Dynamic Contracts Under Loss Aversion

Alejandro Jofre Sofia Moroni Andrea Repetto*

September 25, 2012

Abstract

We analyze a dynamic moral hazard principal-agent model with an agent who is loss averse and whose reference updates according to the previous period's consumption. Under full commitment and when the agent has no access to credit markets, the optimal payment scheme can have flat segments at the reference level. This property implies that there is a positive probability of observing constant wages over time, even though the scheme has memory. Moreover, the model predicts a "status quo bias" -a preference for consuming his full allocation ex post- whenever the agent is allowed to borrow or save after the outcome is realized . This result in turn implies that unlike the canonical model, the optimal contract may be implemented even when the agent has access to a saving technology. We also show that although the optimal contract scheme is renegotiation-proof, it cannot be implemented by a sequence of spot contracts. However, when the agent has access to the credit market and the principal can monitor his savings, the long-term optimal contract is spot contractible. Finally, to deal with non-differentiabilities, we use subdifferential calculus to compute the optimal program.

1 Introduction

In this paper we modify the classical principal-agent model with moral hazard by assuming that the agent is loss averse to payments below the previous period's consumption. We analyze the optimal contracting problem both when the agent has no access to credit markets and when he does have access to credit but the principal can monitor his savings.

The dynamic moral hazard model under the classical risk aversion assumptions

*Universidad de Chile, Yale University and Universidad Adolfo Ibanez, respectively. We have greatly benefited from comments made by Hector Chade, Adam Kapor, Juan Pablo Torres-Martinez, Felipe Balmaceda and Nicolas Figueroa. We also thank seminar participants at University of Chile, Universidad Adolfo Ibanez, Central Bank of Chile, European University, International Conference on Continuous Optimization, Latin American Econometric Society Meetings, North American Econometric Society Summer Meetings.

has been extensively analyzed in the literature. The characteristics of this second best optimal scheme are developed in Rogerson (1985) and Chiappori et al. (1994). Under mild conditions, the restricted savings model predicts that the agent's compensation is strictly increasing in the current period outcome. It also predicts that the contract exhibits memory; i.e., the agent's compensation in one period also depends on the outcomes of all prior periods. Furthermore, as the optimal contract smoothes the expected (inverse) of marginal utility over time, the scheme offers compensations that are front-loaded. Intuitively, since the agent needs to rely on the principal to transfer resources over time, the principal can reduce the cost of providing effective incentives by keeping the marginal utility of consumption low in earlier periods. In turn, this implies that if the agent were ex-post allowed to save or borrow, he would never choose to consume more than the current period's compensation. If anything, he would choose to save. Finally, when the agent has no access to a savings technology, the long-term contract cannot be implemented by a sequence of spot contracts because of the memory-in-wages property.

Under monitored savings, the classical model predicts that the principal will control savings to provide intertemporal consumption smoothing, and will balance insurance and incentives through the payment scheme. Since the long-term contract has now memory in consumption and not in wages, it is spot-implementable. Furthermore, the long-term contract is renegotiation proof.

In this paper we review the robustness of these predictions to the introduction of reference dependence and loss aversion in the agent's utility function. Loss aversion was first proposed by Kahneman and Tversky (1979) as an essential element of their Prospect Theory. Under these preferences, the dislike that consumption below the reference generates is greater than the elation produced by an equally sized gain. A large and growing body of literature in economics and in cognitive psychology supports the hypothesis that references do affect individual decisions.[1] In this paper, we study a dynamic set up in which the reference is updated endogenously. Similar to Bowman et al. (1999), Munro and Sugden (2003) and Dittmann et al. (2010) we assume the agent's reference is equal to the previous period's consumption; that is, in addition to the standard consumption utility, he derives gain-loss utility from comparing current consumption with lagged consumption.[2]

This new framework confirms many of the predictions of the classical dynamic

[1]Rabin (1998) and Della Vigna (2009) survey this literature.

[2]There is little evidence on how reference levels are determined. Our assumption is also similar to models of the endowment effect that posit that willingness to pay for a good depends on recent ownership status. In a dynamic setup like ours, the alternative of assuming that the reference is based upon expectations may lead to solutions in which the agent makes plans that will not be carried out (Koszegi and Rabin, 2009). It is worth emphasizing that our model is one of fully forward-looking behavior: both the agent and the principal take into account the effects of their behavior on future references.

moral hazard model. In particular, we find that the optimal long-term contract is non-decreasing in outcomes, displays memory in wages and provides consumption smoothing. Moreover, the full commitment optimum is ex-post efficient and therefore it is renegotiation proof. We also show that in the monitorable access to credit case the contract is renegotiation-proof, implementable via spot contracts and it coincides with the no-access optimum.

Our main finding, however, is that the optimal contract we derive displays a number of relevant new properties. Most important, these new features of the second best optimal scheme are better in line with observed contracts.

First, optimal wage schedules may be insensitive to outcomes in an interval, offering to pay the reference income for a set of performance measures. Intuitively, the cost of inducing effort by increasing payments right above the reference may be high due to the sudden decrease in marginal utility. Similarly, although effective in providing incentives, a reduction in payment just below the reference increases the cost of inducing participation, again due to the discontinuity in marginal utility. In other words, it is optimal to provide full insurance locally because loss aversion involves first-order risk aversion around the reference (Segal and Spivak, 1990).

Second, for all periods after the first, the optimal wage schedule pays the reference for an interval of outcomes. Except for the last period, the flat segment may even extend for the whole support of the outcomes' distribution. Thus incentives may be optimally provided not by rewards and punishments that are contingent on the current period's result but by the promise of future income.

Third, the fact that the payment schemes exhibit flat segments in most periods implies, under weak assumptions, that there is a positive probability that two consecutive payments are equal; e.g., that wages exhibit time persistence as reported by Dickens et al. (2007). In contrast, the classical model predicts variability in observed wages whenever the outcomes' distribution is continuous. Moreover, there is a positive probability that the agent perceives a fixed wage over $T-1$ periods; therefore all incentives must be deferred to the last period with schemes that are at least partly sensitive to outcomes in T, a prediction consistent with the evidence in Baker, Jensen and Murphy (1988) and in Baker, Gibbs and Holmstrom (1994).

Fourth, when the agent is allowed to save or borrow after the realization of the performance measure, we find that the agent may either choose to save, borrow or consume his wage. In the classical model the optimal payment scheme requires the agent to consume more than he would choose in order to facilitate the provision of incentives in future periods. In our model, this effect is also present. However, there are other effects in play that may make the agent want to borrow or consume his full allocation in some periods. On the one hand, saving and borrowing not only modify the intertemporal allocation of consumption but also change the future references. On the other hand, changes in consumption may lead to current or future losses, motivating the agent to choose to consume his full allocation.

In other words, loss aversion implies that the marginal utility of savings may not be equal to (minus) the marginal utility of borrowing. In fact, they may be both simultaneously negative. Therefore, whenever the agent experiences a loss by either saving or borrowing, he will prefer to consume his full wage in a manner consistent with the "status quo bias" of Samuelson and Zeckhauser (1988). Moreover, whenever the agent is consuming his reference in one period or there is a positive probability of being paid the reference in the next, the smallest interest rate at which the agent would be willing to lend part of his income is strictly higher than the rate he would be willing to pay to borrow. That is, the model predicts a related phenomenon (Bateman et al., 1997): a discrepancy between the willingness to accept and the willingness to pay.

An important consequence of our analysis is that the optimal payment scheme does not always require constrained savings. Many of the canonical model's predictions hinge on the extreme assumption that the agent's borrowing and saving are constrained, either because credit is not available to him or because the principal can monitor his actions in the credit market. In particular, under these assumptions, the optimal long-term contract is renegotiation-proof. However, when the agent does have access to the credit market and his savings and borrowing are private information, the full commitment long-term optimum is not renegotiation proof. In fact, the renegotiation-proof long-term contract cannot provide incentives to exert any effort above the minimum. Since it is unlikely that a court of law would prevent renegotiation towards a Pareto-improving agreement and since constraining savings may be implausible in most contexts, the classical theory cannot explain the existence of long-term commitment contracts (Chiappori et al., 1994). Thus loss aversion and our assumed dynamic update of the reference might give a rationale for the ubiquity of commitment contracts.

Finally, we show that the sequence of optimal spot contracts is not memoryless as in the classical case and will in general not coincide with the full-commitment optimum.

There exists a small but growing literature on moral hazard, optimal contracts and loss aversion. This literature intends to explain the gaps between the rich contracts predicted by theoretical developments and the fairly simple contracts that are actually observed (Prendergast, 1999; Salanie, 2003). De Meza and Webb (2007) first introduced loss aversion to the static principal agent model. When the reference is either exogenous or equal to the certainty equivalent of rewards, their model predicts that over some interval pay may be insensitive to performance. Bonuses may arise when the reference is the median reward. In the latter case, the principal can provide insurance and incentives avoiding the loss area by lowering the median and by rewarding good performance at the same time.

Optimal binary payment schemes also arise in Herweg et al. (2010). Their one-period principal agent model assumes a piece-wise linear gain-loss function and reference formation as in Koszegi and Rabin (2006, 2007); i.e., the agent derives gain-loss utility by comparing the actual payment with his (lagged) rational expectations about rewards. In particular, Koszegi and Rabin's formulation assumes that a stochastic outcome is evaluated according to its expected utility, with each payment being compared to all outcomes in the support of the reference lottery. Under this setting, the optimal scheme is a lump sum bonus contract with the bonus paid whenever a certain level of performance is achieved. The intuition is that under this reference formation process, a contract specifying many possible wages induces the agent to make comparisons that may yield losses or gains. Given loss aversion, on net news hurt the agent, reducing his expected utility and increasing the average payment needed to make him participate. A simple binary contract induces effort and minimizes the cost of inducing the agent to accept the contract.[3]

Iantchev (2011) provides a structural estimation of the one-period principal-agent model under moral hazard and a loss averse agent. The model allows for multiple principals and multiple agents. The reference point is determined by the expected value of the payment under the equilibrium contract in the market, as defined by Rayo and Becker (2007). Under these assumptions, optimal pay may also be insensitive to performance in an interval. Because the agent is assumed risk-loving in the loss space, the payment scheme displays a discontinuity whenever output falls below a threshold.[4]

To our knowledge, Macera (2012) is the only other paper that considers loss aversion in a dynamic setting of the principal-agent model. Macera (2012) extends the static model to a two period model in which the agent is loss averse to changes in beliefs about present and future consumption as in Koszegi and Rabin (2009), an extension of the static environment assumptions of Koszegi and Rabin (2006, 2007). Under certain conditions on the relative strength of current and future gain-loss utility, the optimal contract offers a fixed wage in the first period and an output contingent increasing wage scheme in the second period. That is, the principal defers incentives to the future.

Our model shares many features of this growing body of literature that modifies the principal-agent model by assuming that the agent is loss averse. Our model enriches the setting, however, by allowing for a T period relationship and by analyzing the properties of the optimal contract in addition to its complexity-

[3] Daido and Itoh (2006) also allow for loss aversion and the Koszegi and Rabin (2006, 2007) reference formation. The model assumes a binary measure of performance. Unfortunately, little can be said about the form of the optimal contract under this binary output measure assumption.

[4] The discontinuous drop in payments when the observed outcome measure is low is also found in Dittmann et al. (2010) who also assume risk-loving in the loss space. The paper calibrates a one period principal agent model in order to explain CEO compensation. The calibration assumes, though, that the optimal scheme is piecewise linear.

renegotiation proofness, spot contractibility, wage persistence and the role of the constrained savings assumption. It also shows that predicted contracts may not be as simple once dynamics is considered.

In addition, since loss aversion induces a discontinuity in marginal utility, we rigorously derive the optimality conditions using convex analysis tools. We show that the program the principal faces has a concave objective function and that the feasible set is convex. Therefore, the optimum can be characterized by a "zero belongs to the subgradient set" condition (Rockafellar, 1970 and 1974; Rockafellar and Wets, 1997). This methodology is useful in dealing with non-differentiabilities and the lack of validity of the usual first order conditions.

The remainder of the paper is organized as follow. In Section 2 we present the model. In Section 3 we derive the optimality conditions in the full-commitment case with no access to credit markets to characterize the shape of the optimal second best payment scheme. We also derive the main intra and intertemporal properties of this optimal contract. In Section 4 we analyze the optimal scheme under monitorable access to credit. In Section 5, and in order to illustrate our findings, we develop a numerical example of the solution of a two period model. Finally, we conclude in Section 6.

2 The Model

The model describes a repeated principal-agent problem analogous to the dynamic moral hazard model of Rogerson (1985) and **?**; e.g., we assume finite horizon, discounting and a risk neutral principal who can borrow and save at a fixed interest rate. Our model differs, however, in that we assume a loss averse agent.

The relationship between the principal and the agent lasts $T+1$ periods. In each period $i \in \{0, \ldots, T\}$ the agent exerts an unobservable action $a_i \in \{a_L, a_H\}$ with $a_L < a_H$. The outcome in period i is denoted $x_i \in [\underline{x}_i, \bar{x}_i]$ with a differentiable distribution function $f^i(x_i|a_i)$, where a_i denotes the action chosen. The distributions of outcomes are independent across periods conditional on actions. We assume that these distributions exhibit the Monotone Likelihood Ratio Property (MLRP); that is, if we denote $f^i_{a_i}(x_i|a_i) = f^i(x_i|a_H) - f^i(x_i|a_L)$, then $\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}$ is non-decreasing in x_i. We let the wage schedule in period i depend on the outcomes realized in the current and in all previous periods and denote it $\omega_i(x_0, x_1, \ldots, x_i)$.

Let c_i be the agent's consumption in period i and $R_i > 0$ the corresponding reference point. Also let $\psi_i(\cdot)$ represent an increasing and convex cost function. Then the agent's utility is

$$\tilde{U}(c_i, R_i) - \psi_i(a_i) \tag{1}$$

We denote the cost difference between the low and high action as $\Delta\psi_i = \psi_i(a_H) - \psi_i(a_L)$.

To allow for loss aversion we assume that the agent's preferences are characterized by the following property:

$$\lim_{t\to 0^+}\frac{\tilde{U}_i(R+t,R)-\tilde{U}_i(R,R)}{t} < \lim_{t\to 0^+}\frac{\tilde{U}_i(R-t,R)-\tilde{U}_i(R,R)}{-t}$$

that is, the left-sided derivative of $\tilde{U}_i$ is greater than its right-sided derivative. In other words, we assume that the utility gain of consumption over R is lower than the utility loss of consumption below R in an equally sized amount. We assume that $\tilde{U}_i$ is continuous, concave and differentiable at all points other than R, that $\lim_{c\to 0}\tilde{U}_i(c,R) = -\infty$ and $\lim_{c\to\infty}\tilde{U}_i(c,R) = \infty$.[5]

For $\ell_0 > 0$, an exogenous reference level R_0 and a smooth, concave and strictly increasing function $U(\cdot)$, given R_0, without loss of generality, the period 0 utility, $\tilde{U}_0$ can be expressed as: [6]

$$\tilde{U}_0(c_0,R_0) = U(c_0) - \ell_0\theta(c_0,R_0)\left(U(R_0)-U(c_0)\right)$$

where

$$\theta(c,R) = \begin{cases} 1 & \text{if } c < R \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

A graphical representation of the first period utility function is given in Figure 2. As the reference increases from R to R', utility above R' remains unaffected, while it drops below R'.

For periods $i > 0$, we assume that the reference level depends dynamically on the consumption that took place in the previous period; more specifically, we assume that R_{i+1} equals c_i. This assumption is based on psychological evidence indicating that individual choices depend not only on current consumption but also on previous levels of consumption (see **?**). **?**, **?**, Iantchev (2011) and Dittman et al. (2010) have also assumed similar forms of update.[7]

We assume the utility in period $i+1$ takes the following form:

$$\tilde{U}_{i+1}(c_{i+1},R_i) = \tilde{U}_{i+1}(c_{i+1},c_i) = U(c_{i+1}) - \ell_{i+1}\theta(c_{i+1},c_i)\left(U(c_i)-U(c_{i+1})\right)$$

with $\theta(c_{i+1},c_i)$ defined as in (2) and $\ell_{i+1} > 0$. We assume the agent discounts the future exponentially at rate δ. We also assume that $\ell_i\delta < 1$ in order to ensure that the total utility of two consecutive periods is increasing in the consumption

[5] Note that we do not assume diminishing sensitivity, that is, that the agent is risk averse in gains but risk loving in losses.

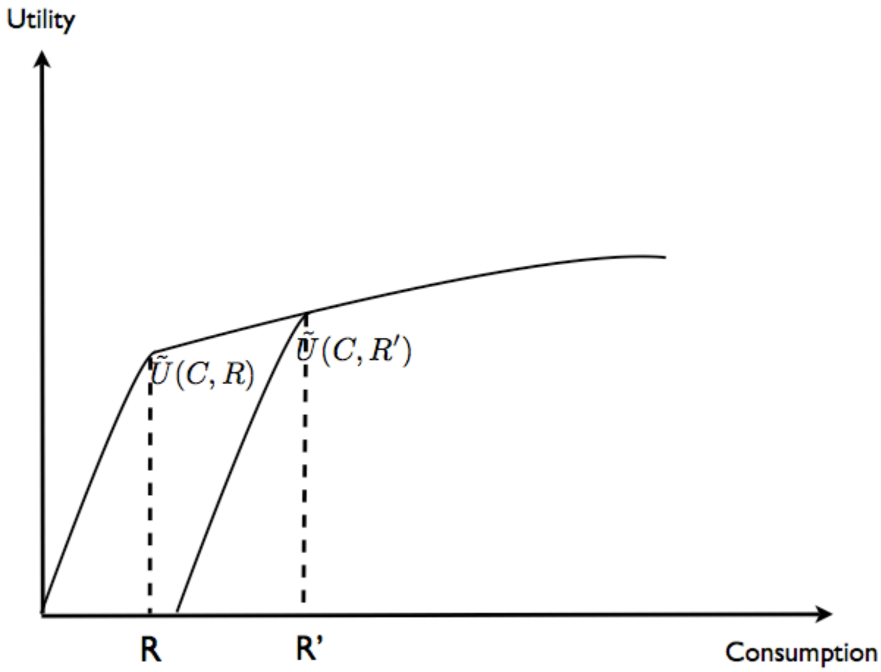
[6] A proof can be found in the appendix.

[7] Other papers assume different processes for reference formation. For instance, Koszegi and Rabin (2006 and 2007) use the rational expectation of consumption. Gul (1991) takes the certainty equivalent. Chetty and Szeidl (2010) derive a reference that partly depends on past consumption in a model of adjustment costs in consumption. The issue of reference formation is still an understudied problem.

of the first.[8]

The principal is risk neutral and therefore, for any given outcome x_i, her utility is $x_i - \omega_i(x_0, x_1, \ldots, x_i)$. Finally, we assume that the principal also discounts the future exponentially at rate δ.

Figure 1: Utility function for different levels of reference.



[8]Recall that utility is decreasing in the reference for consumption levels below the reference. In addition, we let ℓ_i to vary across periods to allow for generality. For instance, Harrison and List (2004) shows that loss aversion can be mitigated by market experience, suggesting that ℓ_i might decrease over time.

3 Full commitment and no access to credit markets

We start by analyzing the optimal contracting problem under the assumptions of full commitment and no access to credit markets. That is, we assume that both the principal and the agent are able to commit to the contract during the whole duration of the relationship, and that the principal can borrow and save at the fixed interest rate $\frac{1}{\delta} - 1$, but that the agent can neither borrow nor save.

Whenever the agent has no access to credit markets, he must consume his current income. Thus, in this case, the principal faces the following program,

$$\max_{(\omega_i(\cdot))_i,(a_i(\cdot))_i} \sum_{i=0}^{T} \delta^i \mathbb{E}\left(x_i - \omega_i | a_0, a_1, \ldots, a_i\right)$$

subject to

$$\sum_{i=0}^{T} \delta^i \mathbb{E}\left(\tilde{U}_i(\omega_i, R_i) - \psi_i(a_i) | a_0, a_1, \ldots, a_i\right) \geq U^* \tag{PC}$$

$$a(\cdot) = (a_0, a_1(\cdot), \ldots a_T(\cdot)) \in \text{argmax}_{a(\cdot)} \sum_{i=0}^{T} \delta^i \mathbb{E}\left(\tilde{U}_i(\omega_i, R_i) - \psi(a_i) | a_0, a_1, \ldots, a_i\right) \tag{IC}$$

where $\mathbb{E}(\cdot | a_0, a_1, \ldots, a_i)$ denotes an expectation given actions $(a_0, a_1, \ldots, a_i)$. [9]

The objective function represents the expected payment to the principal. The first constraint (PC) is the standard participation constraint. It prescribes that the agent must obtain an expected utility of a at least U^* from the relationship. Constraint (IC) states that the effort chosen maximizes the expected utility of the agent, and is henceforth referred to as the incentive compatibility constraint.

3.1 Optimality conditions

In order to find optimality conditions, it is convenient to define a function $h_i(v_0, v_1, \ldots, v_i)$ that represents the cost to the principal of providing a level of utility v_i in period i whenever the utility provisions in the previous periods were $\{v_0, v_1, \ldots, v_{i-1}\}$. The utility provision cost v_i depends on the realized outcomes up to time i, $(x_0, x_1, \ldots, x_i)$. Note that because of the reference dependent preferences and the dynamic update assumption, the provision of any given level of utility affects the shape of the agent's utility in future periods.

[9] This expectation evaluated on an arbitrary function $g(x_0, \ldots, x_i)$ is defined as,

$$\mathbb{E}\left(g | a_0, a_1, \ldots, a_i\right) = \int g(x_0, x_1, \ldots, x_i) f^0(x_0 | a_0) f^1(x_1 | a_1(x_0)) \cdots f^i(x_i | a_i(x_0, \ldots, x_{i-1})) dx_0 dx_1 \ldots dx_i.$$

We can now rewrite the program the principal faces as choosing utility provisions $v_i(x_1, x_2, \ldots, x_i)$ contingent on the outcomes up to period i as follows,

$$\max_{(v_i(\cdot))_i,(a_i)_i} \sum_{i=0}^{T} \delta^i \mathbb{E}\left(x_i - h_i(v_0, v_1, \ldots, v_i)|a_0, a_1, \ldots, a_i\right) \tag{3}$$

subject to

$$\sum_{i=0}^{T} \delta^i \left(\mathbb{E}\left(v_i|a_0, a_1, \ldots, a_i\right) - \psi_i(a_i)\right) \geq U^* \tag{PC'}$$

$$a = (a_0, a_1(x_0), \ldots a_T(x_0, x_1, \ldots, x_{T-1})) \in \text{argmax}_a \sum_{i=0}^{T} \delta^i \left(\mathbb{E}\left(v_i|a_0, a_1, \ldots, a_i\right) - \psi(a_i)\right) \tag{IC'}$$

$h_i(v_0, v_1, \ldots, v_i)$ is an increasing and continuous function given by,

$$h_i(v_0, v_1, \ldots, v_i) = \begin{cases} U^{-1}\left(\frac{v_i + \ell_i U(h_{i-1}(v_0, v_1, \ldots, v_{i-1}))}{1+\ell_i}\right) & \text{if } v_i < U(h_{i-1}(v_0, v_1, \ldots, v_{i-1})) \\ U^{-1}(v_i) & \text{if } v_i \geq U(h_{i-1}(v_0, v_1, \ldots, v_{i-1})) \end{cases} \tag{4}$$

where $U(h_{-1}) = R_0$.

The following three properties determine the optimality conditions for the payment scheme. Property 1 proves that the problem is indeed convex which allows for the use of subdifferential calculus in finding the conditions for the optimal contract. That is to say, we can write a Lagrangian as in the everywhere differentiable case, and find the optimal contract such that zero belongs to the sub-gradient set of the Lagrangian.[10] Property 2 characterizes the sub-gradient set of the cost function which is a crucial step in computing the subgradient set of the Lagrangian. Property 3 describes the optimality conditions.

Property 1. *[Convexity]*

Under the assumptions of the model, the utility provision cost functions $h_i : \mathbb{R}^i \to \mathbb{R}$ *for* $i \in 1, \ldots, T$ *are strictly convex and therefore the optimization problem given by* (3)-(PC')-(IC') *has a strictly concave objective function and a convex feasible set.*

Proof. See appendix.

□

Property 2. *[Subgradient set]*

[10] Note that in the differentiable case we would need to find a zero of the differential with respect to the optimization variables given the multipliers. Since in this case the objective function is not differentiable, optimization can be attained by finding the sub-gradient set. A reference for these ideas is **?**.

The subgradient set[11] *of* $h_i(v_0, v_1, \dots, v_i)$ *is given by*

$$\partial h_i(v_0, v_1, \dots, v_i) = \left(\frac{1}{U'(\omega_i)} \left(\prod_{t=j+1}^{i} \frac{k_t(x_0, x_1, \dots, x_t)\ell_t}{1 + k_t(x_0, x_1, \dots, x_t)\ell_t} \right) \frac{1}{1 + k_j(x_0, x_1, \dots, x_j)\ell_j} \right)_{j=0}^{i} \tag{5}$$

where

$$k_t(x_0, x_1, \dots, x_t) \in \begin{cases} \{1\} & \text{if } \omega_t(x_0, x_1, \dots, x_t) < R_t \\ [0,1] & \text{if } \omega_t(x_0, x_1, \dots, x_t) = R_t \\ \{0\} & \text{otherwise} \end{cases} \tag{6}$$

Proof. See appendix. □

Property 3. *[Optimality conditions]*

There is a unique optimal wage schedule that solves the program faced by the principal and it is characterized by the following optimality conditions,

$$\frac{1}{U'(\omega_i(x_0, x_1, \dots, x_i))} = (1 + k_i(x_0, x_1, \dots, x_i)\ell_i) \left(\lambda_i + \mu_i \frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)} \right) +$$
$$-\delta \ell_{i+1} \int_{\omega_{i+1} \leq \omega_i} k_{i+1}(x_0, x_1, \dots, x_{i+1})(\lambda_{i+1} + \mu_{i+1} \frac{f^{i+1}_{a_{i+1}}(x_{i+1}|a_{i+1})}{f^{i+1}(x_{i+1}|a_{i+1})}) f^{i+1}(x_{i+1}|a_{i+1}) dx_{i+1}. \qquad \forall i < T \tag{7}$$

and

$$\frac{1}{U'(\omega_T(x_0, x_1, \dots, x_T))} = (1 + k_T(x_0, x_1, \dots, x_T)\ell_T) \left(\lambda_T + \mu_T \frac{f^T_{a_T}(x_T|a_i)}{f^i(x_i|a_i)} \right) \tag{8}$$

where

- $\lambda_i = \lambda + \sum_{k=0}^{i-1} \mu_k \frac{f^k_{a_k}(x_k|a_k)}{f^k(x_k|a_k)}$, *with* λ *the multiplier associated to (PC') and* $\mu_i = \mu_i(x_0, \dots, x_{i-1})$ *the multipliers associated to the incentive compatibility constraints.*
- *The function* $k_i(x_0, x_1, \dots, x_i)$ *is associated to the kink in the utility function and is given by*

$$k_i(x_0, x_1, \dots, x_i) \in \begin{cases} \{1\} & \text{if } \omega_i(x_0, x_1, \dots, x_i) < R_i \\ [0,1] & \text{if } \omega_i(x_0, x_1, \dots, x_i) = R_i \\ \{0\} & \text{otherwise} \end{cases}$$

Proof. See appendix. □

Note that whenever $\ell_i = 0 \quad \forall i$ the optimality condition describes the solution to the canonical case. In addition, the optimality condition for a spot contract is given by equation (8) when $T = 0$, as in **?**. In the following subsection we describe the main intra and intertemporal properties of this optimal scheme.

[11] By definition we know that for a generic convex function $f : \mathbb{R}^{n+1} \to \mathbb{R}$ the subgradient set at $x \in \mathbb{R}^{n+1}$ is given by the set of vectors $d = (d_{0,} d_1, \dots, d_{n+1}) \in \mathbb{R}^{n+1}$ such that for any vector $\alpha \in \mathbb{R}^{n+1}$

$$f(x + \alpha) \geq f(x) + d \cdot \alpha$$

3.2 Properties of the optimal payment scheme

Inspection of equations (7) and (8) imply that there are two main properties that distinguish the shape of optimal payment scheme from the classical case. First, it can have flat segments. This is explained by the multiplicative term $(1 + k_i(x_0, x_1, \ldots, x_i)\ell)$ in (7) and (8). At the reference level, $k_i(x_0, x_1, \ldots, x_i)$ is allowed to take any value in $[0, 1]$. Therefore, the right hand side of equations (7) and (8) can remain constant in an interval: as x_i increases, $k_i(x_0, x_1, \ldots, x_i)$ decreases and $\omega_i(x_0, x_1, \ldots, x_i)$ remains at the reference. Intuitively, the cost of inducing effort by increasing payments right above the reference may be high due to the discontinuous fall in marginal utility. Similarly, although effective in providing incentives, a reduction in payment just below the reference increases the cost of inducing participation, again due to the discontinuity in marginal utility.

The second difference relates to the fact that the principal takes into account that each period's payment affects the reference level of the following period. The last term of the right hand side of equation (7), which is strictly positive, represents this effect. In each period i, this term tends to lower the payment scheme and to reduce its growth rate as x_i increases. Intuitively, the term represents the benefit of lowering the payment in the current period, in order to reduce the following period's reference and to increase the utility of the agent in the loss area. Naturally, this effect does not occur in T (equation (8)).

On the basis of these optimality results, in the following property we show that the optimal wage schedule can be insensitive to outcomes in an interval. We then examine memory and the relationship between payments across any pair of consecutive periods.

Property 4. *[Shape of the optimal payment scheme] If $\ell_{i-1} \geq \ell_i \geq \ell_{i+1}$ and $\delta\ell_i \leq 1$, then $\omega_i(x_0, x_1, \ldots, x_i)$ is continuous and non-decreasing in x_i and,*

1. *For $i \in \{0, \ldots, T\}$, $\omega_i(x_0, x_1, \ldots, x_i)$ is non-decreasing in x_j for $j \in \{0, \ldots, i\}$.*
2. *For $i \in \{1, \ldots, T\}$, if $\omega_{i-1}(x_0, x_1, \ldots, x_{i-1}) > R_{i-1}$ then for any value of $(x_0, x_1, \ldots, x_{i-1})$ it must be the case that $\omega_i(x_0, x_1, \ldots, x_i) = R_i$ for some outcome $x_i \in [\underline{x}_i, \bar{x}^i]$. Furthermore the payment scheme has a flat segment at the reference and therefore, ω_i is not strictly increasing.*
3. *For $i \in \{1, \ldots, T-1\}$, if $(\ell_i - \ell_{i+1})\delta \geq \ell_{i-1} - \ell_i$ and $\omega_{i-1}(x_0, x_1, \ldots, x_{i-1}) \leq R_{i-1}$ then, for any value of $(x_0, x_1, \ldots, x_{i-1})$, $\omega_i(x_0, x_1, \ldots, x_i) = R_i$ for some outcome $x_i \in [\underline{x}_i, \bar{x}^i]$. Furthermore the payment scheme has a flat segment at the reference and, therefore, ω_i is not strictly increasing.*
4. *If $\omega_{T-1}(x_0, x_1, \ldots, x_{T-1}) \leq R_{T-1}$ then, for any value of $(x_0, x_1, \ldots, x_{T-1})$, $\omega_T(x_0, x_1, \ldots, x_T) = R_T$ for some outcome $x_T \in [\underline{x}_T, \bar{x}^T]$. The payment scheme has a flat segment at the reference but it cannot be fully flat in x_T.*

Proof. See appendix. □

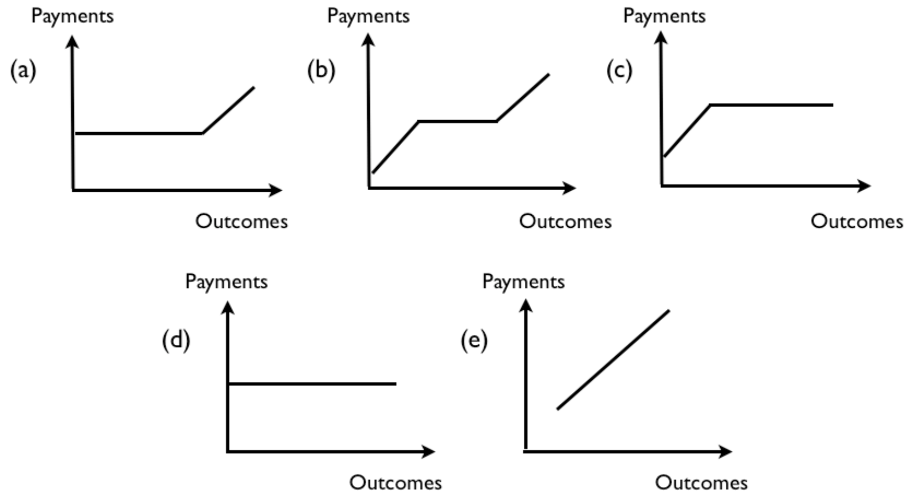


Figure 2: Schematic representation of monotonicity of contracts

The previous property states that the payment scheme is, as in the canonical case, non-decreasing. However, starting in period 1 the reference wage must be paid for an interval of outcomes. Figure 3.2 is an schematic representation of the monotonicity of possible payment schemes. According to Property 4, panels (a), (b) and (c) are possible in all periods. Panel (d) is possible in periods $\{0, \ldots, T-1\}$. Flat schedules are not optimal in period T. Two elements distinguish period T. One refers to the fact that consumption in T does not affect the reference of future periods, so the effect of a high wage in T on the cost of providing utility is limited. The other is that there is a positive probability that the agent perceives a fixed wage over $T-1$ periods. In that case, all incentives must be deferred to the last period with schemes that are at least partly sensitive to outcomes. Macera (2009) finds a similar result in a two-period model. Finally, panel (e) is possible in period 0 only. In period 0, the payment scheme may or may not offer the reference level of consumption for some outcome.

Next we analyze how the payment scheme depends on the history of outcomes. Just as in the canonical model, consumption smoothing requires that a higher payment in one period results in a higher payment in all subsequent periods. In the classical model the implication is that wage schedules are strictly increasing in every realized outcome. However, in our model, wage schedules may overlap for outcomes that pay the reference level. This result is summarized in the following property.

Property 5. *[Dependence across periods]*

Let $x_i' < x_i''$ two possible outcomes in period i.

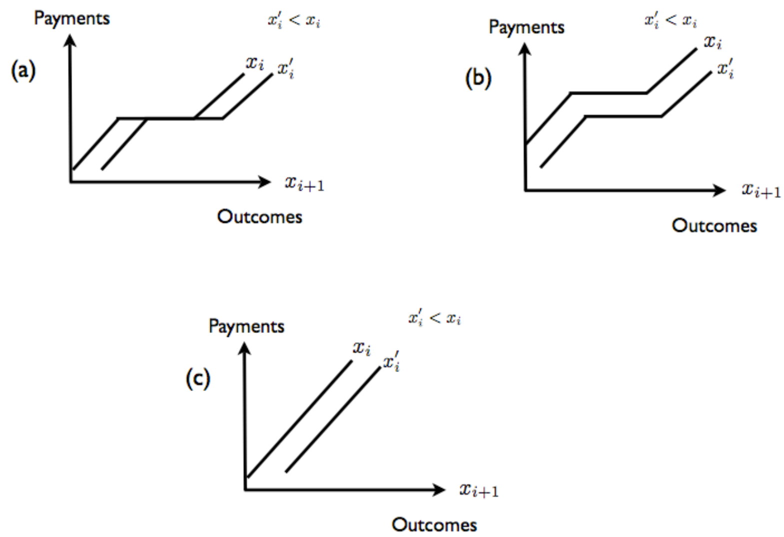


Figure 3: Optimal contracts, outcomes between two consecutive periods

1. *If* $\omega_i(x_0, x_1, \ldots x_i') = \omega_i(x_0, x_1, \ldots x_i'') = R_i$ *then*

   $\omega_j(x_0, x_1, \ldots x_i', x_{i+1}, \ldots, x_j) \leq \omega_j(x_0, x_1, \ldots x_i'', x_{i+1}, \ldots, x_j) \quad \forall j > i \quad \forall x_j \in [\underline{x}_j, \overline{x}_j]$

2. *If* $\omega_i(x_0, x_1, \ldots x_i') < \omega_i(x_0, x_1, \ldots x_i'')$ *then*

   $\omega_j(x_0, x_1, \ldots x_i', x_{i+1}, \ldots, x_j) < \omega_j(x_0, x_1, \ldots x_i'', x_{i+1}, \ldots, x_j) \quad \forall j > i \quad \forall x_j \in [\underline{x}_j, \overline{x}_j]$

Proof. It follows directly from (7) and (8) since $\mu_i > 0 \quad \forall i$.

□

Property 5 implies that panels (a) and (b) in Figure 3.2 are possible in our model. Note also that only (c) is possible in the canonical model. This property implies in turn that there is a positive probability that wages exhibit time persistence, even if realized outcomes differ over time.

Next, we find an analogue to the relationship between the wage schedules offered in any two consecutive periods as was first derived by Rogerson (1985) under classical assumptions. In the classical case, the inverse of the marginal utility of income must equal the conditional expected value of the inverse of the next period's marginal utility of income. This condition is no longer valid in our model. However, an extended condition can be derived as is stated in the following property. Let $\omega_i(x_0, x_1, \ldots, x_i)$ be denoted $\omega_i(x_i)$ to simplify notation, and similarly let $k_i(x_i)$ denote $k_i(x_0, x_1, \ldots, x_i)$.

Property 6. *[Relationship between two consecutive periods] The following relationship between two consecutive periods is fulfilled,*

$$\frac{1}{U'(\omega_{i-1}(x_{i-1}))(1 + k_{i-1}(x_{i-1})\ell_{i-1})} = \int \frac{1}{U'(\omega_i(x_i))(1 + k_i(x_i)\ell_i)} f^i(x_i|a_i)dx_i + c(x_{i-1})$$

where

$$c(x_{i-1}) = -\frac{\ell_i \delta}{1+k_{i-1}(x_{i-1})\ell_{i-1}} \int k_i(x_i) \left(\lambda_i + \mu_i \frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}\right) f^i(x_i|a_i)dx_i + \ell_{i+1}\delta \int\int \frac{k_{i+1}(x_{i+1})}{1+k_i(x_i)\ell_i} \left(\lambda_{i+1} + \mu_{i+1} \frac{f^{i+1}_{a_{i+1}}(x_{i+1}|a_{i+1})}{f^{i+1}(x_{i+1}|a_{i+1})}\right) f^{i+1}(x_{i+1}|a_{i+1}) f^i(x_i|a_i) dx_{i+1} dx_i \quad (9)$$

Proof. Follows directly from equations (7) and (8).

□

Property 6 implies that the inverse of the marginal utility of income might be greater or smaller than the conditional expectation of the marginal utility.[12]

This relationship between any two adjacent periods implies in the classical case that the optimal contract front-loads consumption; i.e., if allowed to save or borrow after the realization of the current period outcome, the agent will choose to save. In our model, however, this will not necessarily be the case: the agent may have incentives to save, borrow or consume her full allocation ex-post. Furthermore, he may face utility losses by either saving or by borrowing and thus might be inclined to consume his full allocation.

Intuitively, assume the loss averse agent faces the possibility of reallocating resources between any two consecutive periods, i and $i+1$. If he borrows, the effect on current marginal utility depends on whether today's income is over or under the reference. Due to loss aversion, whenever the current payment is at or over the reference, the gain in utility is small relative to whenever current income is below the reference. Similarly, the effect on tomorrow's marginal utility might be relatively high if because of previous period's borrowing, consumption falls under the reference. These effects may make borrowing unattractive to the agent. Reinforcing this effect, this intertemporal reallocation of resources increases period $i+1$ reference, reducing utility and increasing the marginal utility of consumption in the loss area. A similar argument applies to the decision of saving. Thus the possibility of facing marginal losses that may be larger than marginal gains, jointly with the intertemporal effects on the reference, imply that the agent will face situations in which he would face a loss in utility by either saving or borrowing. He will thus decide to consume the full allocation.

This discussion is an application of the "status quo bias" as described by **?**. A related phenomenon is that this situation may lead to a gap between willingness to pay (WTP) and willingness to accept (WTA). Assume the agent is in a situation in which he would lose by either saving or borrowing at the market interest rate $1/\delta - 1$. However, if given the possibility of lending a part of his income at interest rate r_l or borrowing at interest rate r_b, indifference between lending and

[12]There is some abuse of language here since we refer as marginal utility of $\tilde{U}_i$ to the term $U'(\omega_i(x_i))(1+k_i(x_i)\ell_i)$ since both quantities are equal for all incomes except the reference. At the reference level the marginal utility is not computable and could take any value between $[U'(\omega_i(x_i)), U'(\omega_i(x_i))(1+\ell_i)]$.

borrowing would require $r_l > \frac{1}{\delta} - 1 > r_b$. That is, the smallest price at which he is willing to lend is strictly greater than the largest price he is willing to pay for borrowing. This is not the case in the canonical model since indifference between lending and borrowing for any set of payment schemes in any two consecutive periods can be attained at a single interest rate. This gap between WTA and WTP has been described in other economical contexts (**?**). Moreover, if the market offers interest rates such that $r_b \geq r_l$, then the agent will find himself inclined to consume his allocation.

An important consequence is that under our assumptions the optimal payment scheme does not always require constrained savings, a problematic property of the canonical model (Chiappori et al., 1994). In particular, renegotiation proofness in the canonical case relies on the assumption of constrained savings, either because credit is not available or because the principal can monitor the agent's actions in the credit market. If the agent has access to the credit market but his savings cannot be contracted upon, then the full commitment long-term optimum is ex-post inefficient. Moreover, the renegotiation-proof long-term contract cannot provide incentives to implement any effort above the minimum. Since it is unlikely that a court of law would prevent renegotiation towards a Pareto-improving agreement and since constraining savings may be implausible in most contexts, the classical theory cannot explain the existence of long-term commitment contracts (Chiappori et al., 1994). Thus loss aversion and our assumed dynamic update of the reference might give a rationale for the existence of commitment contracts.

A formalization of this discussion is presented in the following properties.

Property 7. *[Status quo bias]*

- *If the payment in period i is y and is at the reference and the payment in period $i+1$ is constant such that $y_{i+1}(x_{i+1}) = y \quad \forall x_{i+1} \in [\underline{x}_{i+1}, \bar{x}_{i+1}]$, then the agent will neither want to save nor borrow at the rate $\frac{1}{\delta} - 1$. Moreover, if r_l is the rate that makes the net marginal utility of saving equal to zero and r_b the rate that makes the net marginal utility of borrowing equal to zero, then $r_l > \frac{1}{\delta} - 1 > r_b$.*
- *Let y_i be the payment in period i and $y_{i+1}(x_{i+1})$ the payment scheme in period $i+1$. If y_{i+1} pays the reference with positive probability or y_i is at the reference, then the rate r_l that makes marginal utility of saving equal to zero is strictly greater than the rate r_b that makes the marginal utility of borrowing equal to zero.*

Proof. See appendix. □

Property 8. *[Intertemporal allocation of resources] If the interest rate is $\frac{1}{\delta} - 1$ in period i then the agent may have incentives to save for period $i+1$, to borrow and pay back in period $i+1$ or to consume exactly her income depending on the parameters of the problem. Moreover,*

- *If period's $i+1$ payment scheme is over the reference for all results, then the agent does not have incentives to save in period i.*
- *If period's $i+1$ payment scheme is below the reference for all outcomes in period $i+1$ then*
  - *if period's i payment is strictly above the reference then the agent has incentives to save in period i.*
  - *If period's i payment is at the reference, the agent will not have incentives to borrow in period i.*
  - *If period's i payment is strictly below the reference the agent may have incentives to save, to borrow or to consume her allocation.*

Proof. See appendix. □

The previous property states, among other things, that if the payment scheme is flat for low outcomes and then increasing for larger outcomes, then the agent will not have incentives to save. In the following subsection we show that the payment scheme will not take values under the reference as long as the cost to the principal of the participation constraint λ is sufficiently high.

3.3 The shape of the optimal contract and the shadow cost of participation

In a three period context we analyze how the optimal payment scheme changes if the cost for the principal of the participation constraint were to change. We analyze this case by studying the effects of a change in the multiplier λ.

Property 9. *[Cost of (PC) and shape of optimal contract]*
There is a value $\bar{\lambda}$ such that if $\lambda \geq \bar{\lambda}$ the optimal payment scheme is strictly above the reference, for all realizations, in each period (letting the other multipliers be fixed).

Proof. See appendix.

□

Recall that λ is the shadow cost of relaxing the participation constraint. This property says that if the participation constraint is sufficiently costly then the contract must be backloaded. It suggests that the cost of the participation constraint to the principal is closely related to whether incentives are to be created through rewards or punishments. This result is highly intuitive: providing incentives through the threat of payments below the reference creates a great loss in utility for an agent whose participation is already very costly. The principal is better off thereby providing incentives through rewards only.

3.4 Renegotiation-proofness and spot-implementability

The optimal contract scheme is renegotiation-proof just as in the classical case. As a matter of fact, **?** has shown that the renegotiation-proofness property does not rely on the differentiability of the utility function. Underlying this result is the assumption that the agent is able to predict how her utility updates in each period. If this were not the case, the property might not hold.

However, the optimal sequence of spot contracts exhibit memory, unlike the classical case, because of our assumption on the dynamic update of the reference level. The utility function of the agent changes from period to period with the reference, and so may the reservation utility. In order for the full-commitment contract with no access to credit markets to be implemented by a sequence of spot contracts, the update of the reservation utility after each outcome must exactly correspond to the continuation payoff that the agent expects under this contract. Thus, the optimal commitment contract may not be implemented by a sequence of spot contracts.

Property 10. *[Optimal spot contracting]*
The optimal sequence of spot contracts exhibits memory and it will not implement the full-commitment optimum in general.

This result follows by backwards induction as in **?**.

4 Monitorable access to credit

We now move to the case when the agent can reallocate resources intertemporally by borrowing or saving through the credit market. We further assume that the agent's trades in the credit market can be monitored, so savings can be contracted upon. The program that the principal faces is then the following,

$$\max_{(\omega_i(\cdot))_i(a_i)_i,(s_i)_i(S_i)_i} \sum_{i=0}^{T} \delta^i \mathbb{E}\left(x_i - \omega_i(x_0, x_1, \ldots, x_i) - S_i(x_0, x_1, \ldots, x_i) | a_0, a_1, \ldots, a_i\right)$$

subject to

$$\sum_{i=0}^{T} \delta^i \left(\mathbb{E}\left(\tilde{U}_i(\omega_i(x_0, x_1, \ldots, x_i) - s_i(x_0, x_1, \ldots, x_i), c_{i-1}) | a_0, a_1, \ldots, a_i \right) - \psi_i(a_i) \right) \geq U^* \qquad \text{(PC)}$$

$$(a_0, a_1(x_1), \ldots a_T(x_0, x_1, \ldots, x_T)) \in \operatorname{argmax}_{\vec{a}} \sum_{i=0}^{T} \delta^i \left(\mathbb{E}\left(\tilde{U}_i(\omega_i(x_0, x_1, \ldots, x_i) - s_i(x_0, x_1, \ldots, x_i), c_{i-1}) | a_0, a_1, \ldots, a_i \right) - \psi(a_i) \right) \qquad \text{(IC)}$$

where s_i are the agent's accumulated savings in period i; that is, the net savings of the agent in period i once the endowment derived from previous savings is taken into account. Similarly, S_i are the accumulated savings of the principal. It is easy to see that the previous program is equivalent to one in which the optimization variables are the consumptions of the agent in each period, given by $c_i(x_0, x_1, \ldots, x_i) = \omega_i(x_0, x_1, \ldots, x_i) - s_i(x_0, x_1, \ldots, x_i)$, and the aggregate accumulated savings, given by $s_i + S_i$, and constraint $s_T = -\frac{s_{T-1}}{\delta}$. Since the principal is assumed risk neutral, the optimality conditions are similar to (10) and (8) with c_i replacing reward ω_i. Therefore, the optimality conditions for consumptions with monitorable access to credit are as follows,

$$\frac{1}{U'(c_i(x_0, x_1, \ldots, x_i))} = (1 + k_i(x_0, x_1, \ldots, x_i)\ell_i)\left(\lambda_i + \mu_i \frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}\right) + \\ -\delta\ell_{i+1}\int k_{i+1}(x_0, x_1, \ldots, x_{i+1})(\lambda_{i+1} + \mu_{i+1}\frac{f^i_{a_{i+1}}(x_{i+1}|a_{i+1})}{f^i(x_{i+1}|a_{i+1})})f^i(x_{i+1}|a_{i+1})dx_{i+1}. \qquad \forall i < T \tag{10}$$

and

$$\frac{1}{U'(c_T(x_0, x_1, \ldots, x_T))} = (1 + k_T(x_0, x_1, \ldots, x_T)\ell_T)\left(\lambda_T + \mu_T \frac{f^T_{a_T}(x_T|e_i)}{f^i(x_i|e_i)}\right) \tag{11}$$

where $k_i(x_0, x_1, \ldots, x_i)$ is given by,

$$k_i(x_0, x_1, \ldots, x_i) \in \begin{cases} \{1\} & \text{if } \omega_i(x_0, x_1, \ldots, x_i) < R_i \\ [0,1] & \text{if } \omega_i(x_0, x_1, \ldots, x_i) = R_i \\ \{0\} & \text{otherwise} \end{cases}$$

This result is similar to what is obtained in the classical case: there is a strong relationship between the monitorable access to credit case and the full-commitment with no credit access case. Furthermore, just like in the classical case, monitoring borrowing and savings introduces memory to the principal-agent relationship and therefore the optimal long-term contract will be spot-contractible. These results are summarized in the following property.

Property 11. *[Spot contractibility under monitorable credit]*
Suppose the reservation utility $U_i^*(s_{i-1}, R_i)$ in period i depends on the savings that the agent has in period $i-1$ and on the reference level R_i, and that it is continuous, increasing in s_{i-1}. Then, under monitorable savings, the long-term optimal contract is spot contractible.

Proof. See appendix. □

5 A two period example

In this section we numerically compute the optimal payment scheme in a two period setting in order to illustrate the forms these optimal contracts can take. We assume the distribution function of outcomes $x_i \in [0,1]$ in period $i \in \{1,2\}$, for effort level $a_j \in \{a_L, a_H\}$ is a triangular function given by

$$f^i(x_i|a_j) = \begin{cases} \frac{2x_i}{a_j} & x_i \le a_j \\ \frac{2(1-x_i)}{1-a_j} & x_i > a_j \end{cases} \tag{12}$$

Note that in this case $\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}$ may not be strictly increasing with respect to outcomes x_i. We also assume $a_H = 1$ and $a_L = 0.1$ in which case $\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}$ is constant in $[0, 0.1]$.

We assume that utility is described by the function $U(Y) = \sqrt{Y}$ and therefore

$$\widetilde{U}_i(Y_i, R_i) = \sqrt{Y_i} - \theta(Y_i, R_i)\ell_i(\sqrt{R_i} - \sqrt{Y_i})$$

The optimality conditions are,

$$\frac{1}{U'(\omega_0(x_0))} = 2\sqrt{\omega_0(x_0)} = (1 + k_0(x_0)\ell_0)\left(\lambda_0 + \mu_0 \frac{f^0_{a_0}(x_0|a_0)}{f^0(x_0|a_0)}\right) + \\ - \delta\ell_1 \int_{\omega_1 \leq \omega_0} k_1(x_0, x_1)(\lambda_1 + \mu_1 \frac{f^1_{a_1}(x_1|a_1)}{f^1(x_1|a_1)})f^1(x_1|a_1)dx_1. \quad (13)$$

$$\frac{1}{U'(\omega_1(x_0, x_1))} = 2\sqrt{\omega_1(x_0, x_1)} = (1 + k_1(x_0, x_1)\ell_1)\left(\lambda_1 + \mu_1 \frac{f^1_{a_1}(x_1|a_1)}{f^1(x_1|a_1)}\right) \quad (14)$$

5.1 Case 1: First period payment independent of outcomes

The first example illustrates a case in which the first period payment does not depend on the outcomes that take place in that same period (see Figure 5.1). That is, the first period payment is constant at the reference level. The second period payment scheme is contingent on outcomes obtained on the first and second periods as depicted in Figure 5.1.[13] The values of parameters used in this simulation are given in the following table, along with the multipliers

ℓ_0	ℓ_1	a_H	a_L	$1/U'(R_0)$	λ	μ_0	μ_1
1	1	1	0.1	37	46.1	0.5	2

Note that the second period scheme falls below the reference if low outcomes are realized in the first period. For outcomes in the first period that are greater than a threshold, the agent faces a payment scheme in the following period that is greater or equal than the payment received in the first period. Consequently, according to Property 8, he will not have incentives to save for outcomes above a threshold.

[13]Note that the flat segments for small values of first and second period outcomes are due to the fact that $\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}$ is constant in $[0, 0.1]$.

Figure 4: Case 1. First Period Payment Scheme

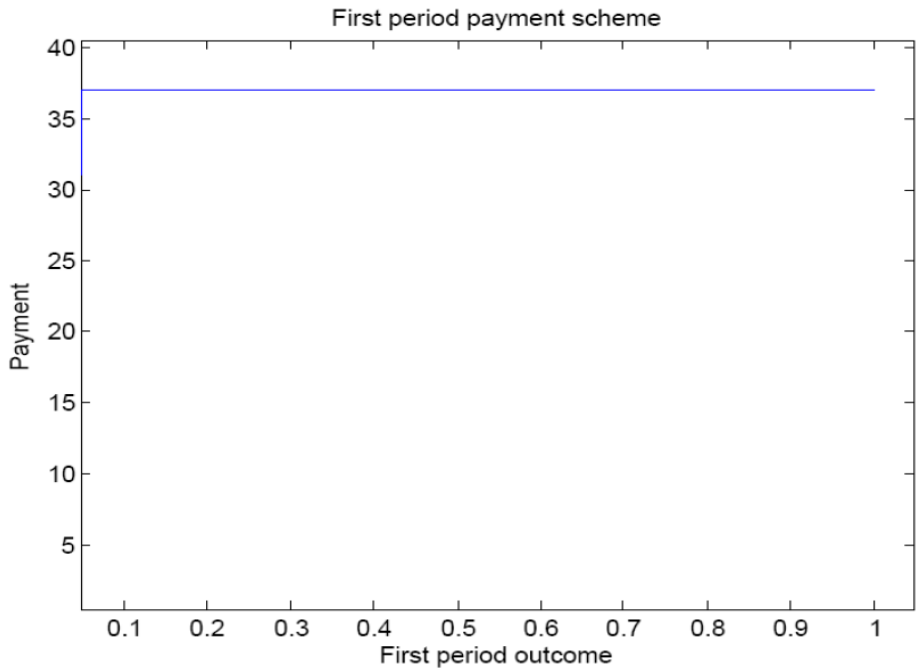


Figure 5: Case 1. Second Period Payment Scheme

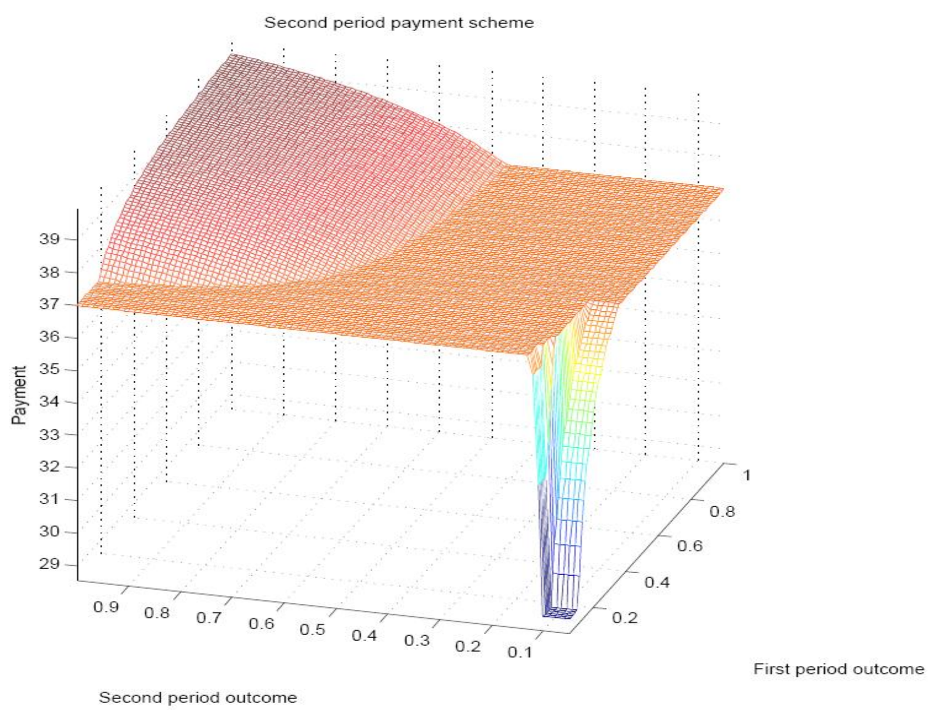
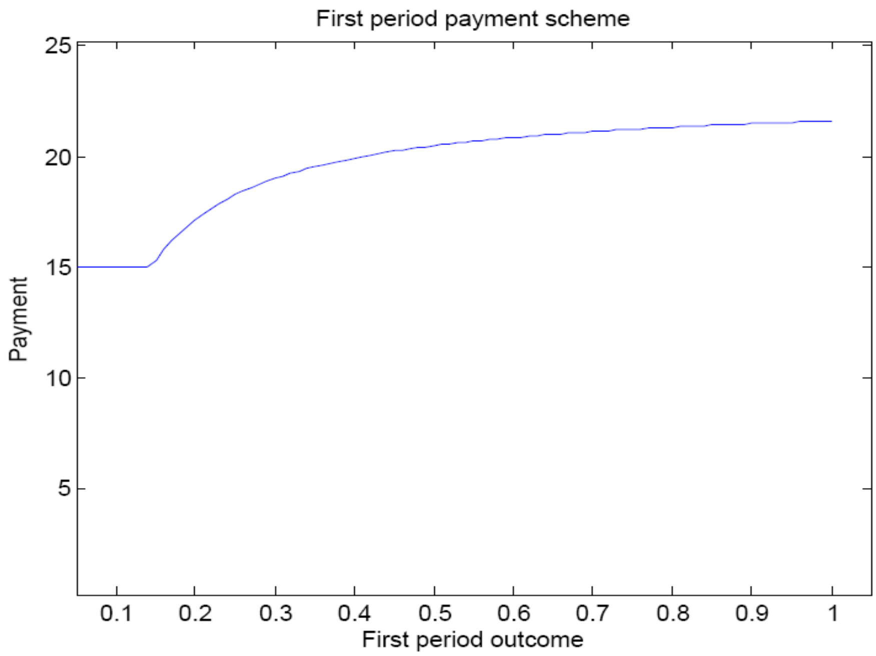


Figure 6: Case 2. First Period Payment Scheme



5.2 Case 2: First period payment greater or equal than R_0

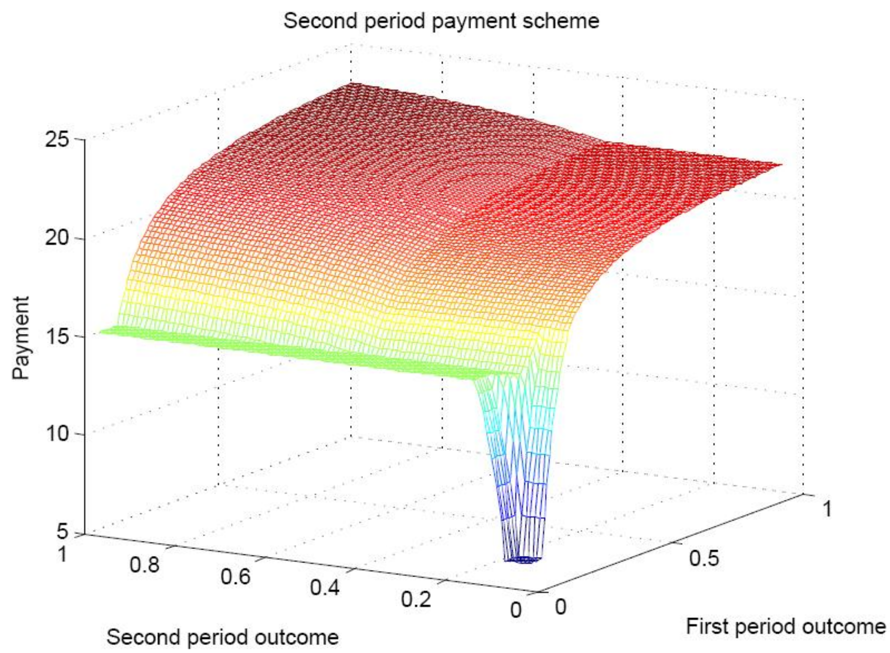
The next example illustrates a case in which the first period payment reaches the reference for low values of the first period outcome (see Figure 5.2). The second period payment scheme is shown in Figure 5.2. The values of parameters used in this simulation are given in the following table, along with the multipliers

ℓ_0	ℓ_1	a_H	a_L	$1/U'(R_0)$	λ	μ_0	μ_1
1	1	1	0.1	15	30.1	1	1

5.3 Case 3: Second period payment over reference for all outcomes

This example illustrates a case in which payments are over the reference for all outcomes in the first and second periods (see Figure 5.3). The parameters used are the same as in Case 2, except that the cost for the principal of inducing participation is higher, which is reflected in a higher λ (Property 9). The second period payment scheme is contingent on outcomes obtained on both the first and second periods as shown in Figure 5.3. According to Property 8, in this case the agent does not have incentives to save for any possible outcome realization in the first period. The payment schemes shown in Figures 5.3 and 5.3 are efficient,

Figure 7: Case 2. Second Period Payment Scheme



renegotiation-proof and they implement the high level of effort in both periods if the agent is restricted to borrow. The values of the parameters used in this simulation are given in the following table, along with the multipliers.

ℓ_0	ℓ_1	a_H	a_L	$1/U'(R_0)$	λ	μ_0	μ_1
1	1	1	0.1	15	40.1	1	1

6 Conclusions and Final Remarks

In this paper we extend the dynamic moral hazard principal-agent model first derived by Rogerson (1985) to allow for an agent who is loss averse and whose reference updates according to previous consumption. We analyze the optimal contracting problem under two scenarios. In the first one the agent has no access to credit markets and thus needs to rely on the principal to transfer resources over time. In the second one the agent has access to credit but the principal can monitor her savings.

When the agent has no access to credit markets and is forced to consume her earnings, we find that the optimal payment scheme can have flat segments at

Figure 8: Case 3. First Period Payment Scheme

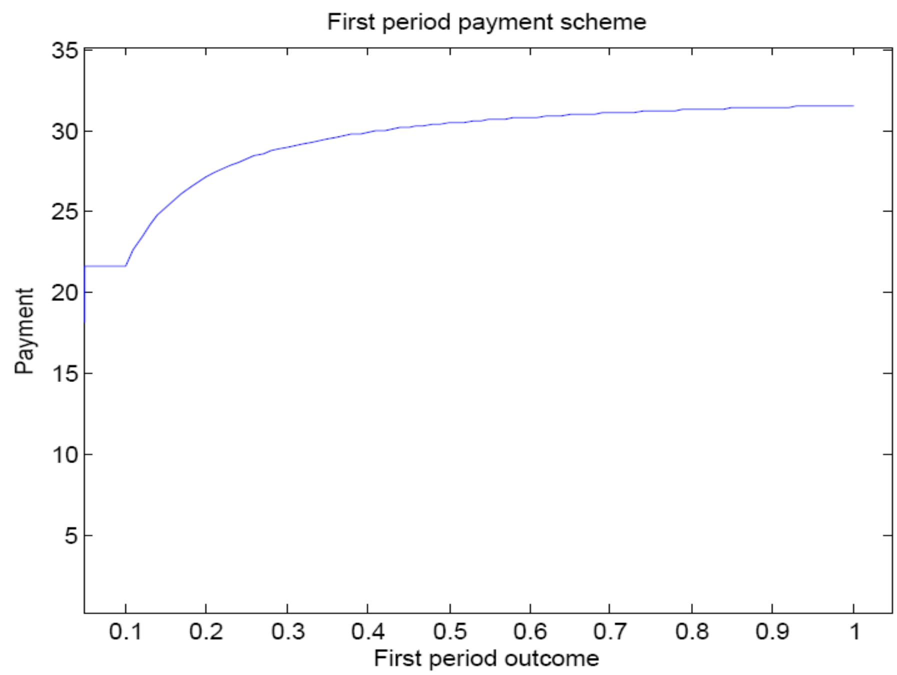
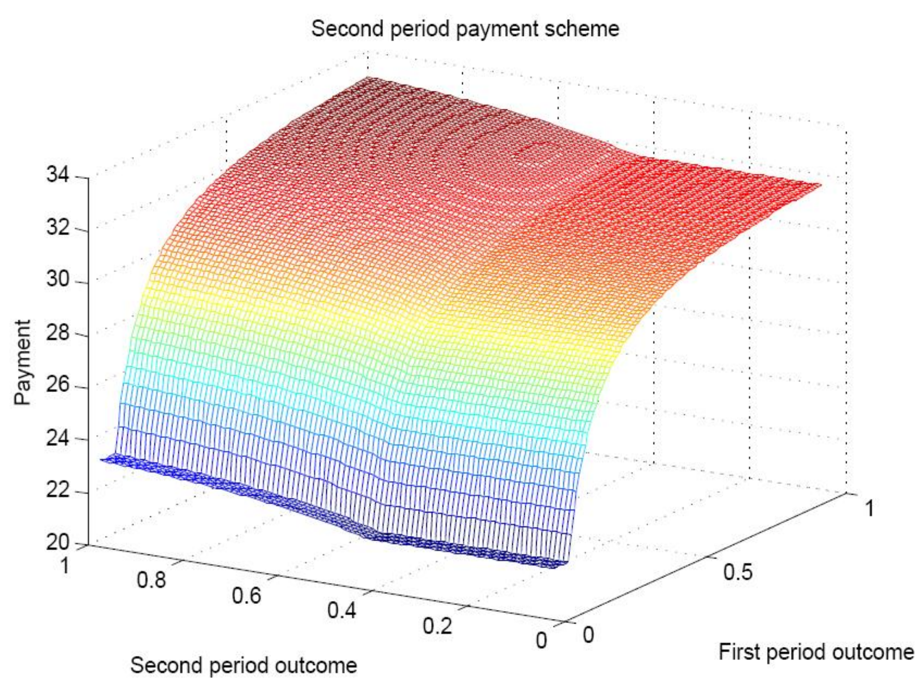


Figure 9: Case 3. Second Period Payment Scheme



the reference; i.e., the wage may be insensitive to outcomes in an interval. This property implies, in turn, that there is a positive probability of observing constant wages over time, even though the contract scheme displays memory –it depends on the full history of outcomes–. Moreover, the model predicts a "status quo bias" whenever the agent is allowed to borrow or save after the outcome is realized, whereas in the canonical model, if anything, he would like to save for future periods. In other words, there is a gap between the interest rates at which the agent is willing to lend and borrow ex-post.

We also show that although the optimal contract scheme is renegotiation-proof, it cannot be implemented by a sequence of spot contracts because of the dynamic update process we assume for the reference. However, when the agent has access to the credit market and the principal can monitor savings, then the long-term optimal contract is spot contractible.

In sum, this paper shows that many of the properties of the classical model hold under our assumptions –consumption smoothing, memory and renegotiation proofness. Moreover, our model predicts new features of the optimal scheme that may help explain many of the facts described by the empirical literature on labor contracts. First it might explain why observed contracts are fairly simpler than those predicted by the classical theory. In fact, many authors have called attention to the simplicity of actual contracts compared to those derived by the theoretical literature (Chiappori and Salanie, 2000; Salanie, 2003; Bolton and Dewatripont, 2005). Our model might also explain why real wages are persistent over time (Dickens et al., 2007) and why incentives tend to be deferred to the future (Baker, Jensen and Murphy, 1988; Baker, Gibbs and Holmstrom, 1994). That is, our model predicts features of the optimal contract that are better in line with the empirical findings, while at the same time conserving many of the properties predicted by the classical model.

Future research should analyze the robustness of our results to a number of assumptions. In particular, a related literature on loss averse preferences has assumed different reference formation processes. In addition, an interesting generalization of our model is to allow for a loss averse principal. In this case, we expect the agent and the principal to protect each other against losses whenever the other party's reference point is reached.

7 Appendix

7.1 $\tilde{U_0}$ expressed in terms of U and ℓ_0

Let's prove that $\tilde{U_0}$ can be expressed as

$$\tilde{U_0}(c_0, R_0) = U(c_0) - \ell_0\theta(c_0, R_0)\left(U(R_0) - U(c_0)\right)$$

Define

$$U(c_0) = \left\{ \begin{array}{cc} \tilde{U_0}(c_0, R_0) & \text{if } c_0 \geq R_0 \\ \tilde{U_0}(c_0, R_0)L + (1-L)\tilde{U}(R_0, R_0) & \text{if } c_0 \leq R_0 \end{array} \right.$$

where $L = \frac{\tilde{U_0}'^{+}(R_0)}{\tilde{U_0}'^{-}(R_0)}$ with $\tilde{U_0}'^{+}(R_0)$ and $\tilde{U_0}'^{-}(R_0)$ denoting the right and left derivative of $\tilde{U_0}$ at R_0.

U as defined is continuous and differentiable with $U'(R_0) = \tilde{U_0}'^{+}(R_0)$. It is concave since we assume $\tilde{U_0}(c_0, R_0)$ concave in c_0.

Define $\ell_0 = \frac{1-L}{L}$, then if $c_0 \leq R_0$ we have

$$U(c_0) - \ell_0 \theta(c_0, R_0) \left(U(R_0) - U(c_0)\right) =$$

$$\tilde{U_0}(c_0, R_0)L + (1-L)\tilde{U}(R_0, R_0) - \frac{1-L}{L}\left(\tilde{U_0}(R_0, R_0) - \tilde{U_0}(c_0, R_0)L - (1-L)\tilde{U}(R_0, R_0)\right) =$$

$$\tilde{U_0}(c_0, R_0)L + (1-L)\tilde{U}(R_0, R_0) - \frac{1-L}{L}L\left(\tilde{U_0}(R_0, R_0) - \tilde{U_0}(c_0, R_0)\right) = \tilde{U_0}(c_0, R_0)$$

7.2 Proof of property 1

Let's see that $h_i(v_0, v_1, \ldots, v_i)$ is strictly convex. Since U is strictly increasing we can write, $h_i(v_0, v_1, \ldots, v_i) = U^{-1}(U(h_i(v_0, v_1, \ldots, v_i)))$, we prove that $U(h_i(v_0, v_1, \ldots, v_i))$ is strictly convex and increasing and we conclude by the strict convexity of U^{-1} (implied by the strict concavity of U).[14] Let $\mathbf{v}^i = (v_0, \ldots, v_i)$ and $\mathbf{v}'^i = (v'_0, \ldots, v'_i)$ be two utility provision vectors. Denote $\mathbf{v}^{i-1} = (v_0, v_1, \ldots, v_{i-1})$. By the definition of convexity, we need to prove,

$$U(h_i(\lambda(v_0, \ldots, v_i) + (1-\lambda)(v'_0, \ldots, v'_i))) < \lambda U(h_i(v_0, v_1, \ldots, v_i)) + (1-\lambda)U(h_i(v'_0, v'_1, \ldots, v'_i)) \quad \forall \lambda \in (0,1) \tag{15}$$

Note that for $i = 0$, by (4) $U(h(v_0))$ is linear by parts, increasing and convex (derivative for $v_0 < R_0$ is $\frac{1}{1+\ell_0} < 1$ and for $v_0 > R_0$ it is 1.). Let's prove (15) assuming true for $i-1$.

If $\lambda v_i + (1-\lambda)v'_i < U(h_{i-1}(\lambda(v_0, \ldots, v_{i-1}) + (1-\lambda)(v'_0, \ldots, v'_{i-1})))$ the utility provision of period i of the convex combination is in the loss area and we have

$$\begin{aligned} U(h_i(\lambda \mathbf{v}^i + (1-\lambda)\mathbf{v}'^i)) &= \left(\frac{\lambda v_i + (1-\lambda)v'_i + \ell_i U(h_{i-1}(\lambda \mathbf{v}^{i-1} + (1-\lambda)\mathbf{v}'^{i-1}))}{1+\ell_i}\right) \\ &\leq \lambda \left(\frac{v_i + \ell_i U(h_{i-1}(\mathbf{v}^{i-1}))}{1+\ell_i}\right) + \\ &\quad (1-\lambda)\left(\frac{v'_i + \ell_i U(h_{i-1}(\mathbf{v}'^{i-1}))}{1+\ell_i}\right) \\ &\leq \lambda U(h_i(\mathbf{v}^i)) + (1-\lambda)U(h_i(\mathbf{v}'^{i-1})) \end{aligned}$$

[14]Note that the composition of a convex increasing function with a convex function is convex.

The first inequality is implied by the induction hypothesis and the second is justified noting that if $v_i > U(h_{i-1})$ then $U(h_i(\mathbf{v}^i)) = v_i > \left(\frac{v_i+\ell_i U(h_{i-1}(\mathbf{v}^{i-1}))}{1+\ell_i}\right)$ and if $v_i \leq U(h_{i-1})$ then $U(h_i(\mathbf{v}^i)) = \left(\frac{v_i+\ell_i U(h_{i-1}(\mathbf{v}^{i-1}))}{1+\ell_i}\right)$.

A similar argument proves (15) for the case in which $\lambda v_i+(1-\lambda)v_i' \geq U(h_{i-1}(\lambda(v_0,\ldots,v_{i-1})+(1-\lambda)(v_0',\ldots,v_{i-1}')))$.

$$U(h_i(\lambda(v_0,\ldots,v_i)+(1-\lambda)(v_0',\ldots,v_i'))) \quad = \quad \lambda v_i+(1-\lambda)v_i' \leq \lambda U(h_i(\mathbf{v}^i))+(1-\lambda)U(h_i(\mathbf{v'}^{i-1}))$$

Where the last inequality is justified by noting $v_i < U(h_{i-1}(\mathbf{v}^{i-1}))$ implies $v_i \leq \left(\frac{v_i+\ell_i U(h_{i-1}(\mathbf{v}^{i-1}))}{1+\ell_i}\right) = U(h_i(\mathbf{v}^i))$.

Finally, the constraints are linear in $v_i(x_0,x_1,\ldots,x_i)$ for each i and, therefore, convexity of the feasible set is straightforward

7.3 Proof of property 2

We have

$$h_i(v_0,v_1,\ldots,v_i) = U^{-1}(U(h_i(v_0,v_1,\ldots,v_i)))$$

where

$$U(h_i(v_0,v_1,\ldots,v_i)) = \begin{cases} v_i & \text{if } v_i \geq U(h_{i-1}(v_0,v_1,\ldots,v_{i-1}) \\ \frac{v_i+\ell_i U(h_{i-1}(v_0,v_1,\ldots,v_{i-1}))}{1+\ell_i} & \text{if } v_i < U(h_{i-1}(v_0,v_1,\ldots,v_{i-1}) \end{cases} \tag{16}$$

By Proposition 4.2.5 in **?** we know that[15]

$$\partial h_i(v_0,v_1,\ldots,v_i) = \left(U^{-1}\right)'\left(\left(U\circ h_i\right)(v_0,v_1,\ldots,v_i)\right)\right)\cdot\partial\left(\left(U\circ h_i\right)(v_0,v_1,\ldots,v_i)\right)$$

.

Now, note that from (16) we have $U(h_i(v_0,v_1,\ldots,v_i)) = F_i((U\circ h_{i-1})(v_0,v_1,\ldots,v_{i-1}),v_i)$ where $F_i(x,y) = \begin{cases} y & \text{if } y \geq x \\ \frac{y+\ell_i x}{1+\ell_i} & \text{if } y < x \end{cases}$

Let $(d_0,\ldots,d_{i-1}) \in \partial\left(U\circ h_{i-1}\right)(v_0,v_1,\ldots,v_{i-1})$ and $(\tilde{d}_0,\tilde{d}_1) \in \partial F_i((U\circ h_{i-1})(v_0,v_1,\ldots,v_{i-1}),v_i)$ let's see that that $(d_0\cdot\tilde{d}_0, d_1\cdot\tilde{d}_0,\ldots,d_{i-1}\cdot\tilde{d}_0,\tilde{d}_1) \in \partial\left(U\circ h_i\right)(v_0,v_1,\ldots,v_i)$. In fact, we have

$$\begin{aligned} (U\circ h_i)(v_0+\alpha_0,v_1+\alpha_1,\ldots,v_i+\alpha_i) \quad &= \\ F_i((U\circ h_{i-1})(v_0+\alpha_0,\ldots,v_{i-1}+\alpha_{i-1}),v_i+\alpha_i) \quad &\geq \quad F_i((U\circ h_{i-1})(v_0,v_1,\ldots,v_{i-1})+d_0\alpha_0+\cdots+d_{i-1}\alpha_{i-1},v_i+\alpha_i) \\ &\geq \quad F_i((U\circ h_{i-1})(v_0,v_1,\ldots,v_{i-1}),v_i)+d_0\tilde{d}_0\alpha_0+\cdots+d_{i-1}\tilde{d}_0\alpha_{i-1}+\tilde{d}_1\alpha_i \end{aligned}$$

[15] A vector $d \in I\!R^n$ is a subgradient of f at a point $x \in I\!R^n$, denoted $d \in \partial f(x)$, if

$$f(z) \geq f(x)+(z-x)'d \qquad \forall z \in I\!R^n$$

Where the first inequality is due to $(d_0, \dots, d_{i-1}) \in \partial (U \circ h_{i-1}) (v_0, v_1, \dots, v_{i-1})$ and F_i increasing in its first variable. The second inequality is implied by

$$(\tilde{d_0}, \tilde{d_1}) \in \partial F_i((U \circ h_{i-1}) (v_0, v_1, \dots, v_{i-1}), v_i)$$

Let's now see that the reverse is also true. That is, we show that an element of $\partial (U \circ h_i) (v_0, v_1, \dots, v_i)$ can be written as $(d_0 \cdot \tilde{d_0}, d_1 \cdot \tilde{d_0}, \dots, d_{i-1} \cdot \tilde{d_0}, \tilde{d_1})$ with $(d_0, \dots, d_{i-1}) \in \partial (U \circ h_{i-1}) (v_0, v_1, \dots, v_{i-1})$ and $(\tilde{d_0}, \tilde{d_1}) \in \partial F_i((U \circ h_{i-1}) (v_0, v_1, \dots, v_{i-1}), v_i)$

Let's compute $\partial F_i(x, y)$. If $x \neq y$ F_i is differentiable and therefore its subgradient set coincides with the derivative.

Otherwise, $y = x$ and the elements of the subgradient set of $\partial F_i(x, y)$ will be the pairs $(\tilde{d_0}, \tilde{d_1})$ such that,

$$F_i(x + \alpha_0, y + \alpha_1) \geq F(x, y) + \alpha_0 \tilde{d_0} + \alpha_1 \tilde{d_1} \qquad \forall \alpha_0, \alpha_1 \in \mathbb{R} \tag{17}$$

If $\alpha_0 \leq \alpha_1$ then $x + \alpha_0 \leq y + \alpha_1$ and (17) becomes

$$y + \alpha_1 \geq y + \alpha_0 \tilde{d_0} + \alpha_i \tilde{d_1}$$

$$\iff (1 - \tilde{d_1})\alpha_1 \geq \tilde{d_0}\alpha_0$$

which is true for all $\alpha_1 \geq \alpha_0$ if and only if $(1 - \tilde{d_1}) = \tilde{d_0} > 0$. [16]

If $\alpha_0 > \alpha_1$ then $x + \alpha_0 > y + \alpha_1$ and (17) becomes

$$\frac{y + \alpha_1 + \ell_i(x + \alpha_0)}{(1 + \ell_i)} \geq y + \alpha_0 \tilde{d_0} + \alpha_1 \tilde{d_1}$$

$$\iff \left(\frac{\ell_i}{1 + \ell_i} - \tilde{d_0} \right) \alpha_0 \geq \left(\tilde{d_1} - \frac{1}{1 + \ell_i} \right) \alpha_1$$

which is true for all $\alpha_0 < \alpha_1$ if and only if $\left(\frac{\ell_i}{1+\ell_i} - \tilde{d_0} \right) = \left(\tilde{d_1} - \frac{1}{1+\ell_i} \right) > 0$. Therefore, summarizing we have established that

$$(\tilde{d_0}, \tilde{d_1}) \in \partial F(x, x) \implies \tilde{d_0} \in \left[0, \frac{\ell_i}{1 + \ell_i} \right], \tilde{d_1} \in \left[\frac{1}{1 + \ell_i}, 1 \right] \text{ and } \tilde{d_0} = 1 - \tilde{d_1}$$

And therefore, we can write,

$$\partial F_i(x, y) = \left\{ \left(\frac{k_i \ell_i}{1 + \ell_i k_i}, \frac{1}{1 + \ell_i k_i} \right); \quad \text{where} \quad k_i(x_0, x_1, \dots, x_i) \in \left\{ \begin{array}{ll} \{1\} & \text{if } \omega_i(x_0, x_1, \dots, x_i) < R_i \\ [0, 1] & \text{if } \omega_i(x_0, x_1, \dots, x_i) = R_i \\ \{0\} & \text{otherwise} \end{array} \right\} \right.$$

Suppose that $(\bar{d_0}, \dots, \bar{d_i}) \in \partial (U \circ h_i) (v_0, \dots, v_i)$, let's see that $(\bar{d_0}, \bar{d_1}, \dots \bar{d}_{i-1}, \bar{d_i}) = (d_0 \cdot \tilde{d_0}, d_1 \cdot \tilde{d_0}, \dots, d_{i-1} \cdot \tilde{d_0}, \tilde{d_1})$ for some vectors $(d_0, \dots, d_{i-1}) \in \partial (U \circ h_{i-1}) (v_0, \dots, v_{i-1})$ and

[16] Note that α_0 and α_1 can take negative values.

$(\tilde{d}_0, \tilde{d}_1) \in \partial F_i((U \circ h_{i-1})(v_0, \dots, v_{i-1}), v_i)$.

Note that by the $F_i((U \circ h_{i-1})(v_0, \dots, v_{i-1}), v_i+\alpha_i) \geq F_i((U \circ h_{i-1})(v_0, \dots, v_{i-1}), v_i)+ \alpha_i \bar{d}_i \quad \forall \alpha_i$ implies we must have $\bar{d}_i = \frac{1}{1+\ell_i k_i}$ with k_i defined by (6) (subgradient set in one variable). We know that in points in which F_i is differentiable its subgradient set coincides with the derivative which will be $(0, 1)$ if $(U \circ h_{i-1})(v_0, \dots, v_{i-1}) < v_i$ and $(\frac{\ell_i}{1+\ell_i}, \frac{1}{1+\ell_i})$ if $(U \circ h_{i-1})(v_0, \dots, v_{i-1}) > v_i$. Therefore, from Proposition 4.2.5 **?** we must have that $\partial U(h_i(v_0, \dots, v_i)) = \big((1-\bar{d}_i) \cdot \partial (U \circ h_{i-1})(v_0, \dots, v_{i-1}), \bar{d}_i\big)$.

If F is not differentiable we have $(U \circ h_{i-1})(v_0, \dots, v_{i-1}) = v_i$. Let $(\alpha_0, \alpha_1, \dots, \alpha_{i-1}) \in \mathbb{R}^i$, we define $\hat{\alpha}_i = (U \circ h_{i-1})(v_0+\alpha_0, \dots, v_{i-1}+\alpha_{i-1}) - v_i$.

Note that $\partial U(h_i(v_0, \dots, v_i)) = F((U \circ h_{i-1})(v_0+\alpha_0, \dots, v_{i-1}+\alpha_{i-1}), v_i+\hat{\alpha}_i) = v_i + \hat{\alpha}_i$. In fact, this is straightforward if $\hat{\alpha}_i \geq 0$. If $\hat{\alpha}_i < 0$ then $F((U \circ h_{i-1})(v_0+ \alpha_0, \dots, v_{i-1}+\alpha_{i-1}), v_i+\hat{\alpha}_i) = \frac{v_i+\hat{\alpha}_i+\ell_i(\hat{\alpha}_i+v_i)}{(1+\ell_i)} = v_i + \hat{\alpha}_i$.

Since $(\bar{d}_0, \dots, \bar{d}_i)$ in $\partial U(h_i(v_0, \dots, v_i))$ we have

$$\begin{aligned}
& v_i + \hat{\alpha}_i \geq v_i + \bar{d}_0\alpha_0 + \cdots + \bar{d}_{i-1}\alpha_{i-1} + \bar{d}_i\hat{\alpha}_i \\
\Longrightarrow\quad & \hat{\alpha}_i(1-\bar{d}_i) \geq \bar{d}_0\alpha_0 + \cdots + \bar{d}\alpha_{i-1} \\
\Longrightarrow\quad & (U \circ h_{i-1})(v_0+\alpha_0, \dots, v_{i-1}+\alpha_{i-1}) - (U \circ h_{i-1})(v_0, \dots, v_{i-1}) \geq \left(\bar{d}_0\alpha_0 + \cdots + \bar{d}_{i-1}\alpha_{i-1}\right)\frac{1}{(1-\bar{d}_i)} \\
\Longrightarrow\quad & (\bar{d}_0, \dots, \bar{d}_{i-1})\frac{1}{(1-\bar{d}_i)} \in \partial (U \circ h_{i-1})(v_0, \dots, v_{i-1})
\end{aligned}$$

We conclude for $(d_0, \dots, d_{i-1}) = (\bar{d}_0, \dots, \bar{d}_i)\frac{1}{(1-\bar{d}_i)} \in \partial (U \circ h_{i-1})(v_0, \dots, v_{i-1})$ and $(\tilde{d}_0, \tilde{d}_1) = \big((1-\bar{d}_i), \bar{d}_i\big) \in \partial F((U \circ h_{i-1})(v_0, \dots, v_{i-1}), v_i)$ that $(\bar{d}_0, \bar{d}_1, \dots \bar{d}_{i-1}, \bar{d}_i) = (d_0 \cdot \tilde{d}_0, d_1 \cdot \tilde{d}_0, \dots, d_{i-1} \cdot \tilde{d}_0, \tilde{d}_1)$.

We may now deduce inductively $\partial U(h_i(v_0, \dots, v_i))$. For the functions k_i defined by (6) we have

$$\partial U(h_0(v_0)) = \left\{\frac{1}{1+k_0\ell_0}\right\}$$

therefore

$$\partial U(h_1(v_0, v_1)) = \left(\frac{k_1\ell_1}{1+\ell_1 k_1} \cdot \frac{1}{1+k_0\ell_0}, \frac{1}{1+\ell_1 k_1}\right)$$

and

$$\partial U(h_2(v_0, v_1, v_2)) = \left(\frac{k_2\ell_2}{1+\ell_2 k_2} \cdot \frac{k_1\ell_1}{1+\ell_1 k_1} \cdot \frac{1}{1+k_0\ell_0}, \frac{k_2\ell_2}{1+\ell_2 k_2} \cdot \frac{1}{1+k_1\ell_1}, \frac{1}{1+\ell_2 k_2}\right)$$

and inductively, (5) is obtained.

$$\partial h_i(v_0, v_1, \dots, v_i) = \left(\frac{1}{U'(\omega_i)}\left(\prod_{t=j+1}^{i} \frac{k_t(x_0, x_1, \dots, x_t)\ell_t}{1+k_t(x_0, x_1, \dots, x_t)\ell_t}\right)\frac{1}{1+k_j(x_0, x_1, \dots, x_j)\ell_j}\right)_{j=0}^{i} \tag{18}$$

7.4 Proof of property 3

We assume that the principal is looking for optimal utility provisions $v = (v_i(\cdot))_{i=0}^T$ in the function spaces $\left(L^1([\underline{x}_0, \bar{x_0}] \times [\underline{x}_1, \bar{x_1}] \times \cdots \times [\underline{x}_i, \bar{x_i}])\right)_{i=0}^T$. Consider the following functions

$$f_0(v_0, v_1, \ldots, v_T) = \sum_{i=0}^T \delta^i \mathbb{E}\left(x_i - h_i(v_0(x_0), v_1(x_0, x_1), \ldots, v_i(x_0, x_1, \ldots, x_i)) | a_0, a_1, \ldots, a_i\right)$$

$$g_0(v_0, v_1, \ldots, v_T) = \sum_{i=0}^T \delta^i \left(\mathbb{E}\left(v_i(x_0, x_1, \ldots, x_i) | a_0, a_1, \ldots, a_i\right) - \psi_i(a_i)\right) - U^*$$

$$h_i(v_0, v_1, \ldots, v_T) = \sum_{j=i}^T \delta^j \left(\Delta_{a_i} \mathbb{E}\left(v_j(x_0, x_1, \ldots, x_j) | a_i, \ldots, a_j\right)\right) - \Delta\varphi_a$$

For $\tilde{u}_0, u_0 \in \mathbb{R}$, $u_i : \mathbb{R}^i \to \mathbb{R} \in L^1(\mathbb{R}^i)$ for $i \in \{1, \ldots, T\}$ endowed with the measure μ^i induced by the outcome probability. Let $u = (\tilde{u}_0, u_0, u_1, \ldots u_T) \in U = \mathbb{R}^2 \times L^1(\mathbb{R}) \times L^1(\mathbb{R}^2) \times \cdots \times L^1(\mathbb{R}^T)$ we define

$$F(v, u) = \begin{cases} f_0(v_0, v_1, \ldots, v_T) & \text{if } g_0(v_0, v_1, \ldots, v_T) \geq \tilde{u}_0, h_i(v_0, v_1, \ldots, v_T) \geq u_i(x_0, x_1, \ldots, x_{i-1}) \\ -\infty & \text{otherwise.} \end{cases}$$

$-F(v, u)$ is closed in u since the sets $\{u | F(v, u) \geq \alpha\}$ are closed for all $\alpha \in \mathbb{R}$ by continuity of f_0, g_0 and h_i for $i \in \{0, \ldots, T\}$. $-F(v, u)$ is also convex in u.[17]

Following **?**, equation 4.2, the Lagrangian function K is defined as

$$K(v, y) = \sup\{F(v, u) + < u, y > | u \in U\}$$

with $y = (\tilde{y}_0, y_0, y_1, \ldots, y_T) \in Y = \mathbb{R}^2 \times (L^\infty)^{T+1}$ and

$$< u, y >= \tilde{y}_0 \tilde{u}_0 + y_0 u_0 + \mathbb{E}_{\mu^1}(y_1 u_1) + \mathbb{E}_{\mu^2}(y_2 u_2) + \ldots + \mathbb{E}_{\mu^T}(y_T u_T)$$

In our case K is equal to

$$K(\cdot, y) = \begin{cases} f_0 + g_0 \tilde{y}_0 + h_0 y_0 + \mathbb{E}_{\mu^1}(y_1 h_1) + \mathbb{E}_{\mu^2}(y_2 h_2) + \ldots + \mathbb{E}_{\mu^T}(y_T h_T) & \text{if } y \geq 0 \\ +\infty & \text{if } y < 0 \end{cases}$$

Where we say $y \geq 0$ if all components are positive almost everywhere. Let $y = (\lambda, \mu_0, \ldots, \mu_T)$, the Lagrangian becomes.

$$\begin{aligned} K(v, y) = & \sum_{i=0}^T \delta^i \mathbb{E}\left(x_i - h_i(v_0(x_0), v_1(x_0, x_1), \ldots, v_i(x_0, x_1, \ldots, x_i)) | a_0, a_1, \ldots, a_i\right) + \\ & \lambda \left(\sum_{i=0}^T \delta^i \left(\mathbb{E}\left(v_i(x_0, x_1, \ldots, x_i) | a_0, a_1, \ldots, a_i\right) - \psi_i(a_i)\right) - U^* \right) + \\ & \sum_{i=0}^T \left(\sum_{j=i}^T \delta^j \left(\Delta_{a_i} \mathbb{E}\left(\mu_i(x_0, x_1, \ldots, x_{i-1}) \cdot v_j(x_0, x_1, \ldots, x_j) | a_0, a_1 \ldots, a_j\right)\right) \right) \end{aligned}$$

[17] We need to check $F(v, u)$ is concave in u. We need to check that for $u^1, u^2 \in U$, if $F(x, u^1) \neq -\infty$, $F(x, u^2) \neq -\infty$ then $F(x, \alpha u^1 + (1 - \alpha) u^2) \neq \infty$. This follows since $\alpha u_j^1 + (1 - \alpha) u_j^2 \geq \min\{u_j^1, u_j^2\}$ for $j \in \{1, \ldots, T + 2\}$.

Where we denote

$$\Delta_{a_i}\mathbb{E}\left(\mu_i(x_0,x_1,\ldots,x_{i-1})\cdot v_j(x_0,x_1,\ldots,x_j)|a_0,a_1,\cdots a_i,\cdots,a_j\right) =$$
$$\mathbb{E}\left(\mu_i(x_0,x_1,\ldots,x_{i-1})\cdot v_j(x_0,x_1,\ldots,x_j)|a_0,a_1,\cdots,a_i=a_{H,}\cdots,a_j\right) \quad +$$
$$-\mathbb{E}\left(\mu_i(x_0,x_1,\ldots,x_{i-1})\cdot v_j(x_0,x_1,\ldots,x_j)|a_0,a_1,\cdots,a_i=a_L,\cdots,a_j\right) \quad =$$
$$\int \mu_i(x_0,x_1,\ldots,x_{i-1})\cdot v_j(x_0,x_1,\ldots,x_j)f^0(x_0|a_0)\cdots f^i_{a_i}(x_i|a_i)\cdots f^j(x_j|a_j)dx_0\cdots dx_i\cdots dx_j$$

We say that y belongs to the subgradient set of a function $\varphi : V \to \mathbb{R}$ at $v \in V$, which we denote $y \in \partial\varphi(v)$ if

$$\varphi(v') \geq \varphi(v) + < v' - v, y > \qquad \forall v' \in U$$

We say that $(\bar{v},\bar{y})$ satisfies the Kuhn-Tucker condition if $0 \in \partial_v(-K(\bar{v},\bar{y}))$ and $0 \in \partial_y(-K(\bar{v},\bar{y}))$.

Since $-F(v,u)$ is closed and convex in u from Theorem 15 in **?** we know that if $(\bar{v},\bar{y})$ satisfies the Kuhn-Tucker condition then $\bar{v}$ solves the principal problem given by equations (3) through (4). Note that $0 \in \partial_y K(\bar{v},\bar{y})$ if and only if $K(\bar{v},y') \geq K(\bar{v},\bar{y}) \quad \forall y' \in Y$ which is equivalent to ask that the constraints (PC') through (4) be satisfied. If one of the constraints is not satisfied at $\bar{v}$ then $K(\bar{v},\cdot)$ is unbounded. Thus we only need to verify that $0 \in \partial_y K(\bar{v},\bar{y})$.

$K(v,y)$, although non-differentiable, is concave in $v = (v_i(\cdot))_i$ and the set of constraints is convex (Property 1), therefore a necessary and sufficient condition for a wage schedule to be optimal is that the subgradient set of $-K(v,y)$ (denoted $\partial\left(-K(v,y)\right)$) contains 0. Since from Property 1 we have that $\mathbb{E}\left(h_i(v_0(x_0),v_1(x_0,x_1),\ldots,v_i(x_0,x_1,\ldots,x\right.$ are convex in $(v_i(\cdot))_i$, from proposition 4.2.4 in **?**. [18]

$$\begin{aligned}\partial\left(-K(v,y)\right) = &\sum_{i=0}^{T}\delta^i\partial\mathbb{E}\left(h_i(v_0(x_0),v_1(x_0,x_1),\ldots,v_i(x_0,x_1,\ldots,x_i))|a_0,a_1,\ldots,a_i\right) + \\ &-\lambda\left(\sum_{i=0}^{T}\delta^i\partial\mathbb{E}\left(v_i(x_0,x_1,\ldots,x_i)|a_0,a_1,\ldots,a_i\right)\right) + \\ &-\sum_{i=0}^{T}\left(\sum_{j=i}^{T}\delta^j\partial\left(\Delta_{a_i}\mathbb{E}\left(\mu_i(x_0,x_1,\ldots,x_{i-1})\cdot v_j(x_0,x_1,\ldots,x_j)|a_i,\ldots,a_j\right)\right)\right)\end{aligned}$$

From Theorem 22 of **?**, we know that a subgradient set of $-K$ is the expectation of the subgradient set of the integrand.

[18] Note that the subgradient set of a constant function is equals to zero

$$\partial\left(-K(v,y)\right) = \sum_{i=0}^{T} \delta^i \mathbb{E}\left(\partial h_i(v_0(x_0), v_1(x_0,x_1), \ldots, v_i(x_0,x_1,\ldots,x_i))|a_0,a_1,\ldots,a_i\right) +$$
$$- \lambda\left(\sum_{i=0}^{T} \delta^i \mathbb{E}\left(\partial v_i(x_0,x_1,\ldots,x_i)|a_0,a_1,\ldots,a_i\right)\right) +$$
$$- \sum_{i=0}^{T}\left(\sum_{j=i}^{T} \delta^j \left(\Delta_{a_i}\mathbb{E}\left(\mu_i(x_0,x_1,\ldots,x_{i-1})\cdot \partial v_j(x_0,x_1,\ldots,x_j)|a_i,\ldots,a_j\right)\right)\right)$$

Therefore from Property 2 making $\partial\left(-K(v,y)\right)$ equal 0 by components corresponds to,

$$0 = \sum_{i=j}^{T} \delta^{i-j}\mathbb{E}\left(\frac{1}{U'(\omega_i(x_0,x_1,\ldots,x_i))}\left(\prod_{t=j+1}^{i}\frac{k_t(x_0,x_1,\ldots,x_t)\ell_t}{1+k_t(x_0,x_1,\ldots,x_t)\ell_t}\right)\frac{1}{1+k_j(x_0,x_1,\ldots,x_j)\ell_j}\cdot g(x_0,x_1,\ldots,x_j)\Big|a_0,a_1,\ldots,a_i\right) +$$
$$- \lambda\mathbb{E}\left(g(x_0,x_1,\ldots,x_j)|a_0,a_1,\ldots,a_j\right) +$$
$$- \sum_{i=0}^{j}\left(\Delta_{a_i}\mathbb{E}\left(\mu_i(x_0,x_1,\ldots,x_{i-1})g(x_0,x_1,\ldots,x_j)|a_0,a_1,\ldots,a_j\right)\right)$$

for every $g \in L^1\left([\underline{x}_0,\bar{x_0}]\times[\underline{x}_1,\bar{x_1}]\times\cdots\times[\underline{x}_j,\bar{x_j}]\right)$, which implies

$$\frac{1}{U'(\omega_j)}\cdot\frac{1}{1+k_j\ell_j} + \sum_{i=j+1}^{T}\delta^{i-j}\mathbb{E}\left(\frac{1}{U'(\omega_i)}\left(\prod_{t=j+1}^{i}\frac{k_t\ell_t}{1+k_t\ell_t}\right)\frac{1}{1+k_j\ell_j}\Big|a_{j+1},\ldots,a_i\right) = \lambda + \sum_{i=0}^{j}\mu_i\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)} \quad (19)$$
$$= \lambda_j + \mu_j\frac{f^j_{a_j}(x_j|a_j)}{f^j(x_j|a_j)}$$

For every $j \in \{1,\ldots,T\}$. Multiplying the equation for $j+1$ by $k_{j+1}\ell_{j+1}$ taking expectation with respect to $f^{j+1}(x_{i+1}|a_{j+1})$ we obtain

$$\mathbb{E}\left(\frac{1}{U'(\omega_{j+1})}\frac{k_{j+1}\ell_{j+1}}{1+k_{j+1}\ell_{j+1}}\Big|a_{j+1}\right) = -\sum_{i=j+2}^{T}\delta^{i-j-1}\mathbb{E}\left(\frac{1}{U'(\omega_i)}\left(\prod_{t=j+2}^{i}\frac{k_t\ell_t}{1+k_t\ell_t}\right)\frac{k_{j+1}\ell_{j+1}}{1+k_{j+1}\ell_{j+1}}\Big|a_{j+1},\ldots,a_i\right)$$
$$\mathbb{E}\left(\left(\lambda_{j+1}+\mu_{j+1}\frac{f^{j+1}_{a_{j+1}}(x_{j+1}|a_{j+1})}{f^{j+1}(x_{j+1}|a_{j+1})}\right)k_{j+1}\ell_{j+1}\Big|a_{j+1}\right)$$

Replacing this last expression in the $j+1$ term of the sum in (19), 7 and 8 are obtained.

7.5 Proof of property 4

The payment scheme must be continuous and non-decreasing. In fact, the multipliers μ_i are strictly positive. If not, then the payment scheme would not depend. It can be seen that the right side of (24) has slope with respect to x_i of at least $\frac{d}{dx_i}\left((1-\delta\ell_{i+1})\mu_i\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}\right)$ and therefore, ω_i must be non-decreasing. A flat segment will arise whenever the payment scheme reaches the first period reference level. (10) and (8) imply that, as x_i increases and reaches the reference if the scheme were to continue increasing it would enter the gain area and the right side of (10) and (8) would jump downwards, therefore contradicting that it increased after reaching the reference income. Something analogous happens when the reference level is reached from above (as x_i decreases), if the optimal scheme were to go below the reference, the optimality characterization would require it to jump upwards. This contradicts that it decreased after reaching the reference from above.

Now, let's see whether the reference will be reached. The following equality must be fulfilled,

$$\frac{1}{U'(\omega_i(x_0,x_1,\ldots,x_i))} = (1+k_i(x_0,x_1,\ldots,x_i)\ell_i)\left(\lambda_i+\mu_i\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}\right)+$$
$$-\delta\ell_{i+1}\int k_{i+1}(x_0,x_1,\ldots,x_{i+1})(\lambda_{i+1}+\mu_{i+1}\frac{f^{i+1}_{a_{i+1}}(x_{i+1}|a_{i+1})}{f^{i+1}(x_{i+1}|a_{i+1})})f^i(x_{i+1}|a_{i+1})dx_{i+1}. \quad (20)$$

The $i-1$ period's payments fulfills the following equation,

$$\lambda_i(1+k_{i-1}(x_0,x_1,\ldots,x_{i-1})\ell_{i-1}) = \frac{1}{U'(\omega_{i-1}(x_0,x_1,\ldots,x_{i-1}))}+$$
$$\delta\ell_i\int k_i(x_0,\ldots,x_i)\left(\lambda_i+\mu_i\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}\right)f^i(x_i|a_i)dx_i.$$

Suppose $\omega_i(x_0,x_1,\ldots,x_i) < \omega_{i-1}(x_0,x_1,\ldots,x_{i-1})\quad \forall x_i$ (except in one point). Then $k_i(x_0,x_1,\ldots,x_i)=1\quad\forall x_i$ and (**??**) becomes $\lambda_i=\frac{1}{U'(\omega_{i-1})(1+k_{i-1}\ell_{i-1}-\delta\ell_i)}$. Therefore, since $\delta\ell_{i+1}\int k_{i+1}(x_0,x_1,\ldots,x_{i+1})(\lambda_{i+1}+\mu_{i+1}\frac{f^i_{a_{i+1}}(x_{i+1}|a_{i+1})}{f^i(x_{i+1}|a_{i+1})})f^i(x_{i+1}|a_{i+1})dx_{i+1}\leq \delta\ell_{i+1}\lambda_{i+1}$ we obtain

$$\frac{1}{U'(\omega_i)}\geq(1+\ell_i-\delta\ell_{i+1})\left(\frac{1}{U'(\omega_{i-1})(1+k_{i-1}\ell_{i-1}-\delta\ell_i)}+\mu_i\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}\right)$$

Therefore, if $(\ell_i-\ell_{i+1})\delta\geq k_{i-1}\ell_{i-1}-\ell_i$ and $\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}>0$ we conclude $\frac{1}{U'(\omega_i)}\geq\frac{1}{U'(\omega_{i-1})}$ which contradicts $\omega_i<\omega_{i-1}$. If $\ell_{i-1}=\ell_i$, $(\ell_i-\ell_{i+1})\delta\geq\ell_{i-1}-\ell_i$ or $k_{i-1}=0$ then $(\ell_i-\ell_{i+1})\delta\geq k_{i-1}\ell_{i-1}-\ell_i$ will be fulfilled.

Suppose $\omega_i(x_0,x_1,\ldots,x_i) > \omega_{i-1}(x_0,x_1,\ldots,x_{i-1})\quad \forall x_i$ (except in one point). Then $k_i(x_0,x_1,\ldots,x_i)=0\quad\forall x_i$ and (**??**) becomes $\lambda_i=\frac{1}{U'(\omega_{i-1})(1+k_{i-1}\ell_{i-1})}$. Therefore, we obtain

$$\frac{1}{U'(\omega_i)}\leq\left(\frac{1}{U'(\omega_{i-1})(1+k_{i-1}\ell_{i-1})}+\mu_i\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}\right)$$

Therefore, if $\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}<0$ we conclude $\frac{1}{U'(\omega_i)}\leq\frac{1}{U'(\omega_{i-1})}$ for which contradicts $\omega_i>\omega_{i-1}$. We conclude that the reference must be reached on an interval for all the cases stated above.

7.6 Proof of property 7

The marginal utility of saving in period i at rate r_l and consuming the savings in period $i+1$ is given by,

$$-(1+1_{\{\omega_i \leq R_i\}}\ell_i)U'(\omega_i) + \delta(1+r_l)\int(1+\ell_{i+1}1_{\{\omega_{i+1}<\omega_i\}})U'(\omega_{i+1}(x_{i+1}))f_{i+1}(x_{i+1}|a_{i+1})dx_{i+1} \tag{21}$$
$$+\ell_{i+1}\delta U'(\omega_i)\int_{\omega_{i+1}<\omega_i} f_{i+1}(x_{i+1}|a_{i+1})dx_{i+1}$$

and the marginal utility of borrowing in period i at rate r and paying back in period $i+1$ is given by,

$$(1+1_{\{\omega_i<R_i\}}\ell_i)U'(\omega_i) - \delta(1+r_b)\int(1+\ell_{i+1}1_{\{\omega_{i+1}\leq\omega_i\}})U'(\omega_{i+1}(x_{i+1}))f_{i+1}(x_{i+1}|a_{i+1})dx_{i+1} \tag{22}$$
$$-\ell_{i+1}\delta U'(\omega_i)\int_{\omega_{i+1}\leq\omega_i} f_{i+1}(x_{i+1}|a_{i+1})dx_{i+1}$$

By (21) the marginal utility of saving at rate $\frac{1}{\delta}-1$ is $-(1+\ell)U'(y)+U'(y)<0$. By (22) the marginal utility of borrowing at rate δ is $U'(y)-(1+\ell_{i+1})U'(y)-\delta\ell_{i+1}U'(y)<0$. Therefore the rate at which the agent would be willing to borrow is smaller than $\frac{1}{\delta}-1$ and the rate at which he would be willing to save must be greater than $\frac{1}{\delta}-1$.

The second point is justified subtracting (21) and (22) with $r_l=r_b=r$ and noting that what is obtained is strictly negative.

7.7 Proof of property 8

Suppose that $\omega_{i+1}(x_{i+1})\geq\omega_i$ for all x_{i+1}. By (21) the marginal utility of saving would be

$$-(1+1_{\{\omega_i\leq R_i\}}\ell_i)U'(\omega_i)+\int U'(\omega_{i+1}(x_{i+1}))f_{i+1}(x_{i+1}|a_{i+1})dx_{i+1} \tag{23}$$

By assumption we will have that $\int U'(\omega_{i+1}(x_{i+1}))f_{i+1}(x_{i+1}|a_{i+1})dx_{i+1}<U'(\omega_i)$ and therefore the agent will not have incentives to save. By 22 the marginal utility of borrowing is

$$(1+1_{\{\omega_i<R_i\}}\ell_i)U'(\omega_i)-\int(1+\ell_{i+1}1_{\{\omega_{i+1}=\omega_i\}})U'(\omega_{i+1}(x_{i+1}))f_{i+1}(x_{i+1}|a_{i+1})dx_{i+1}$$
$$-\ell_{i+1}\delta U'(\omega_i)\int_{\omega_{i+1}=\omega_i} f_{i+1}(x_{i+1}|a_{i+1})dx_{i+1}$$

it may be positive or negative depending on the parameters of the problem

Now, suppose that $\omega_{i+1}(x_{i+1})\leq\omega_i$ for all x_{i+1}. The marginal utility of saving is be

$$-(1+1_{\{\omega_i\leq R_i\}}\ell_i)U'(\omega_i)+\int(1+\ell_{i+1}1_{\{\omega_{i+1}<\omega_i\}})U'(\omega_{i+1}(x_{i+1}))f_{i+1}(x_{i+1}|a_{i+1})dx_{i+1}$$
$$+\ell_{i+1}\delta U'(\omega_i)\int_{\omega_{i+1}<\omega_i} f_{i+1}(x_{i+1}|a_{i+1})dx_{i+1}$$

We must have $\int U'(\omega_{i+1}(x_{i+1}))f_{i+1}(x_{i+1}|a_{i+1})dx_{i+1} > U'(\omega_i)$, therefore, if $\omega_i > R_i$ or ℓ_i is sufficiently small then the marginal utility would be positive and therefore the agent will have incentives to save. The marginal utility of borrowing is,

$$(1+1_{\{\omega_i<R_i\}}\ell_i)U'(\omega_i) - \int(1+\ell_{i+1}1_{\{\omega_{i+1}\leq\omega_i\}})U'(\omega_{i+1}(x_{i+1}))f_{i+1}(x_{i+1}|a_{i+1})dx_{i+1}$$
$$-\ell_{i+1}\delta U'(\omega_i)\int_{\omega_{i+1}\leq\omega_i} f_{i+1}(x_{i+1}|a_{i+1})dx_{i+1}$$

7.8 Proof of property 11

By backwards induction, the optimal spot contract in period T must give the agent the reservation utility $\bar{U}(\omega_{T-1})$ and will depend on ω_{T-1} since it represents the reference in period T. Thus, the optimal sequence of spot contracts has memory. The optimal spot contract in period $T-1$ solves

$$\max_{\omega_{T-1}(\cdot)} \int \left((x_{T-1}-\omega_{T-1}(x_{T-1}))f^{T-1}(x_{T-1}|a_{T-1}) + \delta V(\omega_{T-1}(x_{T-1}))f(x_T|a_T)\right)dx_{T-1}$$

$$\int \left(\tilde{U}_{T-1}(\omega_{T-1}(x_{T-1})) + \delta\bar{U}(\omega_{T-1}(x_{T-1}))\right) f^{T-1}(x_{T-1}|a_{T-1}) \geq \bar{U}(c_{T-2})$$

$$a_{T-1} \in \text{argmax}_a \int \left(\tilde{U}_{T-1}(\omega_{T-1}(x_{T-1})) + \delta\bar{U}(\omega_{T-1}(x_{T-1}))\right) f^{T-1}(x_{T-1}|a_{T-1})$$

where $V(\omega_{T-1}(x_{T-1}))$ represents the profits of the principal under the optimal spot contract in period T. Therefore, unless $\bar{U}(\omega_{T-1}(x_{T-1}))$ coincides with the expectation of the last period contract under the full-commitment optimum the optimal sequence of spot contracts does not implement the full-commitment optimum.

7.9 Proof of property 9

Using the fact that for any period i at the reference R_i (i.e. when $k_i \in [0,1]$) we must have

$$\ell_i k_i(x_0,x_1,\ldots,x_i)\left(\lambda_i + \mu_i\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}\right) = \frac{1}{U'(R_i)} - \left(\lambda_i + \mu_i\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}\right)$$

Noting that the reference in next period's payment scheme is the current's period payment, we have,

$$\delta\ell_2\int_{\omega_2(x_2)<\omega_1} k(x_2)\left(\lambda_2+\mu_2\frac{f^2_{a_2}(x_2|a_2)}{f^2(x_2|a_2)})f^2(x_2|a_2\right)dx_2 = \delta\mathbb{P}_{x_2}(\omega_2(x_2)=\omega_1|a_2)\left(\frac{1}{U'(\omega_1)}-\lambda_2\right) -$$
$$\int_{\omega_2(x_2)=\omega_1}\mu_2 f^2_{a_2}(x_2|a_2)dx_2 + \delta\ell_2\int_{\omega_2(x_2)<\omega_1}\left(\lambda_2+\mu_2 f^2_{a_2}(x_2|a_2)\right)dx_2$$

Replacing the last equation in the optimality condition for period 1 one obtains,

$$\frac{1}{U'(\omega_1(x_0,x_1))}\left(1+\delta\mathbb{P}(\omega_2=\omega_1(x_0,x_1)|a_2)\right)=\Big(1+k_1(x_1)\ell_1+\delta\mathbb{P}(\omega_1(x_0,x_1)=\omega_2|a_2)+$$
$$-\delta\ell_1\mathbb{P}(\omega_1(x_0,x_1)>\omega_2)\Big)\lambda_2-\delta\mu_2\ell_2\int_{\omega_2<\omega_1} f_{a_2}(x_2|a_2)dx_2+\mu_1\delta\int_{\omega_2=\omega_1} f_{a_1}(x_1|a_1)dx_1 \tag{24}$$

By an analogous argument the following expression can be derived,

$$\frac{1}{U'(\omega_0(x_0))}\left(1+\delta\mathbb{P}(\omega_1=\omega_0(x_0)|a_1)+\delta^2\mathbb{P}(\omega_2=\omega_0(x_0)|a_2)\mathbb{P}(\omega_1=\omega_0(x_0)|a_1)\right)=$$
$$(1+k_0(x_0)\ell_0+\delta\mathbb{P}(\omega_1=\omega_0(x_0)|a_1)+\delta^2\mathbb{P}(\omega_2=\omega_0(x_0)|a_2)\mathbb{P}(\omega_1=\omega_0(x_0)|a_1)+$$
$$-\delta\ell_1\mathbb{P}(\omega_0(x_0)>\omega_1|a_1)-\delta^2\ell_2\mathbb{P}(\omega_0(x_0)>\omega_1|a_1)\mathbb{P}(\omega_0(x_0)>\omega_2|a_2))\left(\lambda+\mu_0\frac{f^0_{a_0}(x_0|a_0)}{f^0(x_0|a_0)}\right)+$$
$$+-\delta^2\ell_2\int_{\omega_1=\omega_0,\omega_2<\omega_0}\left(\mu_1\frac{f^1_{a_1}(x_1|a_1)}{f^1(x_1|a_1)}+\mu_2\frac{f^2_{a_2}(x_2|a_2)}{f^2(x_2|a_2)}\right)f^1_{a_1}(x_1|a_1)f^2_{a_2}(x_2|a_2)dx_1dx_2+$$
$$+\delta^2\int_{\omega_1=\omega_0,\omega_2=\omega_0}\left(\mu_1\frac{f^1_{a_1}(x_1|a_1)}{f^1(x_1|a_1)}+\mu_2\frac{f^2_{a_2}(x_2|a_2)}{f^2(x_2|a_2)}\right)f^1_{a_1}(x_1|a_1)f^2_{a_2}(x_2|a_2)dx_1dx_2$$
$$-\delta\mu_1\ell_1\int_{\omega_0<\omega_1} f_{a_1}(x_1|a_1)dx_1+\mu_1\delta\int_{\omega_0=\omega_1} f_{a_1}(x_1|a_1)dx_1 \tag{25}$$

$$\frac{1}{U'(\omega_2(x_0,x_1,,x_2))} = (1+k_2(x_0,x_1,x_2)\ell_2)\left(\lambda_2+\mu_2\frac{f^2_{a_2}(x_2|a_2)}{f^2(x_2|a_2)}\right) \tag{26}$$

From (25) it can be seen that for whatever value of the first period's reference a λ high enough will insure that right hand side is greater than $1/U'(R_0)$ for $k_0=0$. Now, replacing in (24) and (26) the following,

$$\left(\lambda+\mu_0\frac{f^0_{a_0}(x_0|a_0)}{f^0(x_0|a_0)}\right)=\frac{1}{U'(\omega_0(x_0))}+$$
$$-\delta^2\int_{\omega_1=\omega_0,\omega_2=\omega_0}\left(\mu_1\frac{f^1_{a_1}(x_1|a_1)}{f^1(x_1|a_1)}+\mu_2\frac{f^2_{a_2}(x_2|a_2)}{f^2(x_2|a_2)}\right)f^1_{a_1}(x_1|a_1)f^2_{a_2}(x_2|a_2)dx_1dx_2+$$
$$-\mu_1\delta\int_{\omega_0=\omega_1} f_{a_1}(x_1|a_1)dx_1$$

we conclude that if $1/U'(\omega_0(x_0)$ is big enough the right hand sides of both equations will be greater than $1/U'(\omega_0(x_0))$ which implies that the payment schemes will be over the reference in periods 1 and 2.

References