Department of
Economics

University of Pittsburgh
Kenneth P. Dietrich School of Arts and Sciences

WORKING PAPER SERIES

16/013

Dynamic Contracts Under Loss Aversion

Alejandro Jofre, Sofia Moroni and Andrea Repetto

October 18, 2015

Dynamic Contracts Under Loss Aversion

Alejandro Jofre Sofia Moroni Andrea Repetto*

October 18, 2015

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. When there is full commitment and the agent has no access to credit, in every period after the first the optimal payment scheme is insensitive to the current outcome in an interval, offering to pay the reference for a set of performance measures. Therefore, there is a positive probability of observing wage persistence even if outcomes vary over time. Moreover, the model predicts a "status quo bias" –a preference for consuming the full allocation if the agent is allowed to intertemporally reallocate consumption 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 savings technology. We use subdifferential calculus to address the non-differentiable utility function.

Keywords: principal-agent model, moral hazard, dynamic contracts, loss aversion.

JEL Codes: D86, D82.

*Universidad de Chile, University of Pittsburgh and Universidad Adolfo Ibañez, respectively. We have greatly benefited from comments made by Felipe Balmaceda, Dirk Bergemann, Johannes Hörner, Nicolas Figueroa, Adam Kapor, David Pearce, Wolfgang Pesendorfer, Patrick Rey, and Juan Pablo Torres-Martinez. We are particularly grateful to Larry Samuelson and Hector Chade for their advice and detailed comments. We also thank seminar participants at University of Chile, Universidad Adolfo Ibañez, Central Bank of Chile, European University, Universidad Diego Portales, Universidad Alberto Hurtado, International Conference on Continuous Optimization, Latin American Econometric Society Meetings, North American Econometric Society Summer Meetings, SECHI Meetings, Annual Conference SMAI, Stochastic Programming Conference, Yale Summer Workshop and Latin American Workshop in Economic Theory.

1 Introduction

Within employment relationships, workers often do not see their wages change for extended periods of time in spite of the availability of performance measures.[1] These facts are at odds with the classic theory of long-term incentive contracts, which predicts that a worker's wages at each time are finely sensitive to the current and all previous periods' performance. Furthermore, long-term employment contracts are common in spite of workers' access to private savings and the possibility of renegotiation to a mutually beneficial agreement. In contrast, the classical theory predicts that in such a setting efforts above the minimum cannot in general be implemented (Fudenberg and Tirole 1990, Chiappori et al. 1994).

We analyze a dynamic principal-agent interaction with moral hazard in which the agent exhibits loss aversion with respect to a reference given by the previous period's consumption. We solve for the optimal contract when the agent has no access to credit markets and analyze the conditions under which this contract remains optimal if the agent has access to credit markets. We show that the optimal contract with a loss-averse agent exhibits wage persistence, may exhibit downward wage rigidity, and may delay incentives into the future. In addition, it may be possible to incentivize the agent to exert effort even when he has access to savings. These features are not present in the classical dynamic principal-agent model.

Our first finding is that optimal wage schedules may be insensitive to outcomes in an interval, and thus offer to pay the reference income for a set of performance measures. Intuitively, it is optimal to provide full insurance locally because loss aversion involves first-order risk aversion around the reference (Segal and Spivak 1997). In contrast, under mild conditions, the classical restricted savings model predicts that the compensation scheme will be strictly increasing in the current period outcome.

Second, for all periods after the first, the optimal wage schedule *must* pay the reference

[1] Workers' wages tend to stay constant and few workers experience wage cuts in any given year (Baker et al. 1994, Fortin 1996, Wilson 1997, Bewley 1999, Beissinger and Knoppik 2001, Lebow et al. 2003, Fehr and Goette 2005 and Dickens et al. 2007). Furthermore, bonuses are typically awarded on a quarterly or yearly basis and thus depend on infrequent performance evaluations (Healy 1985, Joseph and Kalwani 1998, Oyer 1998, Steenburgh 2008). The theory predicts that if workers have discretion over the choice of effort on a daily basis, for example, contracts should be contingent on all available information about today's and every previous day's performance.

for an interval of outcomes as long as the strength of reference dependence in the agent's preference is relatively stable over time. Except for the last period, this flat segment may even extend for the whole support of the outcomes' distribution. In contrast, the dynamic moral hazard model under classical risk aversion predicts strictly increasing contracts in all prior periods' outcomes. Under loss aversion, only the last period's contract must be sensitive to all periods' outcomes. 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. This prediction is consistent with the observed role of promotions in wage progression (Baker et al. 1994).[2]

Third, because each period's contract is flat across an interval of outcomes around the reference there is a positive probability that two consecutive payments are equal, i.e., that observed wages exhibit time persistence as reported by Dickens et al. (2007). In contrast, under the monotone likelihood ratio assumption the classical model predicts variability in observed wages in the current period's and all previous periods' outcomes.

Fourth, if the loss-averse agent were allowed to transfer resources over time at the going interest rate after the realization of the performance measure, he might choose to save, to borrow or to consume his entire wage. In the classical model, as the optimal contract smoothes the expected inverse of marginal utility over time, the scheme offers payments that are front-loaded. Intuitively, because 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 transfer resources over time, he would choose to save a fraction of his wage in earlier periods. This effect is also present in our model. However, there are other potentially counteracting effects in play. First, transferring resources over time not only modifies the intertemporal allocation of consumption but also changes the future references. Second, changes in consumption may lead to costly current or future losses, motivating the agent to prefer to consume his full allocation.

[2]See Baker, Jensen, and Murphy (1988) for a review of the evidence on promotion based incentives, and Prendergast (1999) for a review of the literature on deferred compensation.

Therefore, loss aversion implies that the marginal utility of postponing consumption need not be equal to (minus) the marginal utility of bringing consumption to the present. In fact, these objects may be both simultaneously negative. Whenever this happens, the agent experiences a loss from transferring resources over time and will prefer to consume his full wage in a manner consistent with the "status quo bias" of Samuelson and Zeckhauser (1988). Moreover, when 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 delay consumption is strictly higher than the rate he would be willing to pay to front load consumption. That is, the model predicts a related phenomenon: a discrepancy between willingness to accept and willingness to pay (Bateman et al. 1997).

An important consequence of our analysis is that the optimal payment scheme may remain optimal if the agent has access to credit and 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 has access to the credit market but his saving and borrowing are private information, the optimal long-term contract is no longer renegotiation proof. In fact, the renegotiation-proof long-term contract generally cannot provide incentives to exert effort above the minimum. Because it is unlikely that a court of law would prevent renegotiation toward a Pareto-improving agreement and because 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 provide a rationale for the existence of commitment contracts.

In addition, because 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, 1974, Rockafellar and Wets 1997). This method is useful when addressing non-differentiability and the lack of validity of the usual first-order conditions.

The remainder of the paper is organized as follows. In Section 2, we review the related literature. In Section 3, we present the main assumptions of the model. In Section 4, we derive the optimality conditions to characterize the shape of the optimal second-best payment scheme and also derive the main intra and intertemporal properties of this optimal contract. Finally, we conclude in Section 5.

2 Relation to Literature

Our framework builds on an extensive literature addressing dynamic moral hazard problems as well as loss aversion. In particular, our model builds on the seminal contributions of Rogerson (1985) and Chiappori et al. (1994), who characterize and analyze the second-best optimal scheme under classical risk aversion assumptions. Consistent with this literature, we find that the optimal long-term contract is non-decreasing in the current period's outcome, displays memory in wages and provides consumption smoothing. Moreover, the full commitment optimum is ex-post efficient and is therefore renegotiation proof.

Additionally, we make use of a large literature on loss aversion.[3] When preferences exhibit loss aversion with respect to a reference, the dislike generated by consumption below the reference is greater than the elation produced by an equally sized gain.

Because loss aversion is defined with respect to a reference, it is necessary to establish what the agent's reference is. Similar to Bowman et al. (1999), Munro and Sugden (2003) and Dittmann et al. (2010), we assume that 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.[4]

[3] Loss aversion was first proposed by Kahneman and Tversky (1979) as an essential element of their Prospect Theory. A large and growing body of literature in economics and cognitive psychology supports the hypothesis that references affect individual decisions. Rabin (1998) and Della Vigna (2009) provide surveys of this literature.

[4] There is little evidence on how reference levels are determined. Our assumption is similar to those made by models of the endowment effect,which posit that willingness to pay for a good depends on recent ownership

There is a gap between the rich contracts predicted by classical contract theory and the fairly simple contracts that are actually observed (Prendergast [1999], Salanié, [2003]). The literature on loss aversion in principal-agent settings, to which we contribute, has attempted to bridge this gap. Because loss aversion generates first-order risk aversion, models with loss aversion predict simpler schemes than those of the classical model, with payments that are insensitive to outcomes over some regions.[5]

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, the model predicts optimal schemes that are continuous. Discontinuous contracts may arise when the reference is the median wage. In this case, the principal can provide insurance and incentives avoiding the loss area by simultaneously lowering the median and rewarding good performance.

Optimal binary payment schemes arise in Herweg et al. (2010), who model reference formation following Kőszegi and Rabin (2006, 2007) in a static setting.[6] In their setting, the optimal scheme is a lump sum bonus contract with the bonus paid whenever a certain level of performance is achieved. Daido and Itoh (2006) also allow for loss aversion, making the Kőszegi and Rabin (2006, 2007) assumption regarding reference formation, but assume a binary measure of performance.

Macera (2013) considers a two-period model in which the agent exhibits loss aversion in present consumption and in beliefs regarding future consumption as in Kőszegi and Rabin (2009), an extension of the static environment assumptions of Kőszegi and Rabin (2006, 2007). Depending on the relative strength of the current and future gain-loss utility component, the optimal contract may offer a fixed wage in one period and an output-contingent

status. Recent consumption also serves as a reference in habit-formation models. In a dynamic setup such as 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 (Kőszegi and Rabin 2009). It is worth emphasizing that our model is characterized by fully forward-looking behavior: both the agent and the principal take into account the effects of their decisions on future references.

[5] Kőszegi (2014) offers a review of contract-theory models built on alternative psychological foundations of behavior.

[6] That is, the agent derives gain-loss utility by comparing the actual payment with the (lagged) rational expectations regarding rewards.

increasing wage scheme in the second period. That is, if contemporaneous gain-loss utility is relatively important, then it may be optimal to provide full insurance in the current period, deferring all incentives to the future. To our knowledge, this is the only other paper that considers loss aversion in a dynamic principal-agent setting with moral hazard.[7]

Iantchev (2011) studies the one-period principal-agent model under moral hazard with a loss-averse agent. 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.[8]

Our model builds on this growing literature on contracting with loss-averse agents. As we introduce dynamics, we allow current-period consumption to affect the agent through two channels: by providing incentives directly and by determining the reference for the next period. Moreover, in addition to characterizing the intertemporal allocation of risk and incentives, we analyze the optimal contract's renegotiation proofness, wage persistence and the role of the constrained savings assumption. Furthermore, our paper rigorously derives the optimality conditions on the basis of convex analysis tools in order address the non-differentiable utility function.

3 Model

The model describes a repeated principal-agent problem analogous to the dynamic moral hazard model of Rogerson (1985) and Chiappori et al. (1994); i.e., we assume a finite horizon,

[7]We also find that incentives may be delayed to the future, as our optimal contracts may be constant on outcomes in all periods but the last. However, the intuition for the result differs in the two models. In both models, the last period contract must depend on its outcomes as there are no future periods to provide incentives for effort in the last period. In Macera (2013), as surprises have net negative effects on utility, the principal may optimally offer an outcome-contingent contract only in the last period. In our model, however, as the reference is not given by expected payments but by the payment in the previous period, the principal can always design a contract that gives payments only at or above the reference so that variation in the payments has a net positive effect on the gain-loss portion of utility. However, it may still not be worth it for the principal to provide incentives in the current period. As the marginal utility of income is relatively low in the gain area, the effect on incentives of an outcome contingent contract may be relatively low, and the principal may optimally decide to delay all incentives until the last period.

[8]Because the agent is assumed to be risk loving in the loss space, the payment scheme displays a discontinuity whenever output falls below a threshold. A discontinuous drop in payment when the observed outcome is low is also found in Dittmann et al. (2010), which similarly assumes the agent is risk loving in the loss space.

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 $f^i(x_i|a_i)$. 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.[9] Additionally, we assume that $\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}$ as a function of x_i is in C^1.[10]

In each period, the agent obtains utility from consumption and incurs a cost from exerting effort. The agent's utility in period i, $\tilde{U}_i(c_i, R_i)$, depends on c_i, the agent's consumption in period i, and on R_i, the agent's reference point in that period. The cost of effort a_i is given by $\psi_i(a_i)$ with ψ_i strictly increasing. Thus, the agent's payoff in period i is

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

To allow for loss aversion, we assume that the agent's utility has a kink at the reference consumption. That is, the left-sided derivative of $\tilde{U}_i$ is greater than its right-sided derivative:

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

In other words, we assume that the utility gain of consumption above R is lower than the utility loss of consumption below R by an equal amount. We assume that $\tilde{U}_i$ is continuous, strictly concave and differentiable at all points other than R and that $\lim_{c \to \infty} \tilde{U}_i(c, R) = \infty$. [11] Given the reference level R_0, there exist constants $\ell_0 > 0$, $\bar{\theta} \in (0,1)$, and a smooth, concave

[9] In other words, we assume that $f^i(x_i|a_H)/f^i(x_i|a_L)$ is non-decreasing in x_i.

[10] That is, it belongs to the set of continuously differentiable functions.

[11] Note that we do not include all aspects of Prospect Theory. In particular, we do not assume diminishing sensitivity; i.e., that the agent is risk loving in losses. Instead, we assume that the agent is risk averse in gains and in losses. Additionally, we do not assume that the agent weighs outcome probabilities.

and strictly increasing function $U(\cdot)$ such that period 0 utility, $\tilde{U}_0$, can be expressed as: [12]

$$\tilde{U}_0(c_0,R_0) = U(c_0) - \ell_0\theta(c_0,R_0)\left(U(R_0) - U(c_0)\right), \tag{2}$$

where

$$\theta(c,R) \quad = \quad \begin{cases} 1 & \text{if } c < R \\ \bar{\theta} < 1 & \text{otherwise.} \end{cases} \tag{3}$$

The proofs are in the appendix. Lemma 1 tells us that if the reference level is exogenously given, all utility functions that are concave and have a kink at the reference can be written as (2) for some concave *underlying utility function* U. For periods $i > 0$, we assume that the reference of the agent corresponds to the previous period's consumption. That is, we set

$$R_i = c_{i-1}.$$

This assumption is based on psychological evidence indicating that individual choices depend not only on current consumption but also on previous levels of consumption and that consumers form habits that affect how they perceive these different levels of consumption (see Kahneman et al. 1991). Bowman et al. (1999), Munro and Sugden (2003), Iantchev (2011) and Dittmann et al. (2010) also assumed similar forms of update.[13,14]

In the canonical repeated moral hazard model, the agent is assumed to have the same preference for consumption over time. In our setting, utility changes from one period to the next because of reference dependence. The reference dependent utility $\tilde{U}_i(c_i,R_i)$ in periods

[12] The function U and constant ℓ_0 depend on R_0.

[13] Our choice of reference update is also related to the habit-formation literature initiated by Gorman (1967) and Pollak (1970).

[14] Other papers assume different processes for reference formation. For instance, Kőszegi and Rabin (2006, 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. Finally, Della Vigna et al. (2015) assume reference consumption equal to average consumption in previous periods. The experimental evidence in Heffetz and List (2014) is consistent with the hypothesis that references depend on endowments rather than on expectations.

$i > 0$ is built on the underlying utility U and is given by,

$$\tilde{U}_i(c_i, R_i) = \tilde{U}_i(c_i, c_{i-1}) = U(c_i) - \ell_i \theta(c_i, c_{i-1}) \left(U(c_{i-1}) - U(c_i)\right). \tag{4}$$

with $\theta(c_i, c_{i-1})$ defined as in (3) and $\ell_i > 0$. Recall that Lemma 1 showed that in period 0 utility takes this form. It is straightforward that the same holds for later periods, though possibly with a different function U_i for period i on the right hand side. We assume that the underlying utility U and θ are the same functions in every period, while ℓ_i can vary across periods, as stated in equation (4).[15]

The agent's utility is decreasing in the reference for gains and losses. Figure 1 illustrates how the utility changes as a function of the reference.[16] If $\bar{\theta} = 0$, the agent's utility does not depend on the reference for consumption in the gain area. However, when $\bar{\theta} < 1$, the agent's utility is decreasing in the reference for all consumption levels.

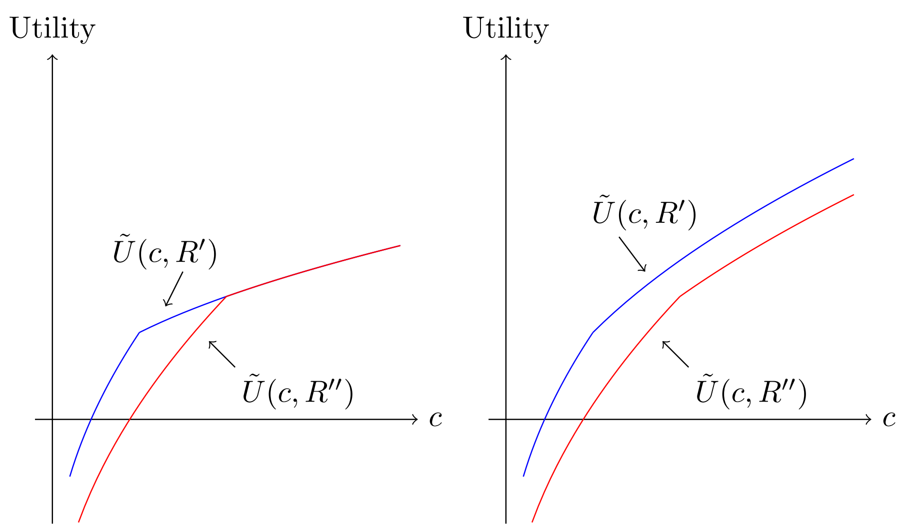

Figure 1: Utility function for reference points R' and R'', with $R' < R''$. Left: $\bar{\theta} = 0$. Right: $0 < \bar{\theta} < 1$.

[15]Most of the literature uses analogous functional forms to express loss-averse utilities. Utility is usually defined as $U(c) + v(c, R)$ with $v(c, R) = \mu(c) - \mu(R)$ if$c > r$ and $v(c, R) = \lambda(\mu(c) - \mu(R))$ if $c \leq R$ with $\lambda > 1$. From Lemma 3, for fixed R, this utility form is fully general. Some examples are Kőszegi and Rabin (2006), Kőszegi and Rabin (2009), Kőszegi and Rabin (2009), Fehr and Goette (2007), Goette et al. (2004), Freund and Özden (2008), Heidhues and Kőszegi (2008) and Card and Dahl (2011).

[16]In a previous version of this paper, we considered the case in which $\bar{\theta} = 0$. The present version extends the analysis to a more general class of utilities in which the reference also affects the agent's preferences in the gain area (e.g., if $R' < R''$, then consuming any level C generates higher utility when the reference is R' than when it is R'', as shown in Figure 1).

At the beginning of the relationship, the principal offers the agent wage schedules $\omega_i(x_0,\dots,x_i)$ for each period i that depend on the realized outcomes up to period i. Each wage schedule $\omega_i(x_0,\dots,x_i)$ specifies the transfer that the agent receives after the realization of the outcome in period i as a function of the outcomes in all periods up to i. The principal has the ability to perfectly commit to these contingent payments.

If the agent accepts the contract, then at the beginning of each period i, he will choose an effort a_i to exert in that period. Thus, the choice of effort in period i may depend on the outcomes up to period $i-1$. The agent chooses the effort to maximize his expected payoff. If the agent does not accept the contract, he obtains his outside option, which gives him payoff U^*.

We assume that the agent and the principal discount the future exponentially at a factor δ. We assume that $\ell_i\delta < 1$ to ensure that the total utility of two consecutive periods is increasing in the consumption of the first.

Finally, we assume that the principal is risk neutral and therefore, for any given outcome x_i, his utility is $x_i - \omega_i(x_0,x_1,\dots,x_i)$.

4 The Optimal Contract

We start by analyzing the optimal contracting problem under the assumptions of full commitment and no access to credit markets for the agent.[17] That is, we assume that both the principal and the agent are able to commit to the contract during the entire duration of the relationship and that the principal can borrow and save at the fixed interest rate $\frac{1}{\delta}-1$, whereas the agent can neither borrow nor save.

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

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

[17] In section 4.4, we discuss the optimality of the present contract when the agent has access to borrowing and saving. Note that in the canonical model in Rogerson (1985) the principal must restrict the agent's savings in the optimal contract.

subject to

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

$$a(\cdot) = (a_0, a_1(x_0), \ldots a_T(x_0, \ldots, x_{T-1})) \in \text{argmax}_{a(\cdot)} \sum_{i=0}^{T} \delta^i \mathbb{E}\left(\tilde{U}_i(\omega_i, R_i) - \psi_i(a_i) | a_0, a_1, \ldots, a_i\right). \tag{IC}$$

where $\mathbb{E}(\cdot|a_0, a_1, \ldots, a_i)$ denotes an expectation given actions $(a_0, a_1, \ldots, a_i)$ and $R_i = \omega_{i-1}$.[18]

The objective function represents the expected payment to the principal. The first constraint, (IR), is the individual rationality constraint. It requires that the agent obtain an expected utility of at least U^* from the relationship. The second constraint, (IC), states that the effort chosen maximizes the expected utility of the agent and is henceforth referred to as the incentive compatibility constraint.

4.1 Preliminaries

In this section, we outline some technical results that allow us to derive the optimality conditions for the principal's problem. To find optimality conditions, it is convenient to use the transformation in Grossman and Hart (1983) and 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 $\{v_0, v_1, \ldots, v_{i-1}\}$. Because the previous periods' outcomes affect the previous period's consumption, and consumption determines the reference, the cost of providing a given utility in period i depends on the utility provisions in previous periods.

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

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

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

$$h_i(v_0, v_1, \ldots, v_i) = \begin{cases} U^{-1}\left(\frac{v_i + \ell_i U(h_{i-1}(v_0, v_1, \ldots, v_{i-1}))}{1+\ell_i}\right) & \text{if } v_i < U(h_{i-1}(v_0, v_1, \ldots, v_{i-1})) \\ U^{-1}\left(\frac{v_i + \ell_i \bar{\theta} U(h_{i-1}(v_0, v_1, \ldots, v_{i-1}))}{\left(1+\ell_i \bar{\theta}\right)}\right) & \text{if } v_i \geq U(h_{i-1}(v_0, v_1, \ldots, v_{i-1})) \end{cases}, \tag{5}$$

with $U(h_{-1}) = R_0$. The agent's utility is weakly above the utility evaluated at the reference whenever $w_i(x_0, \ldots, x_i) = h_i(v_0, v_1, \ldots, v_i) \geq w_{i-1}(x_0, \ldots, x_{i-1}) = h_{i-1}(v_0, v_1, \ldots, v_{i-1})$, that is, whenever $v_i \geq U(h_{i-1}(v_0, v_1, \ldots, v_{i-1}))$.

We can now rewrite the program the principal faces as choosing utility provisions $v_i(x_0, x_1, \ldots, x_i)$ contingent on the outcomes up to period i as follows:

$$\max_{(v_i(\cdot))_i, (a_i)_i} \sum_{i=0}^{T} \delta^i \mathbb{E}\left(x_i - h_i(v_0, v_1, \ldots, v_i) | a_0, a_1, \ldots, a_i\right) \tag{6}$$

subject to

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

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

The following two propositions provide the necessary inputs for the optimality conditions for the payment scheme described in Proposition 3 below. Proposition 1 proves that the problem is convex, which allows for the use of subdifferential calculus in finding the conditions for the optimal contract. That is, 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.[19] Proposition 1 characterizes the sub-gradient set of the cost function, which is a crucial step in computing the sub-gradient set of the Lagrangian.

Proposition 1 (Convexity)**.** *Under the model assumptions, the utility provision cost functions* $h_i : \mathbb{R}^i \to \mathbb{R}$ *for* $i \in \{0, 1, \ldots, T\}$ *are strictly convex, and therefore, the optimization problem*

[19] 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. In this case, because the objective function is not differentiable, optimization can be attained by finding the sub-gradient set. A reference for these ideas is Rockafellar (1974).

given by (6)-(PC')-(IC') *has a strictly concave objective function and a convex feasible set.*

The subgradient set[20] *of $h_i(v_0, v_1, \ldots, v_i)$ is given by*[21]

$$\partial h_i(v_0, v_1, \ldots, v_i)(\cdot) = \left(\frac{1}{U'(\omega_i(\cdot))} \left(\prod_{t=j+1}^{i} \frac{k_t(\cdot)\ell_t}{1+k_t(\cdot)\ell_t} \right) \frac{1}{1+k_j(\cdot)\ell_j} \right)_{j=0}^{i}, \tag{7}$$

where the function $k_t(x_0, x_1, \ldots, x_t)$ is associated with the kink in the utility function and is given by

$$k_t(x_0, x_1, \ldots, x_t) \in \begin{cases} \{1\} & \text{if } \omega_t(x_0, x_1, \ldots, x_t) < R_t \\ [\bar{\theta}, 1] & \text{if } \omega_t(x_0, x_1, \ldots, x_t) = R_t \\ \{\bar{\theta}\} & \text{otherwise} \end{cases}. \tag{8}$$

4.2 Optimality Conditions

Using the results in the previous section we are able to derive the first-order conditions that characterize the optimal contract.

Proposition 2. *An optimal wage schedule that solves the program faced by the principal must satisfy the following optimality conditions:*

$$\frac{1}{U'(\omega_i(x_0, x_1, \ldots, x_i))} = (1 + k_i(x_0, x_1, \ldots, x_i)\ell_i) \left(\lambda_i + \mu_i \frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)} \right) +$$
$$- \delta\ell_{i+1} \int k_{i+1}(x_0, x_1, \ldots, x_{i+1}) \left(\lambda_{i+1} + \mu_{i+1} \frac{f^{i+1}_{a_{i+1}}(x_{i+1}|a_{i+1})}{f^{i+1}(x_{i+1}|a_{i+1})} \right) f^{i+1}(x_{i+1}|a_{i+1}) dx_{i+1} \qquad \forall i < T \tag{9}$$

and

$$\frac{1}{U'(\omega_T(x_0, x_1, \ldots, x_T))} = (1 + k_T(x_0, x_1, \ldots, x_T)\ell_T) \left(\lambda_T + \mu_T \frac{f^T_{a_T}(x_T|a_i)}{f^i(x_i|a_i)} \right) \tag{10}$$

where

[20] For a generic convex function $f : \mathbb{R}^{n+1} \to \mathbb{R}$, the subgradient set at $x \in \mathbb{R}^{n+1}$ is defined as the set of vectors $d = (d_0, d_1, \ldots, d_{n+1}) \in \mathbb{R}^{n+1}$ such that for any vector $\alpha \in \mathbb{R}^{n+1}$

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

[21] In our notation, $\Pi^i_{t=j+1} \frac{k_t(\cdot)\ell_t}{1+k_t(\cdot)\ell_t}$ equals 1 when $j \geq i$.

$\lambda_i = \lambda + \sum_{k=0}^{i-1} \mu_k \frac{f^k_{a_k}(x_k|a_k)}{f^k(x_k|a_k)}$, with λ being a multiplier associated to (PC') and $\mu_i = \mu_i(x_0,\ldots,x_{i-1})$ for $i \in \{0,\ldots,T\}$ being the multipliers associated to the incentive compatibility constraints.

The function $k_i(x_0,x_1,\ldots,x_i)$ is given by equation (8).

Equations (9) and (10) provide necessary and sufficient conditions for the period i wage schedule that the principal offers to the agent. If we set $\ell_i = 0$, we recover the first-order conditions that characterize the solution in Rogerson (1985). If we set $\ell_i = 0$ and $T = 0$, then (10) matches the condition in Grossman and Hart (1983). The optimality condition for a spot contract is given by equation (10) with $T = 0$ as in the one period model by De Meza and Webb (2007).

Inspection of equations (9) and (10) shows that there are two main properties that distinguish the shape of the optimal payment scheme from the classical case. First, the contract can be insensitive to the outcome in period i, x_i, in an interval. This is explained by the multiplicative term $(1 + k_i(x_0,x_1,\ldots,x_i)\ell_i)$ in (9) and (10). At the reference level, $k_i(x_0,x_1,\ldots,x_i)$ is allowed to take any value in $[\bar{\theta},1]$. Therefore, the right hand side of equations (9) and (10) can remain constant in an interval: as x_i increases, $k_i(x_0,x_1,\ldots,x_i)$ decreases so that $\omega_i(x_0,x_1,\ldots,x_i)$ remains at the reference. Intuitively, the cost of inducing effort by increasing payments just above the reference may be high because of 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.

More formally, if the contract pays the reference at an interior outcome $\tilde{x}_i \in (\underline{x}_i, \bar{x}_i)$ and is strictly increasing in an interval around $\tilde{x}_i$[22], the principal can change the contract slightly and reduce the expected wage. By allowing the contract to be flat in a small interval around $\tilde{x}_i$ and punishing low outcomes more harshly, the principal can obtain a first-order gain at a second-order cost.[23] Thus, the contract must be insensitive to outcomes if it pays the reference for

[22]We show in Proposition 4 that the optimal contract must indeed be non-decreasing.

[23]If the principal makes the contract flat in an interval of length 2ε around x_i, the agent's utility experiences a first-order gain of approximately $\ell(1-\bar{\theta})U'(\omega_i(x_0,\ldots,x_i))\int_{x_i-\varepsilon}^{x_i+\varepsilon} f_i(x_i|a_i)\,dx_i$. The cost to the principal is of second order. If the payment scheme is further modified by lowering the payment that the agent receives around the lowest outcomes to keep the agent's utility equal to his utility in the original contract,

an outcome $x_i \in (\underline{x}_i, \bar{x}_i)$.

The second difference relates to the fact that the principal takes into account that each period's payment determines the reference level of the following period. The last term of the right hand side of equation (9), which is strictly positive, shows this effect. In each period i, this term tends to lower the payment scheme and reduce its growth rate as x_i increases. Intuitively, the term represents the benefit of lowering the payment in the current period to reduce the following period's reference and hence increase the utility of the agent. Naturally, this effect does not occur in period T (equation (10)).

These optimality conditions also directly influence the consumption-smoothing property in Rogerson (1985) that relates the wage schedules offered in any two consecutive periods. In Rogerson (1985), the inverse of the marginal utility of income must equal the expected value of the inverse of the next period's marginal utility of income. This exact condition does not hold in our model. However, an extended condition can be derived; in the following proposition, we show that the inverse of the marginal utility of income in any period might be greater or smaller than the expectation of the inverse of the marginal utility in the following period.

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 $k_i(x_0, x_1, \ldots, x_i)$.

Proposition 3 (Relationship between two consecutive periods)**.** *The following relationship between two consecutive periods is fulfilled:*

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

the principal achieves first-order savings. Furthermore, the new contract also satisfies the *IC* constraint. To see this, note that if the new contract gives the same utility to the agent, then we have approximately $\ell_i(1-\bar{\theta})U'(\omega_i(x_0,\ldots,x_i)) \int_{x_i-\varepsilon}^{x_i+\varepsilon} f^i(x_{i,}|a_i)\,dx_i = \ell_i U'(\omega(x_0,\ldots,\underline{x}_i)) \int_{\underline{x}_i}^{x_i+\tilde{\varepsilon}} f^i(x_{i,}|a_i)\,dx_i$ where $\tilde{\varepsilon}$ is chosen so that the previous equality holds. The left hand side of the *IC* constraint changes by

$$-\ell_i(1-\bar{\theta})U'(\omega_i(x_0,\ldots,x_i)) \int_{x_i-\varepsilon}^{x_i+\varepsilon} \frac{f^i_{a_i}(x_{i,}|a_i)}{f^i(x_i|a_i)} f^i(x_i|a_i)\,dx_i + \ell_i U'(\omega(x_0,\ldots,\underline{x}_i)) \int_{\underline{x}_i}^{x_i+\tilde{\varepsilon}} \frac{f^i_{a_i}(x_{i,}|a_i)}{f^i(x_i|a_i)} f^i(x_i|a_i)\,dx_i,$$

compared to the original contract. This quantity is strictly positive by MLRP.

where

$$c_i(x_{i-1}) = -\frac{\ell_i \delta}{1+k_{i-1}(x_{i-1})\ell_{i-1}} \int k_i(x_i) \left(\lambda_i + \mu_i \frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)} \right) f^i(x_i|a_i) dx_i + \\ \ell_{i+1}\delta \int\int \frac{k_{i+1}(x_{i+1})}{1+k_i(x_i)\ell_i} \left(\lambda_{i+1} + \mu_{i+1} \frac{f^{i+1}_{a_{i+1}}(x_{i+1}|a_{i+1})}{f^{i+1}(x_{i+1}|a_{i+1})} \right) f^{i+1}(x_{i+1}|a_{i+1}) f^i(x_i|a_i) dx_{i+1} dx_i. \tag{11}$$

Proof. Follows directly from equations (9) and (10).

□

Ignoring the multipliers k_{i-1} and k_i, the condition that links the wages in two consecutive periods in Rogerson (1985) is obtained by setting $c_i = 0$. Intuitively, the marginal cost of providing utility in period i, $h'(v_i) = \frac{1}{\tilde{U}'(\omega_i(x_i),R_i)}$, must equal the expected marginal cost of providing this utility in the following period. In our setting, however, the marginal cost of providing utility in period $i-1$ is not only accrued in that period because a higher utility in a period lowers the utility of the agent in future periods. The term $c_i(x_{i-1})$ represents the portion of the cost of utility provision that is not accounted for by the current period's marginal cost in i and $i-1$. Furthermore, because the utility function is not differentiable, it is as if the principal computes the marginal cost of providing utility $\frac{1}{U'(\omega(x_i))(1+k_i(x_i)\ell_i)}$ with the corresponding multiplier $k_i(x_i) \in (\bar{\theta}, 1)$ at the point of non-differentiability such that the marginal costs and benefits are equal.

The term c_i consists of the marginal effects of providing utilities in periods $i-1$ and i that are incurred as a result of the reference update. c_i may be positive or negative when $i < T$. It is positive when the marginal cost of utility provision arising from the effect on the reference is greater in period i than in period $i-1$. Because the relationship between the principal and agent ends in period T, the marginal cost resulting from a reference update is zero in that period. Thus, $c_T < 0$, which tends to lower the wage in period $T-1$ relative to the wage in period T. In section 4.4, we show that because c_i is not zero, the principal does not necessarily have to restrict savings to implement the optimal contract. This result is in contrast with Rogerson (1985), in which the condition analogous to Proposition 3 implies that the agent has incentives to save after each period and that the principal cannot implement the full-commitment contract without restricting the agent's access to savings.

In the following subsection, we describe the main intra- and intertemporal properties of this optimal scheme.

4.3 Properties of the Optimal Payment Scheme

In the previous section, we argued that the optimal contract may be flat as a function of the period's outcome. The following Proposition shows that the optimal contract *must* indeed have flat segments at the reference in all periods after period 0 when the intensity of the reference dependence is relatively stable over time. We also show that the contracts must be continuous and non-decreasing in the period's outcome. These results imply that there is a positive probability that an agent receives the same wage period after period. In contrast, in the classical setting, the agent receives the same wage in any two consecutive period with zero probability.[24]

Our first step is to establish some basic properties of the optimal contract.

Proposition 4 (Increasing and continuous)**.** *For each $i \in \{0,\dots,T\}$, the optimal contract $\omega_i(x_0,\dots,x_i)$ is non-decreasing and continuous in x_i. Additionally, for each $j < i$, the expectation of ω_i with respect to $x_{j+1},\dots,x_i$ after outcomes $x_0,\dots,x_j$, denoted $\mathbb{E}_{x_{j+1},\dots,x_i}(\omega_i|x_0,\dots,x_j)$, is non-decreasing and continuous with respect to x_j.*

To gain some intuition for why the contract must be non-decreasing in each period's outcome note that the principal can always improve on a decreasing contract by offering a contract that decreases more slowly and simultaneously gives the same utility to the agent. The new contract weakly improves the slackness of the *IC* constraint and is less risky for the agent. That is, the utility provision of the original contract is a mean-preserving spread of the utility provided by the new contract. As the cost of providing utility in all periods $j \geq i$ is

[24] This observation highlights the difference between our dynamic setting and the one period setting with an exogenous reference in De Meza and Webb (2007). While the contract may have have flat segments in the one period case, it must have flat segments in all periods after the first in the dynamic case when the agent has sufficiently stable preferences.

convex in the utility provided in period i (by Proposition 1), the cost of the new contract must be strictly less than the cost of the original contract.

An analogous argument proves that the contract is continuous in the outcome of each period. If the contract were to jump, the principal can improve upon this contract by offering a contract that jumps by slightly less. By reducing the agent's risk, the new contract reduces the principal's cost to first order while violating the IC constraint by an amount that is at most second order.

Proposition 5 (Shape of the optimal contract)**.** *The optimal contract $\omega_i(x_0, x_1, \ldots, x_i)$ is such that,*

1. *For $i \geq 1$, if $(\ell_i - \ell_{i+1})\delta \geq \bar{\theta}\ell_{i-1} - \ell_i$, $\ell_{i-1} \geq \delta\ell_i$ and $\omega_{i-1}(x_0, x_1, \ldots, x_{i-1}) > R_{i-1}$, 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 $x_i \in [\underline{x}_i, \bar{x}^i]$. Moreover, the payment scheme has a flat segment at the reference, and therefore, ω_i is not strictly increasing in x_i.*

2. *For $i \geq 1$, if $(\ell_i - \ell_{i+1})\delta \geq \ell_{i-1} - \ell_i$, $\ell_{i-1} \geq \delta\ell_i\bar{\theta}$ 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 in x_i.*[25]

3. *The payment scheme has a flat segment at the reference, but it cannot be completely flat in x_T. Furthermore, the last period contract is sensitive to all previous periods' outcomes.*

The conditions $(\ell_i - \ell_{i+1})\delta \geq \bar{\theta}\ell_{i-1} - \ell_i$, $\ell_{i-1} \geq \delta\ell_i$, $\ell_{i-1} \geq \delta\ell_i\bar{\theta}$ and $(\ell_i - \ell_{i+1})\delta \geq \ell_{i-1} - \ell_i$ imply that the weight on the gain/loss component of the utility does not decrease or increase too quickly from one period to the next. If, for example, $\ell_i = \ell$ for all $i \in \{0, \ldots, T\}$, these conditions hold. Thus, the optimal contract has a flat segment at the reference in every period after the first as long as the intensity of reference dependence is relatively stable over time. Figure 2 illustrates the shapes that the optimal payment schedule may take depending

[25] Where we define $\ell_{T+1} = 0$.

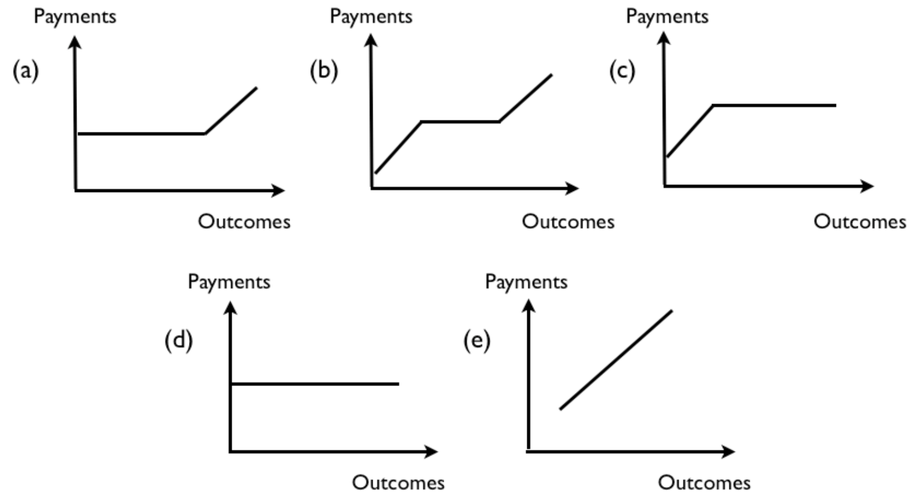

Figure 2: Shapes of optimal contracts: (a) through (e) are possible in period 0. (a) through (d) are possible in periods 1 through $T-1$. (a) through (c) are possible in period T.

on the period when the conditions on the variation of ℓ_i in conditions 1 through 3 in the Proposition hold. From Proposition 5, payment schemes such as those shown in panels (a), (b) and (c) could be optimal in any period. The scheme in (d) could be optimal in periods $\{0,\ldots,T-1\}$. Completely flat schedules are not optimal in period T. Finally, a scheme similar to that shown in panel (e) could only be optimal in period 0, when the optimal payment scheme may or may not offer the reference level of consumption for any outcome.

Several elements distinguish period T. First, consumption in T does not affect the reference of future periods, and thus, the effect of a high wage in T on the cost of providing utility is limited. Second, given that the contract ends in T, incentives in this last period can only be provided by a contract that is at least partly sensitive to x_T. Furthermore, Proposition 5 shows that the last period's contract must be at least partly sensitive to all previous periods' outcomes.

We know that the contract must have a flat segment in period i whenever it pays the same as in period $i-1$. The conditions on ℓ_i guarantee that the contract does so for some outcome. For intuition for why the contract in period i pays the same as in period $i-1$, note that in the classical case, there is a single outcome at which the contract pays the same wage as the

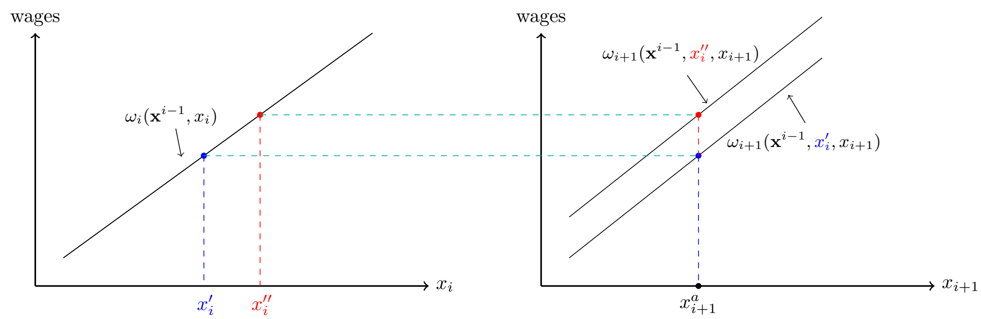

Figure 3: Period i versus period $i+1$ contract when the agent does not have preference dependent preferences (classical case).

previous period. The agent prefers smooth consumption. As a result, it is too costly for the principal to offer wages that vary excessively across periods. At the optimal contract in the classical model, the principal equates the marginal cost of increasing wages in period $i-1$ to the marginal cost of increasing wages for all outcomes in period $i+1$ (see Proposition 3). If these marginal costs are to be equated, the contract in period $i+1$ cannot be strictly above or strictly below the payment in period i. In fact, in the classical case, the integral term in equation (9) is zero. Thus, at x_i^a, which is defined as the outcome such that $\frac{f^i_{a_i}(x^a_i|a_i)}{f(x^a_i|a_i)} = 0$, the agent's wage equals the wage in period $i-1$.[26] Figure 3 illustrates how the contract varies from one period to the next in the classical case. At outcome x^a_{i+1} (and only at that outcome), the payment in period $i+1$is exactly the payment that the agent received in period i.

Now, in our setting, the principal takes into account the cost of increasing the reference, which reduces the agent's utility in the following period, and the shadow marginal cost–given by the multiplier k_i–of providing a wage at the reference (see Proposition 3). When ℓ_{i+1} is large relative to ℓ_i to the point that $(\ell_i - \ell_{i+1})\delta \geq \ell_{i-1} - \ell_i$ or $(\ell_i - \ell_{i+1})\delta \geq \bar{\theta}\ell_{i-1} - \ell_i$ fail to hold, the effect of a high wage in period i on the following period's preference is large. Thus, the principal may optimally lower the period i contract so that it may never give a payment at the reference, as it remains in the loss area for all outcomes. Analogously, when ℓ_{i-1} is large relative to ℓ_i the effect of the period $i-1$ wage on period i's preferences is relatively small.

[26]Note that x^a_i exists for all $i \in \{0,\ldots,T\}$ because $\int_{\underline{x}_i}^{\bar{x}_i} f^i_{a_i}(x_i|a_i)\,dx_i = F(\bar{x}_i|a_H) - F(\bar{x}_i|a_L) = 1-1 = 0$. Thus, because the ratio $f^i_{a_i}(x_i|a_i)/f^i(x_i|a_i)$ is increasing (by the MLRP assumption), $f^i_{a_i}(x_i|a_i)/f^i(x_i|a_i)$ is negative for $x_i \in [\underline{x}_i, x^a_i)$ and positive for $x_i \in (x^a_i, \bar{x}_i]$.

Additionally, it is relatively costly for the principal to provide payments below the reference in period $i-1$. Thus, the contract in period $i-1$ may be higher than the contract in period i to the point that the period i contract may remain in the loss area for all outcomes. Analogously, if ℓ_i is large relative to ℓ_{i-1}, so that $\ell_{i-1} < \delta\ell_i\bar{\theta}$ or $\ell_{i-1} < \delta\ell_i$, the cost of providing payments in period i is high relative to that in period $i-1$. Thus, it may be optimal for the principal to offer a period i contract that is above the reference for all outcomes.

In contrast, when the weight on the loss/gain portion of the utility is sufficiently stable over time, the integral term in condition (9) for period $i+1$ cannot differ substantially from the term in the condition for period i. Thus, the contracts in all periods will pay the previous period's wage for some outcome around which the contract must have a flat segment.

Proposition 5 also indicates that when the preferences are sufficiently stable, each period's contract is flat in an interval of outcomes around the reference. Thus, there is a positive probability that two consecutive payments are equal, i.e., that observed wages exhibit time persistence, as reported by Dickens et al. (2007). This observation is shown in the top panel of Figure 4. Furthermore, for two outcomes that pay a wage at the reference, the wages in the following period overlap, as illustrated in the bottom panel of Figure 4. Thus, under the optimal contract, there is a positive probability that wages do not vary across different histories and different periods.

Figure 5 illustrates how the period T contract depends on the outcomes of the previous period. At outcome x_T^a, the contract is above the previous period's payment and is strictly increasing. In contrast, in the classical case at outcome x_T^a, the period T contract gives the same payment as that received in period $T-1$. Because of the effect of the wages in period $T-1$ on the reference in period T in our setting, contracts are less front-loaded than in the classical model.

4.4 Optimality and Access to Credit Markets

Under classical assumptions, the described relationship between any two adjacent periods implies that the optimal contract front-loads consumption, i.e., after the realization of the

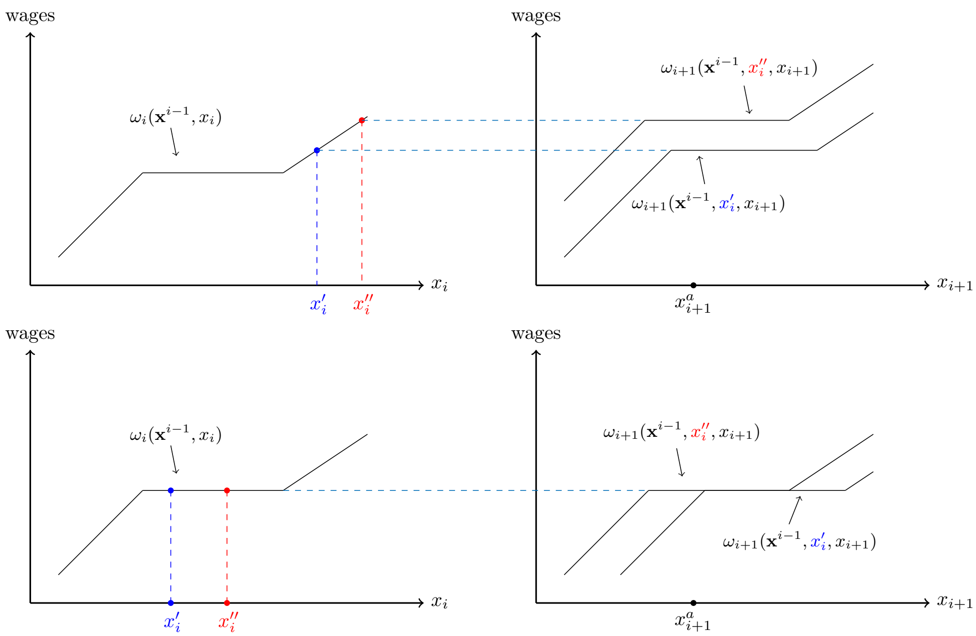

Figure 4: Period i versus period $i+1$ contract.

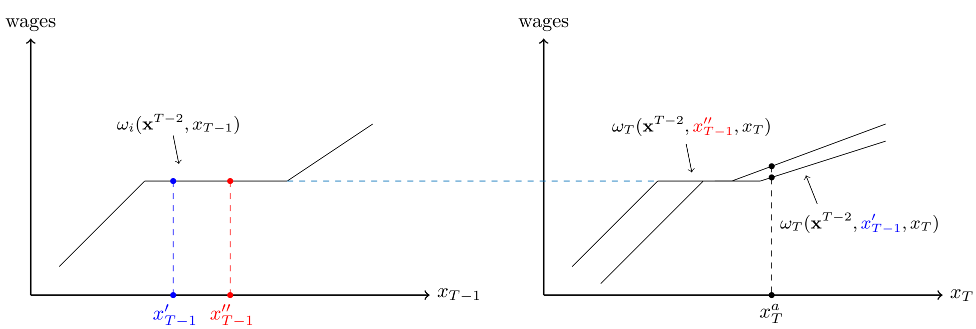

Figure 5: Period $T-1$ versus period T contract.

current period outcome, the agent is left with the desire to transfer resources to the future (Rogerson 1985). In our model, however, this will not necessarily be the case. If the agent had access to credit markets, he might prefer to move resources to or from the future at the going interest rate, depending on the parameters of the problem.

Assume the agent is ex-post given the opportunity to transfer resources over time. Let r^l be the rate that would ex-post leave him indifferent between consuming his full allocation and saving to consume in the following period. Similarly, let r^b denote the rate that would ex-post leave him indifferent between consuming his wage and borrowing. Proposition 7 in the appendix states that under the optimal contract, $r_l > r_b$. Thus, there is a gap between the rate at which the agent is willing to loan his savings and the rate at which he is willing to borrow. This gap between willingness to accept and willingness to pay has been described in other economic contexts (Bateman et al. 1997). Depending on whether $\frac{1}{\delta} - 1$ is below or above r^l and r^b, the agent may prefer to save, borrow or consume his allocation if allowed to save or borrow at rate $\frac{1}{\delta} - 1$. Intuitively, if the agent were to save, current consumption would fall under the reference, and he would incur a loss. If he were to borrow, then future consumption would fall under the reference with positive probability, again incurring a loss. Then, the agent would be better off consuming the current wage in a manner similar to the "status quo bias" described by Samuelson and Zeckhauser (1988) and the "endowment effect" of Knetsch (1989).

An important consequence of this analysis is that under our assumptions, the optimal payment scheme may remain optimal when the agent has access to a savings technology. For example, if the primitives of the repeated moral hazard problem are such that the contract gives payments at the reference for all outcomes in all periods but the last, from Proposition 7, we know that the agent will not want to save if given access to credit markets. Consider now the program the principal solves to provide incentives to an agent who has access to savings at the market interest rate $1 - 1/\delta$ but is unable to borrow. The principal is risk neutral; thus, it is without loss to assume that the principal pays the agent a wage such that the agent does not have incentives to save. Therefore, the program the principal must solve is the original repeated moral hazard problem plus the constraint that zero savings is the optimal decision of the agent. This constraint can be written as

$$0 \in \text{argmax}_{(s_0,s_1,\ldots,s_T)\geq 0, s_{T-1}=s_T/\delta} \sum_{i=0}^{T} \delta^i \mathbb{E}\left(\tilde{U}(w_i - s_i, w_{i-1} - s_{i-1}) - \psi(a_i) | a_0, a_1, \ldots a_i\right)$$

However, note that our original program is a relaxed version of the problem with access to savings. If, under the optimal contract, the constraint is satisfied, the contract must also be optimal under constrained savings. In contrast with the classical model, this program may be solved at $(s_0, \ldots, s_T) = (0, \ldots, 0)$ in our setting.

Analogously, if the agent has access to saving and borrowing at rates of r^l and r^b, respectively, and after each outcome, the agent is better off consuming the entire wage, the optimal contract without access to credit markets is also optimal with unconstrained access to credit.

4.5 Renegotiation Proofness

As defined in Chiappori et al. (1994), we define a contract to be renegotiation proof if at every contracting date, the continuation remains optimal for the remaining periods. In our framework, the optimal contract scheme is renegotiation proof just as in the classical case. In fact, Chiappori et al. (1994) has shown that the renegotiation-proofness proposition does not rely on the differentiability of the utility function. Underlying this result is the assumption that the agent is able to predict how his utility will update in each period. If this were not the case, the proposition might not hold.

Chiappori et al. (1994) shows that in the classical case, renegotiation-proof contracts that imply non-randomized savings decisions generally cannot implement efforts above the minimum. In contrast, in our setting, the optimal contract we derive is renegotiation proof if after each outcome, the agent does not have incentives to reallocate consumption from or to the future. In fact, based on the discussion in 4.4, we know that the optimal contract remains optimal with unconstrained credit under some circumstances.

Renegotiation proofness follows by an argument described in Chiappori et al. (1994). If the original optimal contract is not renegotiation proof then after a positive measure set of outcomes, the principal and agent can write new contracts that leave the agent's utility unchanged and strictly increase the utility of the principal. If we replace the continuation of our original optimal contract by these new contracts, after their respective outcomes, we obtain a contract that gives a strictly better payoff to the principal, contradicting the optimality of the original contract.

5 Conclusions

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.

When the agent has no access to credit markets and is forced to consume his earnings, we find that the optimal payment scheme can have flat segments at the reference, i.e., the wage may be insensitive to outcomes in an interval. This proposition 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, this model predicts a "status quo bias" whenever the agent is allowed to transfer resources over time 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.

In summary, 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 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 Salanié 2000, Salanié 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 Holmström 1994). That is, our model predicts features of the optimal contract that are more in line with the empirical findings, while simultaneously 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.

6 Appendix

6.1 Proof of Lemma 1

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

$$\tilde{U_0}(c_0,R_0) = U(c_0) - \ell_0\theta(c_0,R_0)\,(U(R_0)-U(c_0))\,. \tag{12}$$

Define

$$U(c_0) = \begin{cases} \tilde{U_0}(c_0,R_0) & \text{if } c_0 \geq R_0 \\ L\cdot\tilde{U_0}(c_0,R_0)+(1-L)\cdot\tilde{U_0}(R_0,R_0) & \text{if } c_0 \leq R_0 \end{cases},$$

where $L = \frac{\tilde{U_0}'^{+}(R_0)}{\tilde{U_0}'^{-}(R_0)}$ and $\tilde{U_0}'^{+}(R_0)$ and $\tilde{U_0}'^{-}(R_0)$ denote the right and left derivatives of $\tilde{U_0}$ at R_0, respectively. As defined, U is continuous and differentiable with $U'(R_0) = \tilde{U_0}'^{+}(R_0)$. Define $\ell_0 = \frac{1-L}{L}$.

Let us see that equation (12) holds. By definition, it holds for $c_0 \geq R_0$. It is also concave by the concavity of $\tilde{U_0}(c_0,R_0)$ in c_0.
If $c_0 \leq R_0$ we have

$$\begin{aligned}
&U(c_0) - \ell_0\theta(c_0,R_0)\,(U(R_0)-U(c_0)) = \\
&L\cdot\tilde{U_0}(c_0,R_0)+(1-L)\tilde{U_0}(R_0,R_0) - \frac{1-L}{L}\left(\tilde{U_0}(R_0,R_0)-\tilde{U_0}(c_0,R_0)L-(1-L)\tilde{U_0}(R_0,R_0)\right) = \\
&L\cdot\tilde{U_0}(c_0,R_0)+(1-L)\tilde{U_0}(R_0,R_0) - \frac{1-L}{L}L\left(\tilde{U_0}(R_0,R_0)-\tilde{U_0}(c_0,R_0)\right) = \tilde{U_0}(c_0,R_0),
\end{aligned}$$

which completes the proof.

6.2 Proof of Proposition 1

Let's see that $h_i(v_0, v_1, \ldots, v_i)$ is strictly convex. Because U is strictly increasing we can write $h_i(v_0, v_1, \ldots, v_i) = U^{-1}(U(h_i(v_0, v_1, \ldots, v_i)))$. To establish the convexity of h, 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).[27] Let $\mathbf{v}^i = (v_0, \ldots, v_i)$ and $\mathbf{v}'^i = (v'_0, \ldots, v'_i)$ be two utility provision vectors. Denote $\mathbf{v}^{i-1} = (v_0, v_1, \ldots, v_{i-1})$ for $i \geq 1$ and $\mathbf{v}^{-1} = \emptyset$. By the definition of convexity, we need to show that for all $\lambda \in (0,1)$,

$$U(h_i(\lambda(v_0, \ldots, v_i) + (1-\lambda)(v'_0, \ldots, v'_i))) < \lambda U(h_i(v_0, v_1, \ldots, v_i)) + (1-\lambda)U(h_i(v'_0, v'_1, \ldots, v'_i)). \tag{13}$$

Note that for $i = 0$, by (5) $U(h_0(v_0))$ is linear by parts, increasing and convex because the derivative for $v_0 < R_0$ is $\frac{1}{1+\ell_0} < 1$ and for $v_0 > R_0$ it is $\frac{1}{1+\bar{\theta}\ell_0}$. By induction, we assume that (13) holds for $i-1$ and prove that it must hold for i.

Note first that

$$v_i \geq U(h_{i-1}) \iff \frac{v_i + \ell_i \bar{\theta} U(h_{i-1})}{1 + \ell_i \bar{\theta}} \geq \frac{v_i + \ell_i U(h_{i-1})}{1 + \ell_i} \tag{14}$$

The utility provision in period i can be written as follows, with $\theta \in \{1, \bar{\theta}\}$ depending on whether i's utility provision is in the gain or loss area.

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

The first inequality is implied by the induction hypothesis. The second inequality is justified by noting that if period i and period $i-1$ utilities are both in the gain or both in the loss area for utility realization $\mathbf{v}^i$, then $U(h_i(\mathbf{v}^i)) = \frac{v_i + \ell_i \theta U(h_{i-1}(\mathbf{v}^{i-1}))}{1+\theta \ell_i}$. Now, if the payment in period $i-1$ is in the loss area and the payment in period i is in the gain area then $\theta = 1$ and

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

$v_i \geq U(h_{i-1}(\mathbf{v}^{i-1}))$ which from (14) implies

$$U(h_i(\mathbf{v}^i)) = \frac{1+\ell_i\bar{\theta}U(h_{i-1}(\mathbf{v}^{i-1}))}{1+\ell_i\bar{\theta}} \geq \frac{1+\ell_i\theta U(h_{i-1}(\mathbf{v}^{i-1}))}{1+\ell_i\theta}.$$

Analogously, if the payment in period $i-1$ is in the gain area and the payment in period i is in the loss area, then $\theta = \bar{\theta}$ and $v_i \leq U(h_{i-1})$, which, in turn, implies from (14) that

$$U(h_i(\mathbf{v}^i)) = \frac{1+\ell_i U(h_{i-1}(\mathbf{v}^{i-1}))}{1+\ell_i} \geq \frac{1+\ell_i\theta U(h_{i-1}(\mathbf{v}^{i-1}))}{1+\ell_i\theta}.$$

Identical arguments can be made for realization $\mathbf{v}^i$, thus, establishing the convexity of $U(h_i)$.

6.3 Proof of Proposition 1

To prove Proposition 1, we first write h_i as $U^{-1}(U(h_i))$ and note that the subdifferential set of h_i can easily be derived from the subdifferential set of $U(h_i)$. Then, we show that that the subdifferential of $U(h_i)$ can be characterized as a function of the subdifferential of $U(h_{i-1})$. By induction, we then characterize the subdifferential of $U(h_i)$.

We have

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

where

$$U(h_i(v_0,v_1,\ldots,v_i)) = \begin{cases} \frac{v_i+\ell_i\bar{\theta}U(h_{i-1}(v_0,v_1,\ldots,v_{i-1}))}{1+\ell_i\bar{\theta}} & \text{if } v_i \geq U(h_{i-1}(v_0,v_1,\ldots,v_{i-1})) \\ \frac{v_i+\ell_i U(h_{i-1}(v_0,v_1,\ldots,v_{i-1}))}{1+\ell_i} & \text{if } v_i < U(h_{i-1}(v_0,v_1,\ldots,v_{i-1})) \end{cases}. \quad (15)$$

By Proposition 4.2.5 in Ozdaglar et al. (2003) we know that[28]

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

[28] A vector $d \in \mathbb{R}^n$ is a subgradient of f at a point $x \in \mathbb{R}^n$, denoted $d \in \partial f(x)$, if

$$f(z) \geq f(x) + (z-x)'d \qquad \forall z \in \mathbb{R}^n$$

Now, note that from (15), we can write

$$U(h_i(v_0, v_1, \dots, v_i)) = F_i((U \circ h_{i-1})(v_0, v_1, \dots, v_{i-1}), v_i)$$

where

$$F_i(x,y) = \begin{cases} \frac{y+\ell_i \bar{\theta} x}{1+\ell_i \bar{\theta}} & \text{if } y \geq x \\ \frac{y+\ell_i x}{1+\ell_i} & \text{if } y < x \end{cases}.$$

Let us now see that the subdifferential of $U(h_i)$ can be written as a function of the subdifferential set of $U(h_{i-1})$, as stated in the following Lemma.

Lemma 1. *A vector* $(\bar{d}_0, \dots, \bar{d}_i) \in \partial (U \circ h_i)(v_0, \dots, v_i)$ *if and only if there are vectors*

$$(d_0, \dots, d_{i-1}) \in \partial (U \circ h_{i-1})(v_0, \dots, v_{i-1})$$

and

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

such that

$$(\bar{d}_0, \bar{d}_1, \dots \bar{d}_{i-1}, \bar{d}_i) = (d_0 \cdot \tilde{d}_0, d_1 \cdot \tilde{d}_0, \dots, d_{i-1} \cdot \tilde{d}_0, \tilde{d}_1).$$

Sufficiency

Let $(d_0, \dots, d_{i-1}) \in \partial (U \circ h_{i-1})(v_0, v_1, \dots, v_{i-1})$ 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 us see that $(d_0 \cdot \tilde{d}_0, d_1 \cdot \tilde{d}_0, \dots, d_{i-1} \cdot \tilde{d}_0, \tilde{d}_1) \in \partial (U \circ h_i)(v_0, v_1, \dots, v_i)$. In fact, we have

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

where the first inequality is a consequence of $(d_0, \dots, d_{i-1}) \in \partial (U \circ h_{i-1})(v_0, v_1, \dots, v_{i-1})$ 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 us 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 $(\tilde{d}_0, \tilde{d}_1) \in \partial F_i((U \circ h_{i-1})(v_0, v_1, \dots, v_{i-1}), v_i)$.

Necessity

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

Otherwise, $y = x = F_i(x,y)$, 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,

$$F_i(x + \alpha_0, y + \alpha_1) \geq F(x,y) + \alpha_0 \tilde{d}_0 + \alpha_1 \tilde{d}_1 \qquad \forall \alpha_0, \alpha_1 \in \mathbb{R}. \tag{16}$$

Equation (16) becomes

$$F_i(x+\alpha_0, y+\alpha_1) = \frac{y+\alpha_1+\ell_i\theta(x+\alpha_0)}{(1+\ell_i\theta)} \geq y+\alpha_0\tilde{d}_0+\alpha_1\tilde{d}_1$$

$$\iff \left(\frac{\ell_i\theta}{1+\ell_i\theta}-\tilde{d}_0\right)\alpha_0 \geq \left(\tilde{d}_1-\frac{1}{1+\ell_i\theta}\right)\alpha_1, \tag{17}$$

where $\theta = \bar{\theta}$ if $y \geq x$, and $\theta = 1$ if $y < x$ for all $\alpha_0, \alpha_1 \in \mathbb{R}$. Inequality (17) holds for all $\alpha_0 < \alpha_1$ if and only if $\left(\frac{\ell_i}{1+\ell_i}-\tilde{d}_0\right) = \left(\tilde{d}_1-\frac{1}{1+\ell_i}\right) > 0$ (note, α_0 and α_1 may take negative values). Inequality (17) holds for all $\alpha_0 \geq \alpha_1$ if and only if $\left(\frac{\ell_i\bar{\theta}}{1+\ell_i\bar{\theta}}-\tilde{d}_0\right) = \left(\tilde{d}_1-\frac{1}{1+\ell_i\bar{\theta}}\right) < 0$.

Therefore, to summarize, we have

$$(\tilde{d}_0,\tilde{d}_1) \in \partial F(x,x) \iff \tilde{d}_0 \in \left[\frac{\bar{\theta}\ell_i}{1+\bar{\theta}\ell_i}, \frac{\ell_i}{1+\ell_i}\right], \tilde{d}_1 \in \left[\frac{1}{1+\ell_i}, \frac{1}{1+\bar{\theta}\ell_i}\right] \text{ and } \tilde{d}_0 = 1-\tilde{d}_1.$$

Define the set

$$K_i(x,y) = \begin{cases} \{1\} & \text{if } y < x \\ [\bar{\theta},1] & \text{if } y = x \\ \{\bar{\theta}\} & \text{otherwise.} \end{cases}$$

We can write

$$\partial F_i(x,y) = \left\{ \left(\frac{k_i\ell_i}{1+\ell_i k_i}, \frac{1}{1+\ell_i k_i}\right) : k_i \in K_i(x,y) \right\}.$$

Let us now see that if $(\bar{d}_0,\ldots,\bar{d}_i) \in \partial\,(U \circ h_i)\,(v_0,\ldots,v_i)$ then $(\bar{d}_0,\bar{d}_1,\ldots\bar{d}_{i-1},\bar{d}_i) = (d_0 \cdot \tilde{d}_0, d_1\cdot\tilde{d}_0,\ldots,d_{i-1}\cdot\tilde{d}_0,\tilde{d}_1)$ for some vectors

$$(d_0,\ldots,d_{i-1}) \in \partial\,(U \circ h_{i-1})\,(v_0,\ldots,v_{i-1})$$

and

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

By the definition of subgradient in the v_i direction, we must have

$$F_i\left((U \circ h_{i-1})\,(v_0,\ldots,v_{i-1}), v_i+\alpha_i\right) \geq F_i\left((U \circ h_{i-1})\,(v_0,\ldots,v_{i-1}), v_i\right) + \alpha_i\bar{d}_i \quad \forall\alpha_i$$

and as before, this implies that $\bar{d}_i = \frac{1}{1+\ell_i k_i}$ with $k_i \in K_i\left((U \circ h_{i-1})\,(v_0,\ldots,v_{i-1}), v_i+\alpha_i\right)$.

In points at which F_i is differentiable, its subgradient set coincides with the derivative, which is $(\tilde{d}_0,\tilde{d}_1) = \left(\frac{\ell_i\bar{\theta}}{1+\ell_i\bar{\theta}}, \frac{1}{1+\ell_i\bar{\theta}}\right)$ if $(U \circ h_{i-1})\,(v_0,\ldots,v_{i-1}) < v_i$ and $(\tilde{d}_0,\tilde{d}_1) = \left(\frac{\ell_i}{1+\ell_i}, \frac{1}{1+\ell_i}\right)$

if $(U\circ h_{i-1})(v_0,\ldots,v_{i-1})>v_i$. From Proposition 4.2.5 in Ozdaglar et al. (2003), when F is differentiable, we have $\partial U(h_i(v_0,\ldots,v_i))=\left((1-\bar{d}_i)\cdot\partial(U\circ h_{i-1})(v_0,\ldots,v_{i-1}),\bar{d}_i\right)$, which establishes necessity at points in which $U(h_i(v_0,\ldots,v_i))$ is differentiable.

F is not differentiable when $(U\circ h_{i-1})(v_0,\ldots,v_{i-1})=v_i$. Let $(\alpha_0,\alpha_1,\ldots,\alpha_{i-1})\in\mathbb{R}^i$ and define

$$\hat{\alpha}_i=(U\circ h_{i-1})(v_0+\alpha_0,\ldots,v_{i-1}+\alpha_{i-1})-v_i.$$

Note that

$$U(h_i(v_0+\alpha_0,\ldots,v_{i-1}+\alpha_{i-1},v_i+\hat{\alpha}_i))=F(v_i+\hat{\alpha}_i,v_i+\hat{\alpha}_i)=v_i+\hat{\alpha}_i$$

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

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

Thus, defining $(d_0,\ldots,d_{i-1})=(\bar{d}_0,\ldots,\bar{d}_i)\frac{1}{(1-\bar{d}_i)}\in\partial(U\circ h_{i-1})(v_0,\ldots,v_{i-1})$ and $(\tilde{d}_0,\tilde{d}_1)=\left((1-\bar{d}_i),\bar{d}_i\right)\in\partial F((U\circ h_{i-1})(v_0,\ldots,v_{i-1}),v_i)$, we find that $(\bar{d}_0,\bar{d}_1,\ldots\bar{d}_{i-1},\bar{d}_i)=(d_0\cdot\tilde{d}_0,d_1\cdot\tilde{d}_0,\ldots,d_{i-1}\cdot\tilde{d}_0,\tilde{d}_1)$.

□

We can now establish Proposition 1. We may now derive $\partial U(h_i(v_0,\ldots,v_i))$ inductively using Lemma 1. Let $\tilde{K}_i(\mathbf{v}^i)$ denote the set $K_i((U\circ h_{i-1})(v_0,\ldots,v_{i-1}),v_i)$ and $\tilde{K}_0(\mathbf{v}^0)$ denote the set $K_0(U(R_0),v_0)$. We have

$$\partial U(h_0(v_0))=\left\{\frac{1}{1+k_0\ell_0}:k_0\in\tilde{K}_0\left(\mathbf{v}^0\right)\right\},$$

therefore

$$\partial U(h_1(v_0,v_1))=\left\{\left(\frac{k_1\ell_1}{1+k_1\ell_1}\cdot\frac{1}{1+k_0\ell_0},\frac{1}{1+k_1\ell_1}\right):k_0\in\tilde{K}_0(\mathbf{v}^0),k_1\in\tilde{K}_1(\mathbf{v}^1)\right\}$$

and

$$\partial U(h_2(v_0,v_1,v_2))=\left\{\left(\frac{k_2\ell_2}{1+k_2\ell_2}\cdot\frac{k_1\ell_1}{1+k_1\ell_1}\cdot\frac{1}{1+k_0\ell_0},\frac{k_2\ell_2}{1+k_2\ell_2}\cdot\frac{1}{1+k_1\ell_1},\frac{1}{1+k_2\ell_2}\right):k_i\in\tilde{K}_i(\mathbf{v}^i),i\in\{0,1,2\}\right\}$$

and inductively, we obtain (7). That is,

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

6.4 Proof of Proposition 2

The proof of Proposition 2 includes the following steps. First, lemma 2 shows that we can consider a smaller set of incentive constraints. Then, we construct a Lagrangian using the techniques illustrated in Rockafellar (1974). Finally, by imposing that zero belongs to the subdifferential set of the Lagrangian evaluated at the optimal contract, we derive the necessary conditions stated in Proposition 2. The necessary conditions for optimality are analogous to the KKT conditions in finite dimensional spaces.

Lemma 2. *The incentive constraint is satisfied if and only if*

$$\sum_{j=i}^{T}\delta^j\left(\int\tilde{U}_j(\omega_j(x_0,\dots x_j),R_j)f^i_{a_i}(x_i|a_i)\cdot f^{i+1}(x_{i+1}|a_{i+1})\cdots f^T(x_T|a_T)dx_i\cdots dx_T\right)-\Delta\psi_i\geq 0 \tag{IC''}$$

for every sequence of outcomes $(x_0,\dots,x_{i-1})$.

Proof. Necessity is straightforward. For sufficiency, note that in the last period, the agent exerts high effort after every history of outcomes. In period $T-1$, the best deviation involves exerting effort a_H in period T after every history. Thus, we must only ensure that there is no deviation to a_L at period $T-1$ after a history given that a_H is exerted in period T. Reasoning recursively, we can conclude that IC" implies *IC*.

□

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

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

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

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

Let $m = (\tilde{m}_0, m_0, m_1, \dots m_T) \in \mathscr{M} = \mathbb{R}^2 \times L^1([\underline{x}_0, \bar{x}_0]) \times L^1([\underline{x}_0, \bar{x}_0] \times [\underline{x}_1, \bar{x}_1]) \times \cdots \times L^1([\underline{x}_0, \bar{x}_0] \times \cdots \times [\underline{x}_T, \bar{x}_T]))$. Following the notation in Rockafellar (1974), we define

$$F(v,m) = \begin{cases} f_0(v_0, v_1, \dots, v_T) & \text{if } g_0(v_0, v_1, \dots, v_T) \geq \tilde{m}_0, \\ & r_i(v_0, v_1, \dots, v_T) \geq m_i(x_0, x_1, \dots, x_{i-1}) \quad \text{a.s for } i = 0, \dots T+2 \\ -\infty & \text{otherwise.} \end{cases}$$

$-F(v,m)$ is closed in m because the sets $\{m | F(v,m) \geq \alpha\}$ are closed for all $\alpha \in \mathbb{R}$. $-F(v,m)$ is also convex in m.[29]

Following Rockafellar (1974), equation 4.2, the Lagrangian function K is defined as

$$K(v,y) = \sup\{F(v,m) + < m, y > | m \in \mathscr{M}\}.$$

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

$$< m, y > = \tilde{y}_0 \tilde{m}_0 + y_0 m_0 + \mathbb{E}_{\tilde{\mu}^1}(y_1 m_1) + \mathbb{E}_{\tilde{\mu}^2}(y_2 m_2) + \dots + \mathbb{E}_{\tilde{\mu}^T}(y_T m_T).$$

In our case K is equal to

$$K(\cdot, y) = \begin{cases} f_0 + g_0 \tilde{y}_0 + r_0 y_0 + \mathbb{E}_{\tilde{\mu}^1}(y_1 r_1) + \mathbb{E}_{\tilde{\mu}^2}(y_2 r_2) + \dots + \mathbb{E}_{\tilde{\mu}^T}(y_T r_T) & \text{if } y \geq 0 \\ +\infty & \text{otherwise.} \end{cases}$$

where $y \geq 0$ if all components are positive almost everywhere. Let $y = (\lambda, \mu_0, \dots, \mu_T)$. Then, the Lagrangian becomes

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

[29] We need to ensure that $F(v,m)$ is concave in m. That is, we must determine that for $m^1, m^2 \in \mathscr{M}$, if $F(v,m^1) \neq -\infty$, $F(v,m^2) \neq -\infty$, then $F(v, \alpha m^1 + (1-\alpha) m^2) \neq -\infty$. This follows because $\alpha m^1_j(x_0, \dots, x_{j-1}) + (1-\alpha) m^2_j(x_0, \dots, x_{j-1}) \leq \max\{m^1_j(x_0, \dots, x_{i-1}), m^2_j(x_0, \dots, x_{i-1})\}$ for $j \in \{0, \dots, T+2\}$ (see Rockafellar (1974) page 7).

where we denote

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

Recall 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 $0 \in \partial_y(-K(\bar{v},\bar{y}))$.

Because $-F(v,u)$ is closed and convex in u from Theorem 15 in Rockafellar (1974) ((e) $\iff$ (f)), we know that if $(\bar{v},\bar{y})$ satisfies the Kuhn-Tucker condition, then $\bar{v}$ solves the principal's problem given by equations (5) and (6). 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 implies that complementary slackness is satisfied for constraints (PC') through (5). 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 $0 \in \partial_y(-K(\bar{v},\bar{y}))$.

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

From Theorem 22 of Rockafellar (1974) (condition c. for $p=1$ and $q=\infty$), we know that a subgradient set of $-K$ is the expectation of the subgradient set of the integrand.

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

The expression $\partial v_j(x_0,x_1,\ldots,x_j)$ denotes the subgradient of the function that associates a vector to the j'th component of the vector evaluated at the vector

$$(v_0(x_0),v_1(x_0,x_2),\ldots,v_T(x_0,x_1,\ldots x_T)).$$

Therefore, the component i of the subgradient of $(\partial v_j)_i$ equals zero whenever $j \neq i$ and is equal to 1 otherwise. Thus, we can see that the kth component of the interior summation in the last term of the previous expression for $k \geq i$ is given by

$$\left(\sum_{j=i}^{T} \delta^j \left(\Delta_{a_i} \mathbb{E}\left(\mu_i(x_0, x_1, \ldots, x_{i-1}) \cdot \partial v_j(x_0, x_1, \ldots, x_j) | a_i, \ldots, a_j\right)\right)\right)_k = \Delta_{a_i} \mathbb{E}\left(\delta^k \mu_i(x_0, \ldots, x_{i-1}) | a_i, \ldots, a_k\right).$$

Therefore, we can write,

$$\left(\sum_{i=0}^{T} \left(\sum_{j=i}^{T} \delta^j \left(\Delta_{a_i} \mathbb{E}\left(\mu_i(x_0, x_1, \ldots, x_{i-1}) \cdot \partial v_j(x_0, x_1, \ldots, x_j) | a_i, \ldots, a_j\right)\right)\right)\right)_k = \sum_{i=0}^{k} \Delta_{a_i} \mathbb{E}\left(\delta^k \mu_i(x_0, \ldots, x_{i-1}) | a_i, \ldots, a_k\right).$$

Substituting in Proposition 1 the expression for the sub-differential of h_i and noting that setting component j of $\partial(-K(v,y))$ equal to 0 means that for every $\psi \in L^1\left([\underline{x}_0, \bar{x}_0] \times \cdots \times [\underline{x}_j, \bar{x}_j]\right)$ the following condition must hold for each $j \in \{0, \ldots, T\}$:

$$\begin{aligned} 0 = \sum_{i=j}^{T} \delta^{i-j} \mathbb{E}\Bigg(& \frac{1}{U'(\omega_i(x_0, x_1, \ldots, x_i))} \left(\prod_{t=j+1}^{i} \frac{k_t(x_0, x_1, \ldots, x_t)\ell_t}{1 + k_t(x_0, x_1, \ldots, x_t)\ell_t} \right) \\ & \times \frac{\psi(x_0, x_1, \ldots, x_j)}{1 + k_j(x_0, x_1, \ldots, x_j)\ell_j} \Big| a_0, a_1, \ldots, a_i \Bigg) \\ & - \lambda \mathbb{E}\left(\psi(x_0, x_1, \ldots, x_j) \big| a_0, a_1, \ldots, a_j\right) \\ & - \sum_{i=0}^{j} \left(\Delta_{a_i} \mathbb{E}\left(\mu_i(x_0, x_1, \ldots, x_{i-1}) \psi(x_0, x_1, \ldots, x_j) \big| a_0, a_1, \ldots, a_j\right)\right). \end{aligned} \tag{19}$$

This is because the jth component of the subdifferential is a linear function from $L^1([\underline{x}_0, \bar{x}_0] \times \cdots \times [\underline{x}_j, \bar{x}_j])$ to $\mathbb{R}$, and a linear function is zero when its evaluation on every test function is zero. Note that we can now write

$$\begin{aligned} & \Delta_{a_i} \mathbb{E}\left(\mu_i(x_0, \ldots, x_{i-1}) \psi(x_0, \ldots, x_j) \big| a_0, a_1, \ldots, a_j\right) = \\ & \int \mu_i(x_0, \ldots, x_{i-1}) \cdot \psi(x_0, \ldots, x_j) \frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)} (f^0(x_0|a_0) \cdot \cdots \cdot f^j(x_j|a_j)) dx_0 \cdots dx_i \cdots dx_j = \\ & \mathbb{E}\left(\mu_i(x_0, x_1, \ldots, x_{i-1}) \cdot \psi(x_0, x_1, \ldots, x_j) \frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)} | a_0, \ldots a_j\right). \end{aligned}$$

Because (19) must be true for every function $\psi \in L^\infty([\underline{x}_0, \bar{x}_0] \times \cdots \times [\underline{x}_j, \bar{x}_j])$, re-arranging terms, we find that (19) is equivalent to

$$\begin{aligned} \frac{1}{U'(\omega_j)} \cdot \frac{1}{1 + k_j \ell_j} + \sum_{i=j+1}^{T} \delta^{i-j} \mathbb{E}\left(\frac{1}{U'(\omega_i)} \left(\prod_{t=j+1}^{i} \frac{k_t \ell_t}{1 + k_t \ell_t} \right) \frac{1}{1 + k_j \ell_j} \Big| a_{j+1}, \ldots, a_i \right) &= \lambda + \sum_{i=0}^{j} \mu_i \frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)} \\ &= \lambda_j + \mu_j \frac{f^j_{a_j}(x_j|a_j)}{f^j(x_j|a_j)}. \end{aligned} \tag{20}$$

where $\lambda_j = \lambda + \sum_{i=0}^{j-1} \mu_i \frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}$ for $j \in \{1,\ldots,T\}$. Multiplying the equation for $j+1$ by $k_{j+1}\ell_{j+1}$ and to $f^{j+1}(x_{i+1}|a_{j+1})$, we obtain

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

Replacing this last expression in the $j+1$ term of the sum in (20), (9) and (10) are obtained.

6.5 Proof of Proposition 5

The proof of Proposition 5 consists of four steps. First, we prove that the contract is non-decreasing. Second, we prove that it must be continuous. Third, we prove that the payment schedule corresponds to the reference in an interval of outcomes. Finally, we prove that the payment scheme of the last period is sensitive to all periods' outcomes.

The optimal contract is non-decreasing

In order to establish that the contract is non-decreasing we first prove some simple properties of functions that are not non-decreasing. Let μ denote the Lebesgue measure in X_i^2.

Definition 1. We say that a function w_i is *non-decreasing almost surely in x_i* if there is a set $E \subseteq X_i$ with $\mu(E) = 1$ such that for every $x,y \in E$ $x \leq y \implies w_i(x_0,x_1,\ldots,x) \leq w_i(x_0,x_1,\ldots,y)$.

Lemma 3. *If w_i is not non-decreasing almost surely in x_j with $j \leq i$, then there is an outcome $\tilde{x} \in X_i$, a positive constant $\nu > 0$ and positive measure sets $A \subseteq [\underline{x}_i,\tilde{x}]$ and $B \subseteq [\tilde{x},\bar{x}_i]$ such that $w_i(x_0,\ldots,x) > w_i(x_0,\ldots,y)+\nu$ for every $x \in A$ and $y \in B$.*

Proof. The negation of the claim "w_i is non-decreasing almost surely" , is that there is a positive measure set $\tilde{A} \subseteq X_i$ and positive measure sets $C_x \subseteq X_i$ for each $x \in \tilde{A}$, for each $y \in C_x$, $x < y$ and $f(x_j) := w_i(x_0,\ldots x_j \ldots,x_i) > w_i(x_0,\ldots y_j \ldots,x_i) = f(y_j)$. Let $\bar{A}$ denote $\{(x,y)|x \in A, y \in C_x\} \subseteq \mathbb{R}^2$ and define

$$\bar{A}_n = \{(x,y)|x \in \tilde{A}, y \in C_x, y > x+\frac{1}{n}, \quad f(y)+\frac{1}{n} < f(x)\}.$$

Because $\bar{A}_n \subseteq \bar{A}_{n+1}$ and $\cup_n \bar{A}_n = \bar{A}$, we must have $\lim_n \mu(\bar{A}_n) = \mu(\bar{A}) > 0$. Thus, there must be an $\varepsilon = 1/n$ such that $\mu(\bar{A}_{1/\varepsilon}) > 0$. Define $\bar{y} = inf\{a \in X_i|\mu((x,y) \in \bar{A}_{1/\varepsilon}, y > a) = 0\}$ and

$\bar{f} = inf\{a \in \mathbb{R} | \mu((x,y) \in \bar{A}_{1/\varepsilon}, y > \bar{y} - \varepsilon, f(x) > a) = 0\}$. Define $\tilde{x} = \bar{y} - \varepsilon$, $A = \{x | (x,y) \in A_{1/\varepsilon}, y > \bar{y} - \varepsilon, f(x) > \bar{f} - \frac{\varepsilon}{2}\}$ and $B = \{y | (x,y) \in A_{1/\varepsilon}, y > \bar{y} - \varepsilon, f(x) > \bar{f} - \frac{\varepsilon}{2}\}$. From the definitions of $\bar{y}$ and $\bar{f}$, both A and B must have positive measures because they are defined as infima. Furthermore, for every $x \in A$ and $y \in B$ we have $y \geq x$ (since in A_n, $y > x + 1/n$) and $f(x) > f(y) + \frac{\varepsilon}{2}$ (since $f(x) > f(y) + 1/n$ in A_n). Finally, from the definitions of A and B, we have $A \subseteq [\underline{x}_i, \tilde{x}]$ and $B \subseteq [\tilde{x}, \bar{x}_i]$.

□

Using Lemma 3 we now prove that the contract is non-decreasing almost surely. This result together with continuity (Lemma 4 below) establishes that the optimal contract is non-decreasing.

Proposition 6. *The optimal wage schedule in period i $\omega_i(x_0, x_1, \ldots, x_i)$ is non-decreasing almost surely in x_i.*

Proof. We will argue by contradiction. If ω_i is not non-decreasing in x_i, then from Lemma 3, there must be an outcome $\tilde{x}$ and $\nu > 0$, a set $A \subseteq [\underline{x}_i, \tilde{x}]$ and a set $B \subseteq [\tilde{x}, \bar{x}_i]$ such that $\omega_i(x_0, x_1, \ldots, x) > \omega_i(x_0, x_1, \ldots, y) + \nu$ for every $x \in A$ and $y \in B$. Suppose that instead of offering $\omega_i(x_0, \ldots, x_i)$, the principal offers a contract $\tilde{\omega}_i(x_0, x_1, \ldots, x_i)$ that coincides with $\omega_i(x_0, \ldots x_i)$ outside of A and B, reduces the wage for outcomes in A by an amount $\nu_A > 0$ and increases the wage for outcomes in B by an amount $\nu_B > 0$. Let $\tilde{v}_i(x_0, \ldots, x_i)$ denote the utility from contract $\tilde{\omega}_i(x_0, \ldots, x_i)$ as a function of the outcomes. The constants ν_A and ν_B can be chosen so that

$$\mathbb{E}_{x_i}\left((v_i(x_0, \ldots, x_i) - \tilde{v}_i(x_0, \ldots, x_i))\, \mathbf{1}\{x_i \in (A \cup B)\}\right) = 0 \tag{21}$$

where $\mathbf{1}\{x_i \in (A \cup B)\}$ denotes the indicator function of the set $(A \cup B)$. The new contract $\tilde{\omega}_i$ is measurable with respect to $\mathbf{v}^{i-1}$. The contracts in periods $j \in \{i+1, \ldots, T\}$ are modified so that the agent receives the same utility after every outcome in periods $i+1$ through T. Because the agent's utility is the same under the new contract as it is under the original contract, his incentives from period $i+1$ on are unaffected by the change in contract in period i.

The difference in the left hand side of the IC constraint between contracts ω_i and $\tilde{\omega}_i$ is given by:

$$\int (v_i(x_0, \ldots, x_i) - \tilde{v}_i(x_0, \ldots, x_i))\, \mathbf{1}\{x_i \in (A \cup B)\} \frac{f_{a_i}(x_i|a_i)}{f(x_i|a_i)} f(x_i|a_i) dx_i =$$
$$\int ((v_i(x_0, \ldots, x_i) - \tilde{v}_i(x_0, \ldots, x_i))\, \mathbf{1}\{x_i \in (A \cup B)\}) \left(\frac{f_{a_i}(x_i|a_i)}{f(x_i|a_i)} - \frac{f_{a_i}(\tilde{x}|a_i)}{f(\tilde{x}|a_i)}\right) \cdot f(x_i|a_i)\, dx_i \leq 0.$$

The first equality follows from equation (21) because from MLRP, $\frac{f_{a_i}(x_i|a_i)}{f(x_i|a_i)} - \frac{f_{a_i}(\tilde{x}|a_i)}{f(\tilde{x}|a_i)} \geq 0$ if and only if $x_i \geq \tilde{x}$ and $\tilde{\omega}_i \geq \omega_i$ for $x_i \geq \tilde{x}$ and $\omega_i \leq \tilde{\omega}_i$ for $x_i \leq \tilde{x}$. Thus, the IC constraint in period i holds for contract $\tilde{\omega}_i$.

Now, because the cost of providing utility is convex in v_i (Proposition 1) and v_i is a mean-preserving spread of $\tilde{v}_i$, the cost of the new contract $\tilde{\omega}_i$ must be strictly below the cost of contract ω_i. This contradicts the optimality of contract ω_i.

□

Lemma 4. *The optimal wage schedule in period i $w_i(x_0,\ldots,x_i)$ is continuous in x_i.*

Proof. Let us denote $F([m,n]) = \int_m^n f(x_i|a_i)\,dx_i$ and $F_{a_i}([m,n]) = \int_m^n f_{a_i}(x_i|a_i)\,dx_i$. Because w_i is non-decreasing, it has at most countably many points of discontinuity. Suppose w_i jumps upward at outcome y. Denote

$$\lim_{x\downarrow y} w_i(x_0,\ldots,x) = w^+ \text{ and } \lim_{x\uparrow y} w_i(x_0,\ldots,x) = w^-$$

with $w^+ > w^-$. Define $\tilde{w}_i$ to be equal to w_i, except that it lowers utility by $\frac{z}{F([y,y+\varepsilon])}$ for outcomes in the interval $[y,y+\varepsilon]$ and increases utility by $\frac{z}{F([y-\varepsilon,y])}$ for outcomes in the interval $[y-\varepsilon,y]$. Define $\tilde{\omega}_j$ for periods $j \in \{i+1,\ldots,T\}$ so that the agent's utility in those periods is the same as in the original contract after every outcome. Let $\tilde{v}_j$, v_j denote the period $j \in \{0,\ldots,T\}$ outcome contingent utilities that correspond to contracts $\tilde{w}_j$ and w_j, respectively. The cost difference between w_i and $\tilde{w}_i$ is given by

$$\sum_{j\geq i} \mathbb{E}\left(h_j(\mathbf{v}^{i-1},\tilde{v}_i,\ldots\tilde{v}_j) - h_j(\mathbf{v}^{i-1},v_i,\ldots v_j)\right).$$

For small z and ε, the cost difference is approximately equal to

$$\sum_{j\geq i}\left(\frac{\partial}{\partial v_i}h_j\left(\mathbf{v}^{i-1},\tilde{v}_i,\ldots\tilde{v}_j\right) - \frac{\partial}{\partial v_i}h_j\left(\mathbf{v}^{i-1},v_i,\ldots v_j\right)\right)z,$$

This expression is strictly positive because of the convexity of h_j established in Proposition 1, and it is of first order in z.

By construction, the new contract $\tilde{w}_i$ satisfies the individual rationality constraint. Now, the left hand side of the incentive constraint changes by

$$F_{a_i}([y-\varepsilon,y])\frac{z}{F([y-\varepsilon,y])} - F_{a_i}([y,y+\varepsilon])\frac{z}{F([y,y+\varepsilon])}.$$

Because $f_{a_i}(x_i|a_i)/f(x_i|a_i)$ and $\frac{d}{dx_i}f_{a_i}(x_i|a_i)/f(x_i|a_i)$ are continuous and, hence, bounded, this expression converges to zero at rate approximately $z\cdot\varepsilon\cdot\frac{d}{dx_i}\left(f_{a_i}(x_i|a_i)/f(x_i|a_i)\right)$ for small ε and z. Thus, by offering $(\tilde{w}_j)_{j\in\{i,\ldots,T\}}$ instead of $(w_i)_{i\in\{i,\ldots,T\}}$ from period i on, the principal can reduce costs by an amount of order z while satisfying the (IR) constraint and violating

the (IC) constraint by an amount that is of, at most, second order.
Contracts from period 1 on have a flat segment at the reference.

□

We now prove that the reference must be paid in an interval of outcomes in all periods after the first one. The following equality must be true at the optimal contract

$$\frac{1}{U'(\omega_i(x_0,x_1,\ldots,x_i))} = (1+k_i(x_0,x_1,\ldots,x_i)\ell_i)\left(\lambda_i+\mu_i\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}\right)+ \\ -\delta\ell_{i+1}\int k_{i+1}(x_0,x_1,\ldots,x_{i+1})\left(\lambda_{i+1}+\mu_{i+1}\frac{f^{i+1}_{a_{i+1}}(x_{i+1}|a_{i+1})}{f^{i+1}(x_{i+1}|a_{i+1})}\right)f^{i+1}(x_{i+1}|a_{i+1})dx_{i+1}. \quad (22)$$

Because $\lambda_i=\lambda_{i-1}+\mu_{i-1}\frac{f^{i-1}_{a_{i-1}}(x_{i-1}|a_i)}{f^{i-1}(x_{i-1}|a_i)}$ by definition, ω_{i-1} satisfies the following equation:

$$\lambda_i(1+k_{i-1}(x_0,x_1,\ldots,x_{i-1})\ell_{i-1}) = \frac{1}{U'(\omega_{i-1}(x_0,x_1,\ldots,x_{i-1}))}+ \\ \delta\ell_i\int k_i(x_0,\ldots,x_i)\left(\lambda_i+\mu_i\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}\right)f^i(x_i|a_i)dx_i. \quad (23)$$

Suppose $\omega_i(x_0,\ldots,x_i)<\omega_{i-1}(x_0,\ldots,x_{i-1})$ for every x_i except at one point at most. Then $k_i(x_0,x_1,\ldots,x_i)=1 \quad \forall x_i$, and (23) becomes[30]

$$\lambda_i=\frac{1}{U'(\omega_{i-1})(1+k_{i-1}\ell_{i-1}-\delta\ell_i)}.$$

Also note that the last term in equation (22) satisfies

$$\delta\ell_{i+1}\int k_{i+1}(x_0,x_1,\ldots,x_{i+1})\left(\lambda_{i+1}+\mu_{i+1}\frac{f^i_{a_{i+1}}(x_{i+1}|a_{i+1})}{f^i(x_{i+1}|a_{i+1})}\right)f^i(x_{i+1}|a_{i+1})dx_{i+1}\le\delta\ell_{i+1}\lambda_{i+1}.$$

where the inequality comes from $\int_{\underline{x}_i}^{x} f^i_{a_i}(x_i|a_i)\,dx_i\le 0$ for $x\in[\underline{x}_i,\bar{x}_i]$ and $k_{i+1}\le 1$. Thus, from (22), we obtain

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

Therefore, if $(\ell_i-\ell_{i+1})\delta\ge k_{i-1}\ell_{i-1}-\ell_i$ and $\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)}>0$, we conclude $\frac{1}{U'(\omega_i)}\ge\frac{1}{U'(\omega_{i-1})}$, which contradicts $\omega_i<\omega_{i-1}$. Note that if $\ell_{i-1}=\ell_i$; then, $(\ell_i-\ell_{i+1})\delta\ge k_{i-1}\ell_{i-1}-\ell_i$ holds.

[30]Note $\int_{\underline{x}_i}^{\bar{x}_i} f^i_{a_i}(x_i|a_i)\,dx_i=F(\bar{x}_i|a_H)-F(\bar{x}_i|a_L)=1-1=0$.

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 at one point). Then, $k_i(x_0,x_1,\ldots,x_i) = \bar{\theta} \quad \forall x_i$ and (23) becomes $\lambda_i = \frac{1}{U'(\omega_{i-1})(1+k_{i-1}\ell_{i-1}-\delta\bar{\theta}\ell_i)}$. Therefore, from (22) and because the last term in (22) is non-negative, we obtain

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

Therefore, if $k_{i-1}\ell_{i-1} - \delta\bar{\theta}\ell_i \geq 0$ (that is, $\ell_{i-1} \geq \delta\ell_i\bar{\theta}$ when $k_{i-1} = 1$ and $\ell_{i-1}\bar{\theta} \geq \delta\ell_i\bar{\theta}$ when $k_{i-1} = \bar{\theta}$), for x_i such that $\frac{f^i_{a_i}(x_i|a_i)}{f^i(x_i|a_i)} < 0$, we obtain $\frac{1}{U'(\omega_i)} \leq \frac{1}{U'(\omega_{i-1})}$, which contradicts $\omega_i > \omega_{i-1}$. Thus, the reference must be reached in an interval in all periods after the first one.

□

The contract of the last period is sensitive to all previous periods' outcomes

Lemma 5.

For $i \in \{0,\ldots,T-1\}$ and any vector of outcomes in periods 0 to i $\mathbf{x}^{i-1} = (x_0,\ldots,x_{i-1})$, there exist two period i outcomes x_i and x'_i and outcomes $(x_{i+1},\ldots,x_T)$ such that $w_T(\mathbf{x}^{i-1},x_i,x_{i+1},\ldots,x_T) \neq w_T(\mathbf{x}^{i-1},x'_i,x_{i+1},\ldots,x_T)$.

Proof. Let us see that $\lambda_k = \lambda + \sum_{j=1}^{k-1} \mu_j \frac{f^j_{a_j}(x_j|a_j)}{f^j(x_j|a_j)}$ for each $k \in \{i+1,\ldots,T+1\}$ varies in x_i for almost every vector of outcomes $(x_{i+1},\ldots,x_{k-1})$.[31] First, $\lambda_{i+1}(\mathbf{x}^{i-1},\tilde{x}_i) = \lambda_i(\mathbf{x}^{i-1}) + \mu_i(\mathbf{x}^{i-1})\frac{f^i_{a_i}(\tilde{x}_i|a_i)}{f^i(\tilde{x}_i|a_i)}$ depends on $\tilde{x}_i$ only through the term $\mu_i(\mathbf{x}^{i-1})\frac{f^i_{a_i}(\tilde{x}_i|a_i)}{f^i(\tilde{x}_i|a_i)}$, which is strictly increasing in $\tilde{x}_i$. Let us see that $\lambda_{i+2}(\mathbf{x}^{i-1},\tilde{x}_i,\tilde{x}_{i+1})$ must vary in $\tilde{x}_i$ for almost every $\tilde{x}_{i+1}$. In fact, if $\lambda_{i+2}(\mathbf{x}^{i-1},x_i,x_{i+1}) = \lambda_{i+2}(\mathbf{x}^{i-1},x'_i,x_{i+1})$ because $\lambda_{i+1}(\mathbf{x}^{i-1},x_i) \neq \lambda_{i+1}(\mathbf{x}^{i-1},x'_i)$, then $\mu_{i+1}(\mathbf{x}^{i-1},x_i) \neq \mu_{i+1}(\mathbf{x}^{i-1},x'_i)$ and

$$\lambda_{i+2}(\mathbf{x}^{i-1},x_i,\tilde{x}_{i+1}) - \lambda_{i+2}(\mathbf{x}^{i-1},x'_i,\tilde{x}_{i+1}) = \left(\mu_{i+1}(\mathbf{x}^{i-1},x_i) - \mu_{i+1}(\mathbf{x}^{i-1},x'_i)\right) \times \left(\frac{f^{i+1}_{a_{i+1}}(\tilde{x}_{i+1}|a_{i+1})}{f^{i+1}(\tilde{x}_{i+1}|a_{i+1})} - \frac{f^{i+1}_{a_{i+1}}(x_{i+1}|a_{i+1})}{f^{i+1}(x_{i+1}|a_{i+1})} \right),$$

which is strictly monotone in $\tilde{x}_{i+1}$. By an analogous argument, we find that λ_{i+3} varies in x_i almost surely in (x_{i+1},x_{i+2}) and reasoning recursively we can conclude that, for every $k > i$,

[31] Almost surely with respect to the measure in the outcome set $[\underline{x}_i,\bar{x}_i] \times \ldots \times [\underline{x}_T,\bar{x}_T]$.

λ_k varies in x_i for almost every vector of outcomes $(x_{i+1},\ldots,x_{k-1})$. Now, fix two vectors of outcomes $\mathbf{x}^T=(x_0,\ldots,x_i,\ldots,x_T)$ and $\left(\mathbf{x}^T\right)'=(x_0,\ldots,x_i',\ldots,x_T)$ and assume without loss that the contract in period T is either in the loss area for both $\mathbf{x}^T$ and $\left(\mathbf{x}^T\right)'$ or the gain area for both outcomes. The contract in period T is given by

$$\frac{1}{U'(w_T(\tilde{\mathbf{x}}^T))}=(1+\ell_T\theta)\left(\lambda_{T+1}(\tilde{\mathbf{x}}^T)\right),$$

for $\tilde{\mathbf{x}}^T\in\{\left(\mathbf{x}^T\right)',\mathbf{x}^T\}$ and fixed $\theta\in\{1,\bar{\theta}\}$. Now, we know that $\lambda_{T+1}(\mathbf{x}^T)\neq\lambda_{T+1}(\left(\mathbf{x}^T\right)')$ for almost every vector $(x_{i+1},\ldots,x_T)$. Thus, $w_T(\mathbf{x}^T)\neq w_T(\left(\mathbf{x}^T\right)')$ almost surely in the interval $(x_{i+1},\ldots,x_T)$, which establishes the result.

□

6.6 Status Quo Bias and Intertemporal Allocation of Resources

Let $r_i^l(x_0,\ldots,x_i)$ be the rate that would ex-post make the net marginal utility of transferring consumption from period i to period $i+1$ equal to zero, and $r_i^b(x_0,\ldots,x_i)$ be the rate that would ex-post make the net marginal utility of transferring consumption from period i to period $i+1$ equal to zero.

Proposition 7. *Assume the agent is given the opportunity to transfer resources over time after receiving the wage in period i. Then for every vector of outcomes $(x_0,\ldots,x_i)$, $r_i^l(x_0,\ldots,x_i)>r_i^b(x_0,\ldots,x_i)$. Additionally, depending on the parameters of the problem, the agent may either wish to delay consumption to period $i+1$, bring consumption forward to period i or consume exactly his allocation.*

Proof. If the agent saves on the margin, ω_i is above the reference in period $i+1$. Analogously, if the agent borrows in period i, ω_i is below the reference in period $i+1$. Thus, the marginal utility of saving in period i at rate r_i^l and consuming the savings in period $i+1$ is given by

$$\begin{gathered}-(1+\mathbb{I}\{\omega_i\le R_i\}\ell_i+\mathbb{I}\{\omega_i>R_i\}\ell_i\bar{\theta})U'(\omega_i)+\\ \ell_{i+1}\delta U'(\omega_i)\int(1+\ell_{i+1}\mathbb{I}\{\omega_{i+1}<\omega_i\}+\ell_{i+1}\bar{\theta}\mathbb{I}\{\omega_{i+1}\ge\omega_i\})f_{i+1}(x_{i+1}|a_{i+1})dx_{i+1}\\ +\delta(1+r_i^l)\int(1+\ell_{i+1}\mathbb{I}\{\omega_{i+1}<\omega_i\}+\ell_{i+1}\bar{\theta}\mathbb{I}\{\omega_{i+1}\ge\omega_i\})U'(\omega_{i+1}(x_{i+1}))f_{i+1}(x_{i+1}|a_{i+1})dx_{i+1}.\end{gathered}\tag{24}$$

The first term is the marginal effect of savings in period i, the second term is the effect of period i's savings on the period $i+1$ reference and the last term is the marginal effect of the excess income from savings in period $i+1$. The marginal utility of borrowing in period i at rate r_i^b and paying back in period $i+1$ is given by

$$
\begin{aligned}
&(1+\mathbb{I}\{\omega_i < R_i\}\ell_i + \mathbb{I}\{\omega_i \geq R_i\}\ell_i\bar{\theta})U'(\omega_i) \\
&-\ell_{i+1}\delta U'(\omega_i)\int(1+\ell_{i+1}\mathbb{I}\{\omega_{i+1} < \omega_i\} + \ell_{i+1}\bar{\theta}\mathbb{I}\{\omega_{i+1} \geq \omega_i\})f_{i+1}(x_{i+1}|a_{i+1})dx_{i+1} \\
&-\delta(1+r_i^b)\int(1+\ell_{i+1}\mathbb{I}\{\omega_{i+1} \leq \omega_i\} + \ell_{i+1}\bar{\theta}\mathbb{I}\{\omega_{i+1} > \omega_i\})U'(\omega_{i+1}(x_{i+1}))f_{i+1}(x_{i+1}|a_{i+1})dx_{i+1}|a_{i+1})dx_{i+1}.
\end{aligned}
\tag{25}
$$

Now, $r_i^l > r_i^b$ because the addition of expressions (24) and (25) with $r_l = r_b = r$ is strictly negative.To justify why the agent may have incentives to save, borrow or consume his allocation, note that if $\omega_i = \omega_{i-1} = y$ and $\omega_{i+1} = \omega_i = y$ (which is a shape that the optimal contract may take), the agent does not have incentives to save for a small enough $\bar{\theta}$ because for $\bar{\theta} = 0$, the marginal utility of saving at rate $r_i^l = \frac{1}{\delta} - 1$ is $-(1+\ell_i)U'(y) + U'(y) < 0$. Additionally, the marginal utility of borrowing at rate $r_i^b = \frac{1}{\delta} - 1$ when $\bar{\theta} = 0$ is given by $U'(y) - (1+\ell_{i+1})U'(y) - \delta\ell_{i+1}U'(y) < 0$.

□

References

BAKER, G., M. GIBBS, AND B. HOLMSTRÖM (1994): "The Internal Economics of the Firm: Evidence from Personnel Data," *The Quarterly Journal of Economics*, 109, 881–919.

BAKER, G. P., M. C. JENSEN, AND K. J. MURPHY (1988): "Compensation and Incentives: Practice vs. Theory," *Journal of Finance*, 43, 593–616.

BARBERIS, N. AND M. HUANG (2001): "Mental Accounting, Loss Aversion, and Individual Stock Returns," *The Journal of Finance*, 56, 1247–1292.

BARBERIS, N., M. HUANG, AND T. SANTOS (2001): "Prospect Theory and Asset Prices," *Quarterly Journal of Economics*, 116, 1–53.

BATEMAN, I., A. MUNRO, B. RHODES, C. STARMER, AND R. SUGDEN (1997): "A Test of the Theory of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 112, 479–505.

BEISSINGER, T. AND C. KNOPPIK (2001): "Downward Nominal Rigidity in West German Earnings, 1975?95," *German Economic Review*, 2, 385–417.

BEWLEY, T. F. (1999): *Why wages don't fall during a recession*, Harvard University Press.

BOLTON, P. AND M. DEWATRIPONT (2005): *Contract Theory*, MIT Press.

BOWMAN, D., D. MINEHART, AND M. RABIN (1999): "Loss Aversion in a Consumption Savings Model," *Journal of Economic Behavior & Organization*, 38, 155–178.

CARD, D. AND G. B. DAHL (2011): "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior*," *The Quarterly Journal of Economics*, 126, 103–143.

CHETTY, R. AND A. SZEIDL (2010): "Consumption Commitments : A Foundation for Reference-Dependent Preferences and Habit Formation," *Mimeo Harvard University*.

CHIAPPORI, P., I. MACHO, P. REY, AND B. SALANIÉ (1994): "Repeated Moral Hazard: The Role of Memory, Commitment, and the Access to Credit Markets," *European Economic Review, Elsevier, October, Vol*, 38, 1527–1553.

CHIAPPORI, P. A. A. AND B. SALANIÉ (2000): "Testing Contract Theory: a Survey of Some Recent Work," *Invited Lecture World Congress of the Econometric Society Seattle*, 35, 1–39.

DAIDO, K. AND H. ITOH (2007): "The Pygmalion and Galatea Effects : An Agency Model with Reference-Dependent Preferences and Applications to Self-Fulfilling Prophecy," *Working Paper*.

DE MEZA, D. AND D. WEBB (2007): "Incentive Design under Loss Aversion," *Journal of the European Economic Association, MIT Press*, 5, 66–92.

DELLA VIGNA, S. (2009): "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47, 315–372.

DELLAVIGNA, S., A. LINDNER, B. REIZER, AND J. F. SCHMIEDER (2015): "Reference-Dependent Job Search: Evidence from Hungary," *mimeo UC Berkeley*.

DICKENS, W. T., L. GOETTE, E. L. GROSHEN, S. HOLDEN, J. MESSINA, M. E. SCHWEITZER, J. TURUNEN, AND M. E. WARD (2007): "How Wages Change: Micro Evidence from the International Wage Flexibility Project," *Journal of Economic Perspectives*, 21, 195–214.

DITTMANN, I., E. MAUG, AND O. SPALT (2010): "Sticks or Carrots? Optimal CEO Compensation when Managers are Loss Averse," *Journal of Finance*, 65, 2015–2050.

ENGELMANN, D. AND G. HOLLARD (2010): "Reconsidering the Effect of Market Experience on the "Endowment Effect"," *Econometrica*, 78, 2005–2019.

FARBER, H. (2008): "Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers," *American Economic Review*, 98, 1069–1082.

FEHR, E. AND L. GOETTE (2005): "Robustness and real consequences of nominal wage rigidity," *Journal of Monetary Economics*, 52, 779–804.

——— (2007): "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment," *The American Economic Review*, 97, 298–317.

FORTIN, P. (1996): "The Great Canadian Slump," *The Canadian Journal of Economics / Revue canadienne d'Economique*, 29, 761–787.

FREUND, C. AND C. ÖZDEN (2008): "Trade Policy and Loss Aversion," *The American Economic Review*, 98, 1675–1691.

FUDENBERG, D. AND J. TIROLE (1990): "Moral hazard and renegotiation in agency contracts," *Econometrica*, 1279–1319.

GOETTE, L., D. HUFFMAN, AND E. FEHR (2004): "Loss Aversion and Labour Supply," *Journal of the European Economic Association*, 2, 216–228.

GORMAN, W. M. (1967): "Tastes, Habits and Choices," *International Economic Review*, 8, 218–222.

GROSSMAN, S. J. AND O. D. HART (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51, 7–45.

GUL, F. (1991): "A Theory of Disappointment Aversion," *Econometrica*, 59, 667–686.

HEALY, P. M. (1985): "The effect of bonus schemes on accounting decisions," *Journal of Accounting and Economics*, 7, 85–107.

HEFFETZ, O. AND J. A. LIST (2014): "It the Endowment Effect an Expectations Effect?" *Journal of the European Economic Association*, 12, 1396–1422.

HEIDHUES, P. AND B. KŐSZEGI (2008): "Competition and Price Variation When Consumers Are Loss Averse," *The American Economic Review*, 98, 1245–1268.

HERWEG, F., D. MÜLLER, AND P. WEINSCHENK (2010): "Binary Payment Schemes: Moral Hazard and Loss Aversion," *American Economic Review*, 100, 2451–2477.

IANTCHEV, E. P. (2011): "Risk or Loss Aversion? Evidence from Personnel Records," *Unpublished manuscript*.

JEWITT, I., O. KADAN, AND J. M. SWINKELS (2008): "Moral hazard with bounded payments," *Journal of Economic Theory*, 143, 59–82.

JOSEPH, K. AND M. U. KALWANI (1998): "The Role of Bonus Pay in Salesforce Compensation Plans," *Industrial Marketing Management*, 27, 147–159.

KAHNEMAN, D., J. KNETSCH, R. THALER, AND A. TVERSKY (1991): "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias," *The Journal of Economic Perspectives*, 5, 193–206.

KAHNEMAN, D. AND A. TVERSKY (1979): "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 263–292.

KŐSZEGI, B. (2014): "Behavioral Contract Theory," *Journal of Economic Literature*, 52, 1075–1118.

KŐSZEGI, B. AND M. RABIN (2006): "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 121, 1133–1165.

——— (2007): "Reference-Dependent Risk Attitudes," *American Economic Review*, 97, 1047–1073.

——— (2009): "Reference-Dependent Consumption Plans," *American Economic Review*, 93, 909–936.

KNETSCH, J. L. (1989): "The Endowment Effect and Evidence of Nonreversible Indifference Curves," *The American Economic Review*, 79, 1277–1284.

LEBOW, D., R. SAKS, AND B. A. WILSON (2003): "Downward Nominal Wage Rigidity: Evidence from the Employment Cost Index," *Advances in Macroeconomics*, 3, 1117.

LIST, J. A. (2004): "Neoclassical Theory versus Prospect Theory: Evidence from the Markeplace," *Econometrica*, 72, 615–625.

MACERA, R. (2013): "Intertemporal Incentives with Expectation-Based Reference-Dependent Preferences," *Mimeo, University of California at Berkeley*.

MUNRO, A. AND R. SUGDEN (2003): "On the Theory of Reference Dependent Preferences," *Journal of Economic Behavior and Organization*, 50, 407–428.

OYER, P. (1998): "Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality," *The Quarterly Journal of Economics*, 113, 149–185.

OZDAGLAR, A. E., D. P. BERTSEKAS, AND A. NEDIC (2003): *Convex Analysis and Optimization*, vol. 3 of *Athena Scientific Optimization and Computation Series*, Athena Scientific.

POLLAK, R. A. (1970): "Habit Formation and Dynamic Demand Functions," *Journal of Political Economy*, 78, 745–763.

PRENDERGAST, C. (1999): "The Provision of Incentives in Firms," *Journal of Economic Literature*, 37, 7–63.

RABIN, M. (1998): "Psychology and Economics," *Journal of Economic Literature*, 36, 11–46.

RAYO, L. AND G. S. BECKER (2007): "Evolutionary Efficiency and Happiness," *Journal of Political Economy*, 115, 302–337.

ROCKAFELLAR, R. T. (1970): *Convex Analysis*, vol. 28 of *Princeton Mathematics Series*, Princeton University Press.

——— (1974): *Conjugate Duality and Optimization*, SIAM.

ROCKAFELLAR, R. T. AND R. J.-B. WETS (2009): *Variational analysis*, vol. 317, Springer Science & Business Media.

ROGERSON, W. P. (1985): "Repeated Moral Hazard," *Econometrica*, 53, 69–76.

SALANIÉ, B. (2003): "Testing Contract Theory," *CESifo Economic Studies*, 49, 461–477.

SAMUELSON, W. AND R. ZECKHAUSER (1988): "Status Quo Bias in Decision Making," *Journal of Risk and Uncertainty*, 1, 7–59.

SEGAL, U. AND A. SPIVAK (1997): "First-Order Risk Aversion and Non-Differentiability," *Economic Theory*, 9, 179–183.

STEENBURGH, T. (2008): "Effort or timing: The effect of lump-sum bonuses," *QME*, 6, 235–256.

WILSON, B. A. (1997): "Movements of wages over the business cycle: an intra-firm view," *Finance and Economics Discussion Series 1997-1. Board of Governors of the Federal Reserve System (U.S.).*