.

H. Implementable FO-IC Contracts

Lemma H.1. For each $q_0 < 0$, the following hold:

(i) If $\lambda > 0$, then under any FO-IC contract $k_t \in (0, \theta)$ for all $t \geqslant 0$. Conversely, for each $k_\circ \in (0, \theta)$ there exists an IC contract for which $k_t \equiv k_\circ$.

(ii) If $\lambda = 0$, then under any FO-IC contract $k_t = \theta$ for all $t \geqslant 0$. Conversely, there exists an IC contract inducing $k_t = \theta$ for all $t \geqslant 0$.

Proof. Point (i): Let $\lambda > 0$. The converse direction of the claim follows from the SDE [3.13] for promised utility under Self-Insurance Contracts, and the fact that the range of $f(\cdot;\lambda)$ is $(0, \theta)$. We prove the necessity of $k_t \in (0, \theta)$. Fix an FO-IC contract inducing flow utility u_t, promised utility q_t, and marginal promised utility p_t. Since $\lambda > 0$, we have $e^{-\lambda t} < 1$ for all $t > 0$. Because $u(\cdot) < 0$, the flow utility process satisfies $u_t < e^{-\lambda t} u_t < 0$ for all $t > 0$. By [2.2] and [4.1], it follows that

$$p_t = \mathbf{E}_t^* \left[\int_t^\infty e^{-(\rho+\lambda)(\tau-t)} \theta u_\tau \, \mathrm{d}\tau \right] > \mathbf{E}_t^* \left[\int_t^\infty e^{-\rho(\tau-t)} \theta u_\tau \, \mathrm{d}\tau \right] = \theta q_t$$

Dividing through by $q_t < 0$ yields $k_t = p_t/q_t < \theta$, while $k_t > 0$ follows from the fact that $q_t, p_t < 0$.

Point (ii): Let $\lambda = 0$. The converse direction of the claim follows from the existence of Contract W. We prove the necessity of $k_t \equiv \theta$. The proof mirrors that of point (i) above. Fix an FO-IC contract inducing flow utility u_t, promised utility q_t, and marginal promised utility p_t. Since $\lambda = 0$, we have $e^{-\lambda t} \equiv 1$. Thus, the flow utility process satisfies $u_t = e^{-\lambda t} u_t$. By [2.2] and [4.1], it follows that

$$p_t = \mathbf{E}_t^* \left[\int_t^\infty e^{-(\rho+\lambda)(\tau-t)} \theta u_\tau \, \mathrm{d}\tau \right] = \mathbf{E}_t^* \left[\int_t^\infty e^{-\rho(\tau-t)} \theta u_\tau \, \mathrm{d}\tau \right] = \theta q_t$$

Dividing through by $q_t < 0$ yields $k_t = p_t / q_t = \theta$. □

Point (i) of Lemma H.1, which does not appear in W11, allows us to define the *domain* D of implementable (q, p) pairs by

$$D := \begin{cases} \left\{ (q,p) \in \mathbb{R}^2_{--} : p/q \in (0, \theta) \right\}, & \text{if } \lambda > 0 \\ \left\{ (q,p) \in \mathbb{R}^2_{--} : p = \theta q \right\}, & \text{if } \lambda = 0. \end{cases}$$

I. Material for Subsection 5.1

I.1. Formal Description of Hidden Savings Model

The model builds on that described in Subsection 2.1. We continue to interpret $\rho > 0$ as the risk-free market short rate. However, we now assume that the agent also has direct access to the bond market, and that his trading activity therein is unobservable to the principal. Thus, in a direct-revelation contract the agent chooses both a reporting strategy y and a consumption strategy $\hat{c}$. The agent's trading activity is analogous to that in the self-insurance problem described in Subsection 3.1, except that the agent's "effective endowment" is now $b_t + s_t$ (and is partially controlled by his reporting strategy). He starts with zero assets $A_0 = 0$ but controls their evolution

$$\mathrm{d}A_t = (\rho A_t + b_t + s_t - \hat{c}_t)\,\mathrm{d}t$$

directly via his consumption strategy $\hat{c}$ and indirectly via his misreporting strategy m, which determines the distribution of transfers s. The asset process A is required to satisfy a no-Ponzi condition analogous to [3.3] (with $r = \rho$) almost surely under $\mathbf{P}^m$. Both his endowment b and assets A are private information.

A contract is *NS-IC* if it is optimal for the agent to report truthfully (i.e., choose m^* among all admissible misreporting strategies) and follow the contract's recommended consumption (i.e., set $\hat{c}_t \equiv s_t + b_t$ among all feasible consumption strategies, as defined above). A contract is *first-order NS-IC* if (i) it is optimal for the agent to follow the contract's recommended consumption *conditional on following the truthful reporting strategy*, and (ii) the contract satisfies [**FO-IC**].

I.2. Proof of Theorem 4

Point (i): It was shown in Subsections 3.2–3.3 that Contract W can be implemented as a Self-Insurance Contract with shadow rate $r = \rho$, i.e., a zero tax rate. Although in the hidden savings model the agent also has the option to save directly via the market instead of via the principal, he is indifferent between these options because he faces the same interest rate ρ in either case. By letting the agent break this tie in favor of saving only via the principal, we see that Contract W is still implementable as a Self-Insurance Contract, witnessing the fact that it is NS-IC.

Point (ii): Fix any first-order NS-IC contract. Because the agent does not want to covertly save or borrow conditional on truthful reporting, standard arguments imply that the recommended consumption process $c := s + b$ must satisfy the agent's Euler equation [**3.8**] given the market rate $r = \rho$, meaning that the agent's flow utility u_t defines a martingale (because $u'(c) = -\theta u(c)$ under exponential utility). Plugging this into [**2.2**] and [**4.1**] yields

$$q_t = \int_t^\infty e^{-\rho(\tau - t)}\, \mathbf{E}_t^* \left[u_\tau\right] \mathrm{d}\tau = u_t/\rho$$

$$p_t = \int_t^\infty e^{-(\rho+\lambda)(\tau - t)} \theta\, \mathbf{E}_t^* \left[u_\tau\right] \mathrm{d}\tau = \theta u_t/(\rho + \lambda)$$

where, in each line, the first equality follows from Tonelli's Theorem and the second equality follows from the Euler equation. Thus, the utility rate process β is constant with $\beta_t \equiv \rho$, and by [**FO-IC**] the geometric sensitivity must be constant with $k_t \equiv \rho\theta/(\rho + \lambda) =: k_\circ^*$. Thus, the contract is precisely Contract W.

J. Material for Theorem 6

Let $\lambda = 0$ throughout the present appendix without further mention. We require some preliminary definitions and observations. First, recall (from [**2.2**]) that the agent's promised utility q is a y-adapted process where

[**J.1**] $$q_t := \mathbf{E}_t^* \left[\int_t^\infty e^{-\rho(\tau - t)} u_\tau \,\mathrm{d}\tau \right]$$

denotes the agent's continuation utility conditional on $(y_\tau)_{\tau \in [0,t]}$, assuming that he had reported truthfully in the past and is going to report truthfully in the future (i.e., $y = b$), where $u_t(y) := u(s_t(y) + y_t)$. By [J.1] and the definition of the probability measure $\mathbf{P}^*$ in Online Appendix E, the q induced by any FO-IC contract solves the SDE

$$\mathrm{d}q_t = [\rho q_t - u_t]\,\mathrm{d}t - q_t\theta\sigma\,\mathrm{d}W_t \tag{J.2}$$

Second, for an admissible misreporting strategy m, define the b-adapted process $q^m = (q_t^m)_{t\geqslant 0}$ by

$$q_t^m := \mathbf{E}_t^{m^{\langle t\rangle}}\left[\int_t^\infty e^{-\rho(\tau-t)}u(s_\tau(y) + y_\tau - m_\tau)\,\mathrm{d}\tau\right] \tag{J.3}$$

which denotes the agent's continuation utility conditional on $(b_\tau)_{\tau\in[0,t]}$ under the *truncated* misreporting strategy $m^{\langle t\rangle}$ defined by

$$m_\tau^{\langle t\rangle} := m_{\tau\wedge t} \tag{J.4}$$

That is, under the strategy $m^{\langle t\rangle}$ the agent follows m up until time t, after which his misreport is frozen at m_t. We then have

$$q_t^m = e^{\theta m_t}\,\mathbf{E}_t^{m^{\langle t\rangle}}\left[\int_t^\infty e^{-\rho(\tau-t)}u(s_\tau(y) + y_\tau)\,\mathrm{d}\tau\right] \tag{J.5}$$

$$= e^{\theta m_t}\,\mathbf{E}_t^{*}\left[\int_t^\infty e^{-\rho(\tau-t)}u(s_\tau(y) + y_\tau)\,\mathrm{d}\tau\right] \tag{J.6}$$

$$= e^{\theta m_t} q_t \tag{J.7}$$

In the above display, [J.5] follows from the definition of $m^{\langle t\rangle}$, the definition of q_t^m in [J.3], and the assumption of exponential utility. The second line [J.6] relies on the assumption that $\lambda = 0$, given which the expression follows from the fact that (fixing the initial condition y_t) the law of $(y_\tau)_{\tau\in(t,\infty)}$ is identical under the measures $\mathbf{P}^*$ and $\mathbf{P}^{m^{\langle t\rangle}}$ by [E.3]. Finally, [J.7] follows from the definition of q_t in [J.1]. The equality $q_t^m = e^{\theta m_t} q_t$ will be used in the sequel.

J.0.1. *Well-Behaved Reporting Strategies*

Definition J.1. Let $\lambda = 0$. Misreporting strategy m is *well-behaved given contract s* if:

(i) It is admissible.

(ii) It induces a process q^m (as defined in [J.3]) such that

$$\lim_{T\to\infty} e^{-\rho T}\,\mathbf{E}_0^m\left[q_T^m\right] = 0.$$

(iii) The induced process $Z^m = (Z_t^m)_{t \geqslant 0}$ defined by

$$Z_t^m \equiv \int_0^t e^{-\rho\tau} q_\tau^m \, \mathrm{d}W_\tau^m \tag{J.8}$$

is a martingale under P^m.

The following lemma shows that the class of well-behaved strategies is sufficiently large. In particular, it includes (i) the truthful strategy and (ii) any admissible strategy run for "a short enough period of time" before reversion to truthful reporting. Intuitively, strategies of type (ii) correspond to misreporting for finitely-many periods in the analogous discrete-time model, though additional integrability conditions are required in the present continuous-time setting.

Lemma J.2. Let s be a FO-IC contract. If $\lambda = 0$, then the following misreporting strategies are well-behaved given s:

(i) The truthful strategy $m^* \equiv 0$.

(ii) Given any admissible strategy m and $n \in \mathbb{N}$, the *localized strategy* $m^{(n)}$ defined by

$$m_t^{(n)} := m_{t \wedge \tau_n}$$

where the b-adapted stopping time τ_n is defined by

$$\tau_n := \inf \left\{ t \geqslant 0 : \ \max \left[\Gamma_t^m e^{\theta m_t}, e^{-\rho t} e^{\theta m_t} |q_t| \right] \geqslant n \right\}$$

with Γ^m the density process defined in [E.1].

The proof of Lemma J.2 is in Online Appendix J.0.3 below.

J.0.2. *Proof of Theorem 6*

We may now present the proof of the theorem, which largely mirrors the proof of incentive compatibility in Sannikov (2008, Proposition 2), but accounts for the fact that deviations from truthtelling have persistent effects on the agent's continuation utility (an effect that is absent in Sannikov (2008)).

Proof of Theorem 6. Fix a FO-IC contract s that delivers lifetime utility q_0 and a misreporting strategy m that is well-behaved given s. Let $V^m = (V_t^m)_{t \geqslant 0}$ denote the b-adapted process for which V_t^m is the agent's *total* expected lifetime payoff from the truncated strategy $m^{\langle t \rangle}$ defined in [J.4], conditional on $\mathcal{F}_t^b$. (Notice that as t varies, V_t^m computes expected lifetime utility under *different* truncations of the strategy m.) By definition, it is given by

$$V_t^m = \int_0^t e^{-\rho\tau} e^{\theta m_\tau} u_\tau \,\mathrm{d}\tau + e^{-\rho t} q_t^m$$

[**J.9**]
$$= \int_0^t e^{-\rho\tau} e^{\theta m_\tau} u_\tau \,\mathrm{d}\tau + e^{-\rho t} e^{\theta m_t} q_t$$

where q_t^m is as defined in [**J.3**] and the second line follows from [**J.7**]. Applying Itô's lemma to [**J.9**] yields

$$\begin{aligned} \mathrm{d}V_t^m &= e^{-\rho t} e^{\theta m_t} u_t - \rho e^{-\rho t} e^{\theta m_t} q_t \,\mathrm{d}t + \theta \Delta_t e^{-\rho t} e^{\theta m_t} q_t \,\mathrm{d}t \\ &\quad + e^{-\rho t} e^{\theta m_t} \underbrace{\left[(\rho q_t - u_t)\ \mathrm{d}t - \theta\sigma q_t \,\mathrm{d}W_t\right]}_{=\ \mathrm{d}q_t \text{ under } \mathbf{P}^*} \\ &= \theta e^{-\rho t} e^{\theta m_t} q_t \left[\Delta_t \,\mathrm{d}t - \sigma \,\mathrm{d}W_t\right] \end{aligned}$$

where the second equality follows from canceling terms and, in the first line, $\mathrm{d}q_t$ is expressed under $\mathbf{P}^*$ because the agent truthfully reports increments on $[t, \infty)$ by definition of $m^{\langle t \rangle}$ and the fact that $(y_\tau)_{\tau \in (t,\infty)}$ has identical laws under $\mathbf{P}^*$ and $\mathbf{P}^{m^{\langle t \rangle}}$ because $\lambda = 0$ (see [**E.3**]). Because m is admissible by Definition J.1(i), we have the change-of-measure formula $\sigma \,\mathrm{d}W_t = \Delta_t \,\mathrm{d}t + \sigma \,\mathrm{d}W_t^m$ from [**E.2**], which implies that

[**J.10**]
$$\mathrm{d}V_t^m = -\theta\sigma e^{-\rho t} e^{\theta m_t} q_t \,\mathrm{d}W_t^m = -\sigma\theta Z_t^m$$

where Z^m is as defined in [**J.8**]. Thus, V^m is a local martingale under $\mathbf{P}^m$, the measure induced by the *untruncated* strategy m. By Definition J.1(iii), Z^m is a true martingale under $\mathbf{P}^m$. It follows that for any $T > 0$,

$$V_0^m = \mathbf{E}_0^m \left[V_T^m\right] = \mathbf{E}_0^m \left[\int_0^T e^{-\rho t} e^{\theta m_t} u_t \,\mathrm{d}t\right] + e^{-\rho T}\, \mathbf{E}_0^m \left[q_T^m\right]$$

Taking the limit $T \to \infty$ yields

[**J.11**]
$$V_0^m = \mathbf{E}_0^m \left[\int_0^\infty e^{-\rho t} e^{\theta m_t} u_t \,\mathrm{d}t\right] + \lim_{T\to\infty} e^{-\rho T}\, \mathbf{E}_0^m \left[q_T^m\right] = \mathbf{E}_0^m \left[\int_0^\infty e^{-\rho t} e^{\theta m_t} u_t \,\mathrm{d}t\right]$$

where the first equality follows from the Monotone Convergence Theorem (since $u_t < 0$) and the second equality follows from Definition J.1(ii). Note that [**J.11**] is the agent's expected payoff from the (non-truncated) strategy m. However, the definitions of V_0^m and q_0 imply that $V_0^m = q_0 = \mathbf{E}_0^* \left[\int_0^\infty e^{-\rho t} u_t \,\mathrm{d}t\right]$. We conclude that the agent is indifferent among all well-behaved reporting strategies and truthful reporting (which is itself well-behaved by Lemma J.2). □

J.0.3. *Proof of Lemma J.2*

Point (i): Note that m^* admissible by definition (see Example 1 in Online Appendix E). To see that it satisfies Definition J.1(ii), note that $q = q^{m^*}$, so that integrating [J.2] (with $W^m \equiv W$ and $\Delta \equiv 0$) over $[0, T]$ and taking expectations yields

$$q_0 = \mathbf{E}_0^* \left[\int_0^T e^{-\rho t} u_t \, \mathrm{d}t \right] + e^{-\rho T} \, \mathbf{E}_0^* \left[q_T \right]$$

Taking the limit $T \to \infty$ yields

$$q_0 = \mathbf{E}_0^* \left[\int_0^\infty e^{-\rho t} u_t \, \mathrm{d}t \right] + \lim_{T \to \infty} e^{-\rho T} \, \mathbf{E}_0^* \left[q_T \right] = \mathbf{E}_0^* \left[\int_0^\infty e^{-\rho t} u_t \, \mathrm{d}t \right]$$

where the first equality follows from the Monotone Convergence Theorem (since $u_t < 0$) and the second equality follows from [J.1] (with $t = 0$). Finally, to see that m^* satisfies Definition J.1(iii), note that V^{m^*} (as defined in [J.9]) is a martingale under $\mathbf{P}^*$ because $V_t^{m^*} = \mathbf{E}_t^* [X]$ for the $\mathbf{P}^*$-integrable random variable $X := \int_0^\infty e^{-\rho t} u_t \, \mathrm{d}t$. Moreover, by [J.10], we have $Z_t^{m^*} \equiv -V_t^{m^*}/(\sigma\theta)$.[38] Thus, Z^{Δ^*} is also a martingale under $\mathbf{P}^*$.

Point (ii): Let admissible m and $n \in \mathbb{N}$ be given. Admissibility of $m^{(n)}$ follows from the observation that the density process $\Gamma^{m^{(n)}}$ satisfies $\Gamma_t^{m^{(n)}} \equiv \Gamma_{t \wedge \tau_n}^m$ and the fact that a stopped UI martingale (in this case Γ^m) is itself a UI martingale. For Definition J.1(ii), observe that

[J.12] $$\mathbf{E}_0^{m^{(n)}} \left[q_T^{m^{(n)}} \right] = \mathbf{E}_0^* \left[\Gamma_{T \wedge \tau_n}^m \exp\left(\theta m_{T \wedge \tau_n} \right) q_T \right] \geqslant n \, \mathbf{E}_0^* \left[q_T \right],$$

where the equality follows from admissibility of $m^{(n)}$ and the inequality follows from (a) the definition of τ_n and (b) the facts that $\Gamma_{T \wedge \tau_n}^m \exp\left(\theta m_{T \wedge \tau_n} \right) > 0 > q_T$. Because m^* is well-behaved by point (i) of the lemma, it follows from [J.12] that $\liminf_{T \to \infty} e^{-\rho T} \, \mathbf{E}_0^{m^{(n)}} \left[q_T^{m^{(n)}} \right] \geqslant 0$. Thus, $\lim_{T \to \infty} e^{-\rho T} \, \mathbf{E}_0^{m^{(n)}} \left[q_T^{m^{(n)}} \right] = 0$ because $q^{m^{(n)}} < 0$. Finally, the localized strategy satisfies Definition J.1(iii) because, by construction, the integrand of the stochastic integral defining $Z_t^{m^{(n)}}$ in [J.8] is uniformly bounded by n; hence, the stochastic integral defines a true martingale under $\mathbf{P}^m$.

(38) There is no circular argument in the previous two sentences. Although the referenced displays [J.9] and [J.10] appear in the proof of Theorem 6 in Online Appendix J.0.2, which in turn references the present Lemma J.2, the derivation of these displays depends only on the admissability of m^*, which has already been independently established.

K. Discussion Omitted from the Main Text

K.1. Restricted Reporting Strategies

We briefly describe how our results extend to the case in which the agent's reports are required to satisfy $m \leqslant 0$, the assumption that W11 attempts to impose (recall Remark 1). In that case, the first-order condition [**FO-IC**] would be relaxed to the inequality $\gamma_t + p_t \geqslant 0$ (as on p. 1244 of W11), and the definition of FO-IC contracts adjusted accordingly. The definition of (first-order) NS-IC contracts in the hidden savings model from Subsection 5.1 would require analogous adjustment. In the statement of Theorem 4, Contract W would remain NS-IC, and we conjecture (but have not shown) that it would remain the optimal (but not the unique) first-order NS-IC contract. Theorem 6 would *not* hold as stated.

K.2. Discussion of Errors in W11

As noted in the main text, the formal analysis in W11 contains two main errors.[39] First, Contract W is asserted to be optimal for $\lambda > 0$, which our Theorem 2 shows to be false. Second, it is asserted that requiring reports to satisfy $m \leqslant 0$ implies that they must satisfy $\Delta \leqslant 0$, which is clearly false and impacts results in W11 pertaining to the verification that FO-IC contracts are in fact IC (recall Remark 1). In addition to these formal errors, W11 provides incorrect explanations for the optimality of Contract W and its properties, as well as the relation between that paper's results and those of the prior discrete-time literature. These explanations remain incorrect even when $\lambda = 0$, in which case Contract W is indeed the optimal FO-IC contract (as confirmed by our Theorem 5).

(1) Derivation of Contract W. When $\lambda > 0$, W11 attempts to derive Contract W as the optimal full-commitment contract through the following steps:
Step 1. Conjecture that the principal's value function depends on p only through the ratio $k = p/q$, which enters via a function $h(k)$ that is additively separable from all terms involving y and all other terms involving q (see display 26, p. 1252).
Step 2. Assume that $h(\cdot)$ is smooth, plug the conjectured form of the value function from Step 1 into the the HJB equation, and derive a second-order ODE that $h(\cdot)$ must satisfy (display A.11, p. 1269) in order for the conjectured form of the value function to be consistent with the HJB.

(39) As described in Remark 3, the argument given in W11 that [**FO-IC**] necessarily follows from [**IC**] appears to be incomplete, but we conjecture that this technical gap can be closed.

Step 3. Use the policy functions derived from the HJB, together with Itô's lemma, to determine that the resulting k process must satisfy a particular SDE (display 28, p. 1253).
Step 4. Numerically solve the ODE for $h(\cdot)$ derived in Step 2, and observe that $h(\cdot)$ appears to be minimized at the value $k^*_\circ$ (pp. 1270-71).
Step 5. Observe that, given the initial condition $k_0 = k^*_\circ$, the k process is necessarily constant. Conclude that the policy function from the HJB generates Contract W.
Step 6. Conclude that Contract W is the optimal FO-IC contract and, since it is IC, also the optimal full-commitment contract.

The main error in this derivation is the numerical observation in Step 4. (Step 5 is correct conditional on the preceding steps.) Relatedly, in Step 2, W11 does not specify either the domain of $h(\cdot)$ or the boundary conditions for its ODE, which are needed to verify that a suitable solution to the HJB exists in Step 2 and to (numerically) solve the ODE for $h(\cdot)$ in Step 4. One can see that Step 4 is the main error as follows (details are in the Supplementary Appendix, Bloedel, Krishna, and Strulovici (2020)). Begin by deducing from first principles that the principal's first-order value function (defined in Bloedel, Krishna, and Strulovici (2020)) takes the form conjectured in W11's Step 1. As long as the h function is twice continuously differentiable, one can follow the analytical calculations underlying Steps 1–3 and 5 to show that, if the optimal FO-IC contract is in fact stationary, then it must have initial geometric volatility $k_0 = k^*_\circ$ and therefore be Contract W. But this is impossible because our Theorem 2 implies that Contract W is strictly dominated.

Separately, W11's Step 6 is also incomplete without a standard verification argument for the candidate optimal contract derived from the HJB equation and, more importantly, because that paper's argument that Contract W is IC is incorrect without the restriction that the agent's reports satisfy $\Delta \leqslant 0$. We turn the latter issue next.

(2) Restriction on Reporting Strategies. Remark 1 and footnotes 13–14 in the main text explain that W11's assertion that $m \leqslant 0$ implies $\Delta \leqslant 0$ is incorrect, and that the $\Delta \leqslant 0$ restriction is used throughout the proof of W11's Theorem 1, calling the validity of that result into question. W11 correctly verifies that Contract W is IC in the special case where $\lambda = 0$, in that the argument given remains valid when Δ is unconstrained and independently of whether $m \leqslant 0$ is required (cf. the discussion following Theorem 6 in Subsection 5.2). We now show that the corresponding verification for Contract W in the $\lambda > 0$ case is incorrect, in that W11's analysis ceases to be valid when $m \leqslant 0$ is imposed but the $\Delta \leqslant 0$ restriction is relaxed. (A fortiori, it is also invalid when both m and Δ are unconstrained.)

Let $\lambda > 0$. Suppose the agent is restricted to reporting strategies satisfying $m \leqslant 0$ and admissibility (but place no other constraints on Δ). W11 attempts to verify (in Appendix A.3.2, pp. 1271-72) that Contract W is IC by explicitly solving for the value function in the agent's reporting problem, which is shown on p. 1272 to take the form $V(q, m) = qe^{\theta m} \cdot (\rho + \lambda)/(\rho + \lambda + \theta\lambda m)$.[40] However, notice that $V(q, \cdot)$ is only well-defined on $(\underline{M}, 0]$ and that $\lim_{m \downarrow \underline{M}} V(q, m) = -\infty$, where $\underline{M} := -(\rho + \lambda)/(\theta\lambda)$. (Notice that $V(q, m) > 0$ when $m < -\underline{M}$, which is impossible because $u(\cdot) < 0$.) Intuitively, when $m_t < 0$ the principal expects stronger positive mean-reversion than actually occurs, and so punishes the agent for not reporting increments $\mathrm{d}y_t > \mathrm{d}b_t$. Meanwhile, the constraint $\Delta \leqslant 0$ forces the agent to report increments $\mathrm{d}y_t \leqslant \mathrm{d}b_t$, meaning that he cannot avoid such punishments by making up for past under-reports. As $m \downarrow \underline{M}$, the punishments become so severe that the agent's continuation value decreases without bound.

By contrast, as described above in Online Appendix F.2 the analysis in our paper (and that in Strulovici 2020) shows that the agent's value function under Contract W takes the form $V(q, m) = qe^{\theta m}$. It is easy to show that, under this assumption, his value function takes the same functional form (but on a restricted domain) when the requirement that $m \leqslant 0$ is imposed. Note that this value function is well-defined for all $m \leqslant 0$, and differs from that in W11 by the multiplicative factor $(\rho + \lambda)/(\rho + \lambda + \theta\lambda m)$ that causes W11's value function to explode as $m \downarrow \underline{M}$. This establishes that — at least at non-truthful histories under Contract W — $\Delta \leqslant 0$ is a strictly more binding constraint for the agent than $m \leqslant 0$. The preceding argument assumes that the agent is allowed to report jumps in the endowment. We conjecture that his value function is unchanged when we disallow jumps by requiring that reports be admissible.[41]

(3) Explanation for Contract W and its Properties. As noted in Subsection 5.2.1, the explanation given in W11 for the optimality of Contract W and why its properties differ from those in the classic discrete-time literature is misleading in several respects. This is true even in the special case that $\lambda = 0$, in which case Contract W is, in fact, optimal. The explanation given in W11 hinges on three assertions (pp. 1235-36, 1257-58, 1264). (i)

(40) W11 also verifies (on p. 1271) that Contract W satisfies the (infinite-horizon analogues of the) bounds in that paper's Theorem 1. However, this approach is not valid without a proof of W11's Theorem 1 that does not rely the $\Delta \leqslant 0$ restriction.

(41) Here is a sketch: Initialize Contract W at state (q_{t-}, m_{t-}) with $m_{t-} < 0$. When jumps are allowed, we have seen that the agent wants to immediately re-set to $m_t = 0$. By Example 2 in Online Appendix E, for each $\varepsilon > 0$ there exists an admissible strategy specifying $\Delta_\tau = m_{t-}/\varepsilon$ for $\tau \in [t, t + \varepsilon]$ and $\Delta_\tau = m_\tau = 0$ thereafter. Calculations in Strulovici (2020) suggest that the agent's payoff to such an admissible strategy converges to the upper bound $V(q_{t-}, m_{t-}) = q_{t-}e^{\theta m_{t-}}$ as $\varepsilon \to 0$.

The agent's incentive constraints are qualitatively different in continuous- as opposed to discrete-time models. (ii) This difference is the reason why the well-known *inverse Euler equation (IEE)*, which requires that $1/u'(c_t)$ define a martingale, holds in some related discrete-time models but fails in W11's solved examples. (iii) The failure of the IEE is responsible for, or at least closely related to, the failure of immiseration. We describe below why all three assertions are misleading. All three seemed to have caused confusion in the literature about the role of continuous-time modeling in dynamic contracting.[42]

(i) Role of Continuous Time. W11 (p. 1235) explains the role of continuous time as follows:

> "[T]hese differences [between Contract W, which leads to bliss, and the optimal contract in Thomas and Worrall (1990), which leads to immiseration,] rely at least partly on differences in the environments. In the discrete analogue of my model, when deciding what to report in the current period, the agent trades off current consumption and future promised utility. In my continuous-time formulation, the agent's private state follows a process with continuous paths and the principal knows this. Thus in the current period the agent only influences the future increments of the reported state. Thus current consumption is independent of the current report and all that matters for the reporting choice is how future transfers are affected … [T]he reporting problem and, hence, the incentive constraints become fully forward-looking …"

However, the continuous- vs. discrete-time distinction cannot explain why Contract W differs from the optimal contract in Thomas and Worrall (1990) because these contracts differ *even in the discrete-time model*. To see this, first note that precise analogues to Contract W and our Theorems 1–3 and 5–6 exist in the discrete-time version of the model (as described in Section 6 of the main text), showing that the discrete-time analogue Contract W is strictly dominated whenever shocks are transient and is optimal in the permanent shock case, in which that the agent is indifferent among (essentially) all reporting strategies under any FO-IC contract. In particular, the analogue to Contract W will be strictly suboptimal when types are iid in the discrete-time model, as in Thomas and Worrall (1990). Consequently, the (sub)optimality of Contract W and its difference from the full-commitment optimum does not depend on whether time is discrete or continuous,

(42) For instance, Kapička (2013) studies a discrete-time optimal taxation model with persistent types. Both that paper (on p. 1029) and W11 (on pp. 1262-63) attribute the differences in their findings, in large part, to the continuous- vs. discrete-time distinction. Zhang (2009) studies a related continuous-time taxation model with binary persistent types in which the IEE and immiseration both hold, writing on p. 652 that "because the inverse Euler equation is no longer valid in Williams (2011), the immiserization does not hold and consumption has a positive drift and increasing variability."

but rather hinges on the persistence of the agent's information, as we have argued in Subsection 5.2.1.

More broadly, while W11's above description of the agent's continuous-time reporting problem itself is technically correct (cf. footnote 11), it is also misleading. Although the agent's choice of Δ_t doesn't affect c_t because it doesn't affect y_t, it *does* affect $y_{t+\varepsilon}$ and hence $c_{t+\varepsilon}$ for all $\varepsilon > 0$, no matter how small; because continuous time is not well-ordered, this is tantamount to the agent trading off consumption "today" (i.e., c_τ for $\tau \in [t, t+\varepsilon)$ with $\varepsilon > 0$ small) and promised utility "tomorrow" (i.e., $q_{t+\varepsilon}$). Conversely, as the period length in discrete time becomes short, the impact of current consumption on the agent's incentives vanishes, so it is without loss of optimality for the principal to make period-t consumption depend only only on *prior* reports $(y_\tau)_{\tau=0}^{t-1}$ but *not* the current report y_t (cf. Sadzik and Stacchetti 2015). These observations suggest (without proving) that the continuous-time model behaves similarly to the discrete-time model with short period length. More concretely, it can be shown that the qualitative properties of the Self-Insurance Contracts on which we focus — and, under permanent shocks, all FO-IC contracts — are independent of the discrete-time period length and are retained in the continuous-time limit studied here.[43]

One might object that the real distinction between the discrete- and continuous-time models arises from the restriction to admissible strategies in the latter, reasoning that in discrete time the agent is permitted to choose any path of misreports $(m_t)_{t \in \mathbb{N}}$, while in continuous time the agent is restricted to choosing "small" misreports at each instant (because the mapping $t \mapsto m_t$ must be absolutely continuous). Even if this distinction is meaningful in principle, it is not relevant to Contract W or any other Self-Insurance Contract: we have shown in Online Appendix F.2 that such contracts remain incentive compatible even when the agent is allowed to report jumps in his endowment, which do impact his instantaneous consumption and flow utility, exactly as in the discrete-time model.

(ii)–(iii) Role of the IEE. The IEE arises as an optimal inter-temporal cost-smoothing condition *for the principal* in certain dynamic agency settings, most notably in *separable optimal dynamic taxation models* in which the agent's private information concerns his labor productivity and the agent's preferences over consumption and labor are additively separable.[44] Golosov, Kocherlakota, and Tsyvinski (2003) show that, in such settings, the

(43) We have not shown that the same is true for all FO-IC contracts or for the optimal full-commitment contract (though we conjecture that it is). A complete analysis of the relationship between discrete- and continuous-time contracting models with persistent private information — analogous to Sadzik and Stacchetti's (2015) study of the iid hidden action case — is an important topic for future research.

(44) Formally, the agent's flow utility is $u(c_t) - v(b_t, l_t)$ where c_t is consumption, b_t is labor productivity,

IEE holds for very general discrete-time private information processes. It also generally holds in continuous-time taxation models with both continuous sample paths (Farhi and Werning 2013) and jumps (Zhang 2009). Thus, the IEE is independent of whether a model is cast in continuous or discrete time.

However it depends sensitively on the agent's preferences, and is neither necessary nor sufficient for immiseration. When the agent's marginal utility of consumption is independent of his private information — as in the separable taxation models — his incentives are determined entirely by the present value of consumption utility, which serves as a "payment rule" that implements an "allocation rule" (of, e.g., labor effort) as is familiar from static mechanism design. In such settings, the IEE arises from the principal choosing the cost-minimizing consumption process inducing any such incentive compatible payment rule and, consequently, is independent of the (sub)optimality of the labor allocation that it supports.[45] However, it is well known that the IEE fails when when the agent's utility over consumption is not independent of his private information, in which case consumption utility cannot be interpreted as a payment rule (Golosov, Kocherlakota, and Tsyvinski 2003; Farhi and Werning 2013; Golosov, Troshkin, and Tsyvinski 2016). This includes the discrete-time models on which the solved examples in W11 are based (Thomas and Worrall 1990; Atkeson and Lucas 1992). Since immiseration occurs in those models, it follows that the IEE is not necessary for immiseration. It is also not sufficient, as can be seen by noting that the IEE holds under a wide range of suboptimal (but IC) contracts in the separable taxation models.[46]

and l_t is labor effort. The contract offered by the principal specifies processes for c and l, adapted to the agent's reports.

(45) Pavan, Segal, and Toikka (2014, p. 620) discuss this point in the context of their dynamic generalization of the Revenue/Payment Equivalence Theorem.

(46) For instance, in the separable taxation model, the IEE holds under the IC contract in which the c and l processes are constant, which clearly does not induce immiseration. In a distinct but related hidden action model, Prat and Jovanovic (2014, p. 883, fn. 31) give an example in which, under the *optimal* contract, the IEE holds and yet consumption converges to its *upper* bound.