Econometrica, Vol. 72, No. 3 (May, 2004), 701–746

MORAL HAZARD CONTRACTING AND PRIVATE CREDIT MARKETS

BY IN-UCK PARK[1]

This paper studies the impact of credit markets on optimal contracting, when the agent's "interim preference" over upcoming contracts is private information because personal financial decisions affect it via the wealth effect. The main result is a severe loss of incentive provision: equilibrium contracts invariably cause the agent to shirk (i.e., exert minimal effort) if the agent's private financial decision precedes moral hazard contracting. The basic intuition is that committing on another private variable, other than the effort level, exposes the parties to further exploitation of efficient risk-sharing by relaxing the incentive constraint that was binding ex ante, unless the risk-sharing was fully efficient to begin with.

KEYWORDS: Moral hazard, endogenous adverse selection, wealth effect.

1. INTRODUCTION

THE ESSENCE of an optimal contract in a principal-agent relationship is achieving efficiency in risk-sharing while providing appropriate performance incentives. The wealth position of the agent is a relevant factor in designing an efficient contract, because the agent's attitude towards risk involved in the contracts is influenced by his wealth level. Typically, the agent engages in private financial activities in between renegotiations of contracts and/or in anticipation of upcoming contracts, e.g., for intertemporal consumption smoothing. This intrinsic link between an agent's financial and production activities creates a serious obstacle to effective contracting: the principal does not know about the agent's wealth level as that evolved endogenously, and consequently, she cannot know about the agent's interim preference on subsequent contracts.

In this paper, we study the impact of such an endogenous informational asymmetry upon contractual relationships. Note that the agent's interim preferences are no longer common knowledge if the principal does not observe the agent's savings. Our main finding is that the disintegration of this common knowledge almost completely undermines the contract's capacity to provide incentives for the agent. Specifically, we show that shirking (i.e., exerting minimal effort) is generically the only possible equilibrium outcome in an extended

[1]I am grateful to Dan Berkowitz, Andreas Blume, Kim-Sau Chung, John Conlon, Dean Corbae, Anthony Dukes, Esther Gal-Or, Srihari Govindan, Maija Halonen, Ian Jewitt, Narayana Kocherlakota, Andy McLennan, Jack Ochs, Tongwook Park, Bernard Salanie, Drew Saunders, and Jan Werner for helpful conversations at various stages, and to an editor and two anonymous referees for thorough and insightful comments that improved the paper a lot. Preliminary versions of this paper have been presented at the Econometric Society Summer Meetings in 1996, the Second ESRC Research Seminars in Game Theory in Kenilworth, UK, the Midwest Theory conference at Purdue University, Seoul National University, University of Essex, University of Keele, University of Minnesota, University of Pittsburgh, and the Federal Reserve Bank in Richmond, and I thank participants. I am responsible for any remaining errors.

moral hazard model in which the agent privately makes an initial consumption/savings decision prior to standard moral hazard contracting.[2]

To get the main intuition for this inefficiency result, first observe that for optimal intertemporal consumption smoothing, the agent chooses an initial consumption/savings level such that the marginal utility of initial consumption is equal to the expected marginal utility of future consumption. Since the latter depends on the prospects of future income, the optimal level of initial consumption varies according to the specific contract that the agent anticipates and the effort level he plans to exert, because these are the factors that determine future income prospects.

Ex ante efficiency requires that the incentive constraint binds for an incentive contract designed to induce a high effort level, i.e., this contract provides just enough incentive so that the agent is indifferent between exerting high and low effort for this contract. Given an ex ante efficient contract, if the agent privately controls a variable prior to choice of effort level, such as initial consumption, then the incentive constraint may no longer bind after the first private variable is committed, hence the contract fails to be efficient at the interim stage. Such a discrepancy of incentive contracts between ex ante and interim constraints is the basic source of the aforementioned inefficiency result.

In our specific context, by choosing an initial consumption (type) optimally for a particular contract-effort pair to be adopted in equilibrium, the agent fixes his wealth level and thereby, his risk attitude optimally for the uncertain prospects that this contract-effort pair defines. So he may no longer derive the same utility level from other contract-effort pairs that were equivalent ex ante because the chosen risk attitude is suboptimal for the prospects that these pairs define. Since this will be the case for each separate type that might arise along the equilibrium path (if the agent follows a mixed strategy), the incentive constraints will not bind at this stage for each contract for which the agent optimized the initial consumption. This means that any contract that the agent anticipated in equilibrium is not the optimal one for the principal to offer at the actual contracting stage, unless the contract was designed to induce the minimal effort so that the incentive constraint was not a concern to begin with.

In standard moral hazard models in which effort level is the only private choice of the agent (i.e., without initial consumption), the need for incentive provision disappears once the effort level is chosen. If the contract can be manipulated at this stage, further exploitation of risk-sharing becomes possible, which also has an effect of undermining the contract's capacity to provide incentives. But, unlike in our model, this does not erode the equilibria in which the agent follows random paths because, the effort being the only hidden choice, the incentive constraint remains binding after the hidden choice is

[2]To be fully precise, this is shown to be always the case if the agent randomizes over finitely many paths, and generically so if continuous randomization is allowed when the agent has monotone absolute risk aversion.

made. Our analysis suggests that an additional hidden choice in another dimension causes the agent's interim preference to vary in a more divergent manner (because this extra dimension affects the preference) so that the incentive constraints get relaxed more easily; hence the eventual impact on the equilibrium contract is more drastic. This is reminiscent of Rochet and Stole (2003) who recognize multi-dimensional dependence of the agent's preference on private information and the endogeneity of binding incentive constraints as the main sources for divergence from the single-dimensional cases in screening models.

This difference can be illustrated by comparing our result with that of Fudenberg and Tirole (1990) who examined the renegotiation-proof solutions of a standard moral hazard where renegotiation may take place after the effort is taken but before the outcome is realized, so the agent's preference is private information at the renegotiation stage. Pure strategy solutions always induce the minimal effort in their model because, if the principal were certain that the agent chose high effort given the incentive of an initial contract, she could then safely revise the contract for mutual benefit by shifting the risk on the agent (which is no longer needed) to herself. However, mixed strategy equilibria also exist in which the high effort is taken with positive probability, alleviating the inefficiency: the agent randomizes between exerting the high effort and then self-selecting an "incentive contract" from a menu, and exerting the low effort and self-selecting a "risk-free contract." The idea is that in this case, revising the incentive contract as before to attract the agent who exerted the high effort, would not be profitable for the principal after all because it would be adversely selected by the agent who exerted the low effort, too.

This logic works in their model because ex ante efficiency requires that the incentive contract be marginal (i.e., equivalent) to the risk-free contract for the agent conditional on the low effort, so the aforementioned adverse selection entails when the incentive contract is revised. But the logic does not carry over to our setting because the agent's optimal initial consumption levels differ for the two contracts (conditional on the same low effort level), so the incentive contract becomes strictly inferior to the risk-free contract once the agent chooses his initial consumption optimally for the latter. Hence, the risk-sharing of the incentive contract can be improved at this stage without being adversely selected by such an agent. This means that high effort cannot be induced even randomly in equilibrium.

The main intuition for our result explained above relies on the direct dependence of the agent's interim preference upon his initial consumption, which exists because the initial consumption determines his risk attitude via wealth level. In the knife-edge cases where the wealth effect does not exist (e.g., when the agent exhibits constant absolute risk aversion), we still obtain the same inefficiency due to another effect: if an agent's initial consumption is not optimal for the outside option, the interim value of the outside option is strictly lower for this agent than the ex ante value. Foreseeing that the principal will try to exploit such an agent by offering a contract short of the ex ante value (albeit worth the interim value) of the outside option, the agent would not pursue a path leading to high effort because such a path would require an initial consumption suboptimal for the outside option. Consequently, the aforementioned inefficiency is a general result that prevails in all cases, even when the wealth effect is absent.

We briefly review other studies dealing with endogenous hidden information on interim preferences in moral hazard contracting. In addition to Fudenberg and Tirole (1990) discussed above, Ma (1991) studies a model in which a single effort choice affects outcomes in two periods and renegotiation is possible after the first-period outcome. Ma does not model capital markets and finds that it can be optimal to induce high effort stochastically. Also related are Ma (1994) and Matthews (1995) who study the same basic issue in a one-shot model similar to Fudenberg and Tirole, but in which the informed party (the agent) initiates renegotiation. In these studies, type selection (effort level) influences the agent's interim preference only to the extent that it defines the exact risk associated with continuation contracts. The current paper differs in that the type selection (savings level) affects the interim utility function itself via the wealth effect.

As for the private financial decisions and the ensuing hidden information on interim preferences, Fudenberg, Holmstrom, and Milgrom (1990), inter alia, provide an example illustrating that the effects can be loss of incentive provision in multiperiod contracting environments with renegotiation. Chiappori et al. (1994) generalize this example and prove that the loss of incentive provision prevails in all pure strategy equilibria of such environments. In light of the findings of Fudenberg and Tirole, however, examination of mixed strategy equilibria appears essential for a better understanding of the topic.[3]

This paper also belongs to the literature on nonexclusive contracting. Most studies in this literature address issues that arise when the agent can make independent contracts with multiple principals for the same moral hazard activity (Arnott and Stiglitz (1991), Bisin and Guaitoli (2000), Bisin and Rampini (2002), Bizer and DeMarzo (1992, 1999), Kahn and Mookherjee (1998), and Pauly (1974)). The current paper addresses a different issue, namely, the impact of personal credit markets (that exist independently of the moral hazard problem) on moral hazard contracting.

The rest of the paper is organized as follows. Section 2 provides an illustration of the key argument of the paper. Section 3 describes the model. Section 4 discusses the equilibrium conditions. Sections 5 and 6 characterize the properties that equilibrium contracts should satisfy. Section 7 identifies the equilibrium paths with these properties and presents the main results. Section 8 contains a summary and some concluding remarks. A lengthy technical proof is relegated to the Appendix.

[3]Park (2001) contains some results showing that the loss of incentive provision generalizes beyond the pure strategy solutions.

2. SIMPLE ILLUSTRATION OF MAIN RESULT

This section attempts to elucidate the essential forces behind the main results informally. Consider a manager and a worker who will engage in a standard moral hazard contract *next* period: if the worker signs a contract offered by the manager, he will privately exert either a high or low effort, which determines the probabilities of good and bad outcomes/profits for the manager in a natural way. The next period's wage is the only income of the worker, so he seeks consumption smoothing by financing this period's consumption in the credit market against the anticipated income in the next period.

The low effort level can be induced most efficiently by a flat-wage scheme that pays a fixed sum w_f regardless of the outcome. If the worker signs a flat-wage scheme in equilibrium, therefore, he will exert the low effort and smooth consumption by an equal split between two periods, i.e., this period's consumption is $c^\ell = w_f/2$.

To induce the high effort level, on the other hand, a bonus/incentive scheme is needed that pays w_g if a good outcome is realized and w_b if a bad outcome is realized ($w_g > w_b$). If the worker adopts such a bonus scheme in equilibrium, he will exert the high effort and smooth consumption by equating marginal consumption across periods; i.e., this period's consumption, c^h, satisfies

$$u'(c^h) = pu'(w_g - c^h) + (1 - p)u'(w_b - c^h),$$

where u is a utility function of the risk-averse agent and p is the probability of a good outcome when high effort is exerted. The most efficient bonus scheme provides just a marginal incentive to induce the high effort, so that the worker may derive the same utility by exerting low effort as well, given the bonus scheme (i.e., the incentive compatibility constraint binds). Optimal consumption smoothing given low effort, however, means that he should consume less than c^h in this period because the equation above needs to be satisfied for a lower p, i.e., because the income prospects are lower.

From this we deduce that the manager cannot induce the high effort with certainty because it is time-inconsistent. In such an equilibrium, the worker would have spent c^h in the initial period, at which point he strictly prefers exerting the high effort for the bonus scheme because he already spent suboptimally for exerting the low effort. Knowing this, the manager would devise a new compensation scheme that Pareto-improves upon the supposed equilibrium bonus scheme, by reducing both the risk and expected wage for the agent in such a way that the risk-averse worker would prefer it to the original wage scheme and would still exert the high effort. Offering such a revised wage scheme would upset the supposed equilibrium. (Note that this offer would not induce the high effort if anticipated initially, because then the agent would find it optimal to consume differently and exert the low effort subsequently.)

Hence, if the high effort is induced at all, it would be induced stochastically. In such an equilibrium, the worker would mix between the two aforementioned paths, one in which he consumes c^h, then self-selects the bonus scheme from a menu and exerts the high effort, and the other in which he consumes c^ℓ, then self-selects the flat-wage scheme and exerts the low effort. Note here that replacing the bonus scheme in the menu with the one that Pareto-improves upon it as described in the previous paragraph would still work the same way to upset the equilibrium, as long as the worker who has spent c^ℓ either prefers the flat-wage scheme to the revised one or selects the latter and then exerts the high effort. (In the latter case, this selection would only reinforce the benefits the manager expects from attracting the worker who spent c^h.) A key result of this paper is that the wealth effect on the agent's preference ensures that this provision indeed holds, thereby upsetting the equilibrium.

To get an intuition, observe that the worker finds the flat-wage scheme and the original bonus scheme equivalent ex ante, but after spending c^ℓ he would weakly prefer the former (because c^ℓ is not necessarily optimal for the latter). If he strictly prefers the flat-wage scheme at this stage, by continuity he would opt for the flat-wage scheme when the bonus scheme is revised only slightly. If, on the other hand, he still finds the two schemes equivalent but he can derive the equilibrium utility from the bonus scheme only by exerting the high effort, then he will in fact exert the high effort in case he adopts the revised bonus scheme. In either case, the supposed equilibrium cannot be sustained because the revised scheme brings about Pareto-improvement as discussed above.

To part from this conclusion, therefore, after spending c^ℓ the worker must remain indifferent between the flat-wage scheme and the bonus scheme *conditional on* exerting the low effort level. Phrased slightly more generally, the following condition is necessary for a mixed equilibrium: The hidden action of the agent (the initial consumption, or equivalently, the wealth level at the point of contracting) should not result in relaxing the incentive constraint via affecting his preference over contracts. In our model this amounts to requiring that the wealth effect should not exist.[4] In the one-shot model of Fudenberg–Tirole, on the other hand, this condition is automatically satisfied because the hidden action is in fact the effort level, which is the core distinction from our multi-period consumption environment.

3. MODEL

Consider a principal (she) and an agent (he) who engage in independent moral hazard contracting for two periods. If income is not transferable between periods, backward induction would imply repetition of the optimal single-period contract. If the agent can freely access the credit market, on the other

[4]Even when the wealth effect is absent, the path of consuming c^h and then exerting high effort is not viable in our setting due to another reason: since the outside option is strictly inferior to the equilibrium path for the agent who consumed c^h (see Lemma 5), the principal can shade the wage scheme for this agent and still induce the high effort, upsetting the supposed equilibrium.

hand, his inter-period financial decisions have implications on subsequent contracting. Such effects can be fully examined by focusing on the continuation game from the point in time that the first period compensation is made but before the consumption/savings decision, because the only strategic link between the two periods is the intertemporal consumption smoothing. For expositional clarity, therefore, we analyze a single-period moral hazard model augmented by an initial stage in which the agent makes a private consumption/savings decision in anticipation of an upcoming contract. The main results in this paper extend qualitatively to multiperiod models where the parties sign a spot contract in each period.

The moral hazard activity is characterized by two *actions* or *effort levels*, the high level h and the low level ℓ, that the agent may take, and two numerical *outcomes,* g (the "good outcome") and b (the "bad outcome"), where $g > b > 0$. The outcomes g and b occur with probabilities p^e and $1 - p^e$, respectively, where $e \in \{h, \ell\}$ is the effort level taken by the agent. We assume $1 > p^h > p^\ell > 0$. A *wage scheme* is a vector $\mathbf{w} = (w_g, w_b) \in \mathbb{R}^2$, where w_g and w_b are *wages* contingent upon the outcomes g and b, respectively.

The order of events is as follows. In the initial period $t = 0$, the agent with an initial wealth w_0 privately consumes $c \geq 0$ and saves the remaining wealth (or borrows the shortage against future income) at zero interest rate. This determines his interim wealth level at the beginning of the main, subsequent period $t = 1$. In this main period, (i) the principal offers a menu, W, of wage schemes, which the agent either accepts or rejects; (ii) if the agent accepts the offer, he selects one wage scheme $\mathbf{w} = (w_g, w_b)$ from the menu W as a binding contract, followed by the usual moral hazard activities: he privately takes an effort level $e \in \{h, \ell\}$, an outcome $x \in \{g, b\}$ is realized, and the principal pays a wage w_x to the agent and retains a *profit* of $x - w_x$; (ii′) if the agent rejects the offer the agent receives an outside option which is a fixed income $\underline{w} > 0$, and the principal's profit is 0; (iii) the agent privately consumes the remaining wealth after paying off any debt from the previous period.

The agent's payoff is measured by a *utility function* $u : \mathbb{R} \to \mathbb{R}$ from consumption and a *disutility function* $d : \{h, \ell\} \to \mathbb{R}$ from effort, which are additively separable. The total payoff of the agent is the undiscounted[5] sum of the two periods' payoffs, i.e., $u(c_0) + u(c_1) - d(e)$, where c_t is the consumption level in period $t = 0, 1$, and e is the effort taken. We make the standard assumptions that u is twice continuously differentiable, increasing, and strictly concave ($u' > 0$ and $u'' < 0$), and $d(h) > d(\ell) = 0$. The principal maximizes the expected profit. The payoff structure reflects our presumption that the agent is risk-averse in consumption and the principal is risk-neutral.

Since the agent's consumption (equivalently, saving) is private, the contract cannot be contingent on the wealth level of the agent. Furthermore, if the agent

[5]The main results of the paper stand for any interest rate and discount rate, as long as both players can borrow and save with unlimited liability.

follows a random consumption path, hidden information on the wealth level of the agent, referred to as his *type*, arises at the point when contracts are being offered. To deal with this problem we model that the principal offers a (*wage*) *menu*, a set of wage schemes, so that the agent may choose a wage scheme from the menu depending on his type.

This is an endogenous adverse selection phenomenon that arises as a part of the agent's optimization behavior, and is different in nature from the more conventional adverse selection exogenously given from outside the model. To focus on the former, we assume away any exogenous uncertainty: the agent's initial wealth is known and is normalized as $w_0 = 0$. The extensive form game described above, denoted by Γ, is common knowledge.

Since the agent finances initial consumption by borrowing before any wage contract is actually signed, there is an issue of potential default. This does not pose a real problem in our analysis, though, because the credit markets would take the agent's incentives into account in equilibrium, and hence would not advance loans that will result in default. To facilitate exposition, however, we abstract from this issue by imposing unlimited liability.[6]

4. EQUILIBRIUM PATHS

The solution concept we use is perfect Bayesian equilibrium, which requires that the strategies be sequentially rational given the players' beliefs, and that the beliefs be obtained whenever possible from equilibrium strategies and observations using Bayes rule. To avoid the uninteresting solution of no contracting in Γ, we assume that the principal strictly prefers employing the agent by paying $\underline{w}$ (the income of an outside option) to no employment:

$$p^{\ell} g + (1 - p^{\ell}) b > \underline{w}. \tag{1}$$

In equilibrium the Participation Constraint is binding, so the agent is indifferent between working for the principal and taking the outside option. We make a standard tie-breaking assumption that the agent chooses to work for the principal in these cases, to avoid inessential analytical complications. Then, the agent signs a contract with probability 1 in equilibrium.

The strategic decisions the agent takes in Γ are depicted by a tuple $(c, \mathbf{w}, e)$, called a *passage*, that specifies an initial consumption level c, and the wage scheme $\mathbf{w}$, and effort level e to adopt. A *passage profile* is a set, $\mathcal{P}$, of passages. A *passage distribution* is a probability measure, F, whose support is a passage profile $\mathcal{P}$. (We say that F is *on* $\mathcal{P}$.) The agent's strategy in an equilibrium induces a passage distribution. The main task in the remainder of the paper is to characterize the passage distributions that can be induced in an equilibrium.

[6]Alternatively, we may assume that $\lim_{c \to 0} u(c) = -\infty$ so that the agent has no incentive to borrow more than he can pay back.

This section formalizes the properties that such distributions must satisfy and the subsequent sections identify the distributions that satisfy them.

In an equilibrium the agent chooses a passage to maximize the (expected) utility, i.e., to solve

$$\max U(c, \mathbf{w}, e) = u(c) + p^e u(w_g - c) + (1 - p^e)u(w_b - c) - d(e) \tag{2}$$

subject to choosing $\mathbf{w}$ from the equilibrium wage menu W and $e \in \{h, \ell\}$. Since W is arbitrary at this point, different passages the agent may adopt in an equilibrium are characterized as solving (2) subject to $\mathbf{w} \in W$ for some common wage menu W. Equivalently stated, an equilibrium induces an ex ante rational profile $\mathcal{P}$ as defined below.

DEFINITION 1: A passage profile $\mathcal{P}$ is *ex ante rational* if every passage in $\mathcal{P}$ solves (2) subject to $\mathbf{w} \in W(\mathcal{P})$ and $e \in \{h, \ell\}$, where $W(\mathcal{P})$ is the wage menu consisting of all the wage schemes in $\mathcal{P}$.

At the point in time when the principal offers a wage menu, which we refer to as the interim stage, the agent has already committed the initial consumption: this fixes his risk attitude, hence his preference over wage contracts. An equilibrium requires that the wage menu be optimal for the principal to offer given such interim preferences of the agent. To formalize this condition in terms of passage profile and distribution, we introduce some terminology.

The consumption component c of a passage $(c, \mathbf{w}, e)$ is private information at the interim stage and is called the agent's *type.* Given a passage distribution F, the *induced type distribution*, denoted by F_C, is the marginal distribution of F on types. Two passage distributions are *type-preserving variations* of each other if their induced type distributions are identical.

The agent's response to a wage menu varies depending on his type, and for some wage menus the outside option may be strictly preferred. So, given an arbitrary wage menu W, a type c agent selects a wage scheme-action pair $(\mathbf{w}, e)$ to solve (2) subject to $(\mathbf{w}, e) \in (W \times \{h, \ell\}) \cup \{(\underline{\mathbf{w}}, \phi)\}$, conditional on c, where $(\underline{\mathbf{w}}, \phi)$ denotes choosing the outside option represented by the flat wage scheme $\underline{\mathbf{w}} = (\underline{w}, \underline{w})$.

An *outside passage* refers to a tuple $(c, \underline{\mathbf{w}}, \phi)$; an *extended* passage profile refers to a set of passages possibly including outside passages; and an *extended* passage distribution is a probability measure whose support is an extended passage profile. The principal's expected profit given an extended passage distribution F is

$$\Pi(F) = \int_{\{(c,\mathbf{w},e) | e \neq \phi\}} \big(p^e(g - w_g) + (1 - p^e)(b - w_b)\big)\, dF. \tag{3}$$

Note that the integration is over nonoutside passages. We extend the definitions of *induced type distribution* and *type-preserving variations* to extended passage distributions in the obvious way.

DEFINITION 2: A passage distribution F on a passage profile $\mathcal{P}$ is *interim rational* if for every wage menu $\tilde{W}$ there exists an extended passage distribution $\tilde{F}$ on an extended passage profile $\tilde{\mathcal{P}}$ such that:

(i) if $(c, \mathbf{w}', e') \in \tilde{\mathcal{P}}$, then $(\mathbf{w}', e')$ solves (2) subject to $(\mathbf{w}, e) \in (\tilde{W} \times \{h, \ell\}) \cup \{(\underline{\mathbf{w}}, \phi)\}$, conditional on c,

(ii) $\tilde{F}$ is a type-preserving variation of F, and

(iii) $\Pi(\tilde{F}) \leq \Pi(F)$.

The interim rationality formalizes the principal's optimality that she cannot increase her expected profit by offering a wage menu other than $W(\mathcal{P})$ at this interim stage, provided that the agent (of various equilibrium types) responds optimally. Combining with ex ante rationality, we now define an equilibrium of Γ in terms of a passage distribution.[7]

DEFINITION 3: A passage distribution F on a passage profile $\mathcal{P}$ is an *equilibrium* of Γ if $\mathcal{P}$ is ex ante rational and F is interim rational. An ex ante rational passage profile $\mathcal{P}$ is an equilibrium passage profile if there exists a passage distribution F on $\mathcal{P}$ that is interim rational, which we refer to as a *supporting* distribution.

5. EX ANTE RATIONALITY

We say that a passage $(c, \mathbf{w}, e)$ is *self-optimal* if (c, e) is optimal for $\mathbf{w}$ (i.e., solves (2) given $\mathbf{w}$): in this case we also say that $\mathbf{w}$ *prompts* the type c, the type-effort pair (c, e), and the passage $(c, \mathbf{w}, e)$. It is immediate from the definition that the passages in an ex ante rational passage profile $\mathcal{P}$ are self-optimal and have the same (expected) utility level, say v^*. Conversely, any passage profile that consists of self-optimal passages with the same utility level is ex ante rational. Let $\mathcal{P}(v^*)$ denote an ex ante rational profile with a common utility level v^*. As the pool of all possible members of $\mathcal{P}(v^*)$, we construct the set of all self-optimal passages with the utility level v^*. First, we find the wage schemes that generate v^* given $e = \ell$, and given $e = h$, separately. Next, among these we select the ones for which the given e is indeed optimal.

Fix $e \in \{\ell, h\}$ and choose c to solve (2) for a given $\mathbf{w}$. Since the objective function is strictly concave in c, there is a unique solution

$$c^*(\mathbf{w}, e) = \arg\max_c U(\cdot, \mathbf{w}, e) \tag{4}$$

[7]Implicit in this definition is that the principal's wage menu offer is not random in equilibrium. If the principal were to make offers randomly, then it would be consistent to model that she may offer random devices (i.e., probabilities over wage menus) according to which a wage menu will be selected if such an offer was accepted by the agent. This would complicate the model and analysis with a much larger strategy space of the principal. An earlier version of this paper (Park (2001)) contains an analysis that shows that nontrivial random offers are not used in equilibrium if the agent's utility function u exhibits strictly monotone absolute risk aversion.

that satisfies the first-order condition:

$$D_c U(c, \mathbf{w}, e) = u'(c) - p^e u'(w_g - c) - (1 - p^e)u'(w_b - c) = 0. \tag{5}$$

Denote the value function by

$$V(\mathbf{w}, e) = U(c^*(\mathbf{w}, e), \mathbf{w}, e). \tag{6}$$

By the Envelope theorem, the gradient is

$$D_{\mathbf{w}} V(\mathbf{w}, e) = \begin{pmatrix} p^e u'(w_g - c^*) \\ (1 - p^e)u'(w_b - c^*) \end{pmatrix}. \tag{7}$$

Denote the set of $\mathbf{w}$ with the same value, say v, by

$$\mathrm{IC}^e(v) = \{\mathbf{w} \in \mathbb{R}^2 : V(\mathbf{w}, e) = v\}. \tag{8}$$

The graph of $\mathrm{IC}^e(v)$ is called an ex ante indifference curve (conditional on e): it consists of wage schemes that are ex ante equivalent. Obviously, $V(\mathbf{w}, e)$ is increasing in $\mathbf{w}$. Since u is strictly concave, $V(\mathbf{w}, e)$ is also strictly concave in $\mathbf{w} \in \mathbb{R}^2$.[8] Therefore, the ex ante indifference curves in the space of wage schemes ($\mathbb{R}^2$) are strictly concave to the origin.

The ex ante indifference curves exhibit the usual properties of the static principal-agent models. In particular, the "single-crossing property" is satisfied as the next lemma asserts, facilitating the subsequent analysis considerably. Unlike in the static model, however, this result is not straightforward in the considered environment, because in principle the intertemporal consumption smoothing can have quite complex influence on the (ex ante) preference over wage schemes. The next lemma shows that this effect is not strong enough to overturn the single-crossing property in the current model. (Strict concavity of u is used to prove this.) Accordingly, the two indifference curves $\mathrm{IC}^\ell(v^*)$ and $\mathrm{IC}^h(v^*)$ for a utility level v^*, illustrated in Figure 1, cross at a single point, denoted by $\mathbf{w}^h = (w_g^h, w_b^h)$, at which $\mathrm{IC}^h(v^*)$ is steeper than $\mathrm{IC}^\ell(v^*)$.

LEMMA 1: *Any two ex ante indifference curves* $\mathrm{IC}^h(v)$ *and* $\mathrm{IC}^\ell(v')$ *meet at most at a single point and at this point* $\mathrm{IC}^h(v)$ *is strictly steeper than* $\mathrm{IC}^\ell(v')$.

PROOF: For this proof, we treat p^e as a continuous variable and denote it by p. Then, c^* and $D_{\mathbf{w}}V$ in (7) are functions of p (instead of e). Differentiate (5) with respect to p to get

$$\frac{\partial c^*}{\partial p} = \frac{u'(w_g - c^*) - u'(w_b - c^*)}{u''(c^*) + pu''(w_g - c^*) + (1 - p)u''(w_b - c^*)}. \tag{9}$$

[8]It is straightforward to verify that the Hessian matrix of $V(\cdot, e)$ is negative definite for all $\mathbf{w} \in \mathbb{R}^2$, by checking the signs of the principal minors.

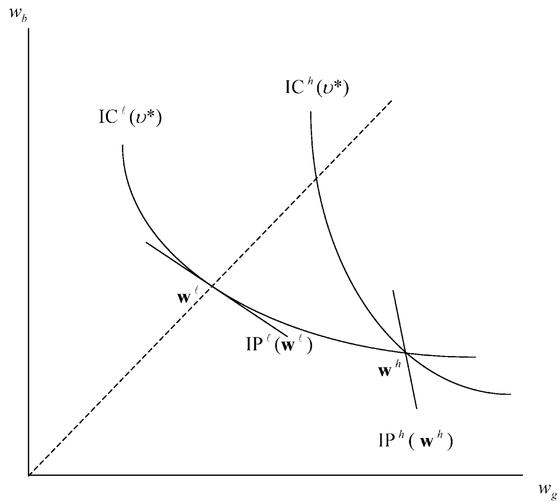

FIGURE 1.

We prove the claim by showing that

$$\frac{\partial}{\partial p}\left(\frac{(1-p)u'(w_b-c^*)}{pu'(w_g-c^*)}\right),$$

the slope of the gradient $D_{\mathbf{w}}V$ defined in (7), decreases in p. Note

$$\begin{aligned}&\frac{\partial}{\partial p}\left(\frac{(1-p)u'(w_b-c^*)}{pu'(w_g-c^*)}\right)p^2u'(w_g-c^*)^2\\&\quad=-u'(w_b-c^*)u'(w_g-c^*)\\&\qquad+p(1-p)\big(u'(w_b-c^*)u''(w_g-c^*)\\&\qquad\qquad-u'(w_g-c^*)u''(w_b-c^*)\big)\frac{\partial c^*}{\partial p}.\end{aligned}$$

After plugging in (9), multiply by the negative of the denominator of (9) and rearrange to get

$$\begin{aligned}&u'(w_b-c^*)u'(w_g-c^*)\big(u''(c^*)+p^2u''(w_g-c^*)\\&\qquad\qquad+(1-p)^2u''(w_b-c^*)\big)\\&\quad+p(1-p)\big(u'(w_g-c^*)^2u''(w_b-c^*)+u'(w_b-c^*)^2u''(w_g-c^*)\big)\\&\quad<0.\end{aligned}$$

This proves

$$\frac{\partial}{\partial p}\left(\frac{(1-p)u'(w_b-c^*)}{pu'(w_g-c^*)}\right)<0,$$

i.e., the slope of $D_{\mathbf{w}}V$ decreases in p. This implies that $\mathrm{IC}^h(v)$ is steeper than $\mathrm{IC}^\ell(v')$ at every point that they meet. Since both of these curves are connected one-dimensional manifolds, it follows that they cannot meet more than once. *Q.E.D.*

For wage schemes on $\mathrm{IC}^h(v^*)$ that lie above $\mathrm{IC}^\ell(v^*)$, the agent's optimal effort is obviously ℓ and he derives a utility level higher than v^*. So these wage schemes are not used in ex ante rational passage profiles $\mathcal{P}(v^*)$. Analogously, wage schemes on $\mathrm{IC}^\ell(v^*)$ that lie above $\mathrm{IC}^h(v^*)$ are not used in $\mathcal{P}(v^*)$, either. Hence, the wage schemes that can be used in $\mathcal{P}(v^*)$ constitute the lower envelope of $\mathrm{IC}^\ell(v^*)$ and $\mathrm{IC}^h(v^*)$, which we call the *indifference frontier* and denote by $\mathrm{IC}(v^*)$. The decomposition

$$\mathrm{IC}(v^*)=\big(\mathrm{IC}^\ell(v^*)\cap\mathrm{IC}(v^*)\big)\cup\big(\mathrm{IC}^h(v^*)\cap\mathrm{IC}(v^*)\big) \tag{10}$$

is trivial. An ex ante rational passage profile $\mathcal{P}(v^*)$ consists of self-optimal passages $(c,\mathbf{w},e)$ with $\mathbf{w}\in\mathrm{IC}(v^*)$. Conversely, any collection of such passages forms an ex ante rational passage profile with a utility level v^*. The next two lemmas are immediate from the fact that the optimal consumption $c^*(\mathbf{w},e)$ is unique given $(\mathbf{w},e)$ and is higher when $e=h$ than when $e=\ell$ (for the same $\mathbf{w}$); hence the proofs are omitted.

LEMMA 2: (i) *The wage scheme* $\mathbf{w}^h=\mathrm{IC}^\ell(v^*)\cap\mathrm{IC}^h(v^*)$ *prompts exactly two self-optimal passages, namely,* $(c^*(\mathbf{w}^h,h),\mathbf{w}^h,h)$ *and* $(c^*(\mathbf{w}^h,\ell),\mathbf{w}^h,\ell)$.
(ii) *Every other wage scheme in* $\mathrm{IC}(v^*)$ *prompts a unique self-optimal passage.*

LEMMA 3: *If* $(c,\mathbf{w},e)$ *is a self-optimal passage, then a c-type agent strictly prefers* e *to the other effort level* $e'=\{h,\ell\}\setminus\{e\}$ *given* $\mathbf{w}$, *i.e.,* $U(c,\mathbf{w},e)>U(c,\mathbf{w},e')$.

Next, isoprofit lines are added in Figure 1. The expected profit of the principal from $(\mathbf{w},e)$ is

$$\Pi(\mathbf{w},e)=p^e(g-w_g)+(1-p^e)(b-w_b) \tag{11}$$

and its gradient is

$$D_{\mathbf{w}}\Pi(\mathbf{w},e)=\begin{pmatrix}-p^e\\-1+p^e\end{pmatrix}. \tag{12}$$

Conditional on e, an isoprofit line containing $\mathbf{w}$ is a straight line perpendicular to the gradient (12), and has an associated expected profit level of $\Pi(\mathbf{w}, e)$. It is expositionally convenient to define isoprofit lines relative to wage schemes rather than profit levels:

$$\text{(13)} \qquad \text{IP}^e(\mathbf{w}) = \{\mathbf{w}' \in \mathbb{R}^2 : \Pi(\mathbf{w}', e) = \Pi(\mathbf{w}, e)\}.$$

Two isoprofit lines, $\text{IP}^\ell(\mathbf{w}^\ell)$ and $\text{IP}^h(\mathbf{w}^h)$, are illustrated in Figure 1 for two important wage schemes in subsequent analysis: $\mathbf{w}^h = \text{IC}^\ell(v^*) \cap \text{IC}^h(v^*)$ as defined earlier and $\mathbf{w}^\ell$, the intersection of $\text{IC}^\ell(v^*)$ and the certainty line. (Although not indicated for notational simplicity, $\mathbf{w}^\ell$ and $\mathbf{w}^h$ are defined for an underlying utility level v^*.) Because the gradients (7) and (12) are linearly dependent when $w_g = w_b$, isoprofit lines are tangent to ex ante indifference curves along the certainty line (conditional on the same e). Hence, $\mathbf{w}^\ell$ and $\mathbf{w}^h$ are the wage schemes that generate the maximum expected profit for the principal subject to the reservation utility level v^* and induce ℓ and h, respectively. The two passages that $\mathbf{w}^\ell$ and $\mathbf{w}^h$ prompt and generate the maximum expected profit are important in subsequent discussion, so we denote them by $(c^\ell, \mathbf{w}^\ell, \ell)$ and $(c^h, \mathbf{w}^h, h)$ as shorthand, i.e., $c^\ell = c^*(\mathbf{w}^\ell, \ell)$ and $c^h = c^*(\mathbf{w}^h, h)$.

6. INTERIM RATIONALITY

The previous section characterizes an ex ante rational passage profile as a collection of self-optimal passages whose wage schemes are on the same indifference frontier $\text{IC}(v^*)$. For such a passage profile $\mathcal{P}$ to indeed be an equilibrium, a supporting passage distribution should exist that is interim rational so that the principal cannot benefit by offering a wage menu different from $W(\mathcal{P})$, the set of wage schemes in $\mathcal{P}$, at the interim stage. In the next section, we exclude various ex ante rational passage profiles by showing that alternative wage menus exist that, if offered at the interim stage, would necessarily increase the principal's expected profit. This section provides two preliminary observations that are useful in this process.

DEFINITION 4: A wage scheme $\hat{\mathbf{w}}$ *Pareto-improves upon* a passage $(c, \mathbf{w}, e)$ if both the principal and a c-type agent are better off when the agent adopts $\hat{\mathbf{w}}$ and selects the effort level optimally, i.e., if $U(c, \hat{\mathbf{w}}, \hat{e}) > U(c, \mathbf{w}, e)$ and $\Pi(\hat{\mathbf{w}}, \hat{e}) > \Pi(\mathbf{w}, e)$, where $\hat{e}$ (which may or may not be e) is the effort level strictly preferred[9] by a c-type agent given $\hat{\mathbf{w}}$. A self-optimal passage $(c, \mathbf{w}, e)$ prompted by $\mathbf{w} \in \text{IC}(v^*)$ is *ex post inefficient* if there exists a wage scheme $\hat{\mathbf{w}}$ that Pareto-improves upon $(c, \mathbf{w}, e)$. A passage is *ex post efficient* otherwise.

[9]That $\hat{e}$ is uniquely optimal for $\hat{\mathbf{w}}$ is innocuous: if both effort levels are optimal conditional on c, $\hat{\mathbf{w}}$ can be perturbed slightly to break the tie in either direction.

The first observation, formalized in the next lemma, is that all non-flat wage schemes $\mathbf{w}$ (i.e., $w_g \neq w_b$) necessarily prompt ex post inefficient passages because, once the type of a self-optimal passage $(c, \mathbf{w}, e)$ is committed, a small improvement of insurance (which is feasible for every non-flat $\mathbf{w}$) would induce the same effort level due to Lemma 3, resulting in a Pareto improvement.

LEMMA 4: *Every self-optimal passage* $(c, \mathbf{w}, e)$ *with* $\mathbf{w} \in \mathrm{IC}(v^*) \setminus \{\mathbf{w}^\ell\}$ *is ex post inefficient. Furthermore, one can find wage schemes arbitrarily close to* $\mathbf{w}$ *that Pareto-improve upon* $(c, \mathbf{w}, e)$.

PROOF: Consider a passage $(c', \mathbf{w}', e')$ prompted by $\mathbf{w}' \in \mathrm{IC}(v^*) \setminus \{\mathbf{w}^\ell\}$. We prove for the case in which $\mathbf{w}'$ lies below the certainty line, but the same argument applies to the other case. It is graphically clear that the ex ante indifference curve $\mathrm{IC}^{e'}(v^*)$ and the isoprofit line $\mathrm{IP}^{e'}(\mathbf{w}')$ cross at $\mathbf{w}'$ where the former has a flatter slope than the latter. That is, the gradient $D_\mathbf{w} V(\mathbf{w}, e')$ in (7) is steeper than the negative of the gradient $D_\mathbf{w} \Pi(\mathbf{w}, e')$ of (12) at $\mathbf{w} = \mathbf{w}'$. Therefore, there exists a vector $\mathbf{r} \in \mathbb{R}_- \times \mathbb{R}_+$, arbitrarily small such that $\mathbf{r} \cdot D_\mathbf{w} \Pi(\mathbf{w}, e') > 0$ and $\mathbf{r} \cdot D_\mathbf{w} V(\mathbf{w}, e') > 0$ when these products are evaluated at $\mathbf{w} = \mathbf{w}'$. Note that, by the Envelope theorem, the gradients $D_\mathbf{w} U(c', \mathbf{w}, e')$ and $D_\mathbf{w} V(\mathbf{w}, e')$ coincide at $\mathbf{w} = \mathbf{w}'$. Hence,

$$(14) \qquad U(c', \mathbf{w}' + \mathbf{r}, e') > v^* \quad \text{and} \quad \Pi(\mathbf{w}' + \mathbf{r}, e') > \Pi(\mathbf{w}', e').$$

Since $U(c', \mathbf{w}', e') = v^* > U(c', \mathbf{w}', e)$ for $e \neq e'$ by Lemma 3, e' is uniquely optimal for the c'-type given $\mathbf{w}' + \mathbf{r}$ for sufficiently small $\mathbf{r}$. Hence, $\mathbf{w}' + \mathbf{r}$ Pareto-improves upon $(c', \mathbf{w}', e')$ for arbitrarily small $\mathbf{r}$, completing the proof. *Q.E.D.*

The second observation concerns the changing value of the outside option. The ex ante value of the outside option is $2u(\underline{w}/2)$ because equal split of consumption is optimal for the outside option. Therefore, the agent's equilibrium utility level is at least $2u(\underline{w}/2)$. As the next lemma states, however, the value of the outside option at the interim stage is strictly lower unless the agent has consumed $\underline{w}/2$ initially, i.e., optimally for the outside option. Consequently, the principal may try to exploit the deteriorated outside option by offering "shaded" wage schemes for the agent of types $c \neq \underline{w}/2$, because the agent would still accept them as long as they are better than the interim value of the outside option. Recall $c^\ell = c^*(\mathbf{w}^\ell, \ell)$.

LEMMA 5: *Let* $(c, \mathbf{w}, e)$ *be a passage in an equilibrium passage profile* $\mathcal{P}(v^*)$. *If* $c \neq c^\ell$, *then the interim value of the outside option for a c-type agent*, $u(c) + u(\underline{w} - c)$, *is strictly less than the equilibrium utility level* $v^* = U(c, \mathbf{w}, e)$.

PROOF: As asserted earlier, $v^* \geq 2u(\underline{w}/2)$, and $2u(\underline{w}/2) \geq u(c) + u(\underline{w} - c)$. If $v^* > 2u(\underline{w}/2)$, therefore, the conclusion is trivial. If $v^* = 2u(\underline{w}/2)$, then

$(\underline{w}/2, (\underline{w}, \underline{w}), \ell)$ is a self-optimal passage in $\mathcal{P}(v^*)$ with a flat wage scheme, so $c^\ell = \underline{w}/2$. Since $u(c)+u(\underline{w}-c) < 2u(\underline{w}/2)$ for any $c \neq \underline{w}/2$ by strict concavity of u, the proof is complete. *Q.E.D.*

7. INEFFICIENCY

Suppose a single passage $(c, \mathbf{w}, e)$ is to be taken with certainty in an equilibrium. Then, this passage must be ex post efficient because otherwise the principal can surely increase her expected profit by offering an alternative wage scheme that Pareto-improves upon it (which exists by Definition 4), failing the interim rationality. Hence, by Lemma 4, this passage must be the one that the flat wage scheme $\mathbf{w}^\ell$ prompts, i.e., $(c^\ell, \mathbf{w}^\ell, \ell)$. In particular, ℓ is the only effort level that can be induced in any deterministic equilibrium.

We show in this section that such inefficiency of an equilibrium contract is a robust conclusion that generalizes beyond deterministic solutions. We first extend the same basic arguments to solutions in which the agent may randomize over a finite number of passages, then to the cases that he may randomize in any manner.

Consider an equilibrium passage distribution with a finite support and let $\mathcal{P}(v^*)$ be the associated passage profile. Let $C^* = \{c^1, \dots, c^I\}$ denote the set of equilibrium types and for each c^i, $i = 1, \dots, I$, let $(c^i, \mathbf{w}^i, e^i)$ denote the "most profitable passage" of the c^i-type in the sense that $\Pi(\mathbf{w}^i, e^i) \geq \Pi(\mathbf{w}, e)$ for all $(c, \mathbf{w}, e) \in \mathcal{P}(v^*)$ with $c = c^i$. Since $\Pi(\mathbf{w}, e)$ decreases as $\mathbf{w}$ moves away from the certainty line along $\mathrm{IC}^e(v^*)$, for each c^i there are at most two most profitable passages, one for each $e = h, \ell$: if indeed there are two for some c^i, let $(c^i, \mathbf{w}^i, e^i)$ be the one associated with h, i.e., $e^i = h$. (In particular, if $c^\ell = c^h = c^i$ for some i and both $(c^\ell, \mathbf{w}^\ell, \ell)$ and $(c^h, \mathbf{w}^h, h)$ are most profitable passages, let $(c^i, \mathbf{w}^i, e^i) = (c^h, \mathbf{w}^h, h)$.) The next lemma states that the incentive constraints generally do not bind for each $(c^i, \mathbf{w}^i, e^i)$ at the interim stage, and identifies the exceptional case.

LEMMA 6: *For each i, a c^i-type agent strictly prefers $\mathbf{w}^i$ to $\mathbf{w}^j$ for all $j \neq i$ unless $\mathbf{w}^j = \mathbf{w}^h$ and $\{c^i, c^j\} = \{c^h, c^*(\mathbf{w}^h, \ell)\}$: in the latter case, he is indifferent between $\mathbf{w}^i$ and $\mathbf{w}^j = \mathbf{w}^h$.*

PROOF: Since $c^i \neq c^j$, it is obvious that each c^i strictly prefers $\mathbf{w}^i$ to $\mathbf{w}^j$ unless c^i is optimal for $\mathbf{w}^j$ as well, which is possible only if $\mathbf{w}^j = \mathbf{w}^h$ and $\{c^i, c^j\} = \{c^h, c^*(\mathbf{w}^h, \ell)\}$ by Lemma 2. If in fact $\mathbf{w}^j = \mathbf{w}^h$ and $c^i = c^h$ or $c^*(\mathbf{w}^h, \ell)$, then the c^i-type can derive the equilibrium utility v^* from $\mathbf{w}^h$; hence he is indifferent between $\mathbf{w}^i$ and $\mathbf{w}^j$. *Q.E.D.*

We now construct an alternative wage menu $\widehat{W}$ that, if offered instead of $W(\mathcal{P}(v^*))$, would surely increase the principal's expected profit: $\widehat{W}$ will primarily consist of wage schemes that Pareto-improve $(c^i, \mathbf{w}^i, e^i)$, $\mathbf{w}^i \neq \mathbf{w}^\ell$, each of which will be adopted by the type it improves upon.

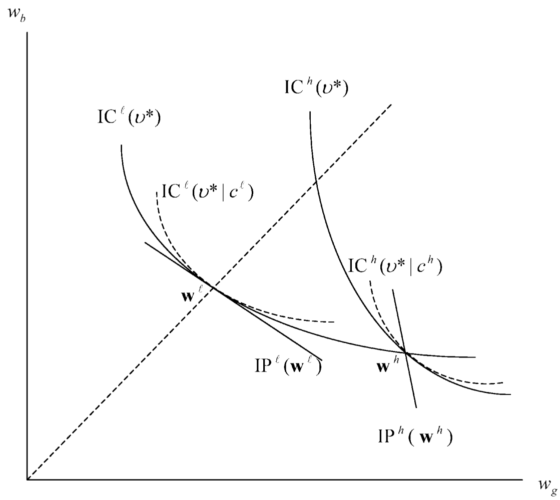

FIGURE 2.

It proves helpful to illustrate the construction of $\widehat{W}$ graphically first. To derive the equilibrium utility level from any wage scheme $\mathbf{w} \in \mathrm{IC}^e(v^*)$, the agent should obtain a type that is optimal for $(\mathbf{w}, e)$. As soon as he commits to this type, however, other wage schemes in $\mathrm{IC}(v^*)$ for which this type is suboptimal become inferior. So the ex ante indifference curve $\mathrm{IC}^e(v^*)$ is a lower envelope of interim indifference curves of various types c, denoted by $\mathrm{IC}^e(v^*|c)$, that are tangent to $\mathrm{IC}^e(v^*)$ at the wage schemes for which c is optimal. Two interim indifference curves, $\mathrm{IC}^\ell(v^*|c^\ell)$ and $\mathrm{IC}^h(v^*|c^h)$, are illustrated as dotted curves in Figure 2 that "curled up" from the ex ante curves.

The phenomenon that indifference curves curl up with type selection reflects the wealth effect that changes the agent's preference over future contracts; in particular, the wage scheme of an equilibrium passage $(c^i, \mathbf{w}^i, e^i)$ is perceived strictly inferior by other equilibrium types. Therefore, unless $\mathbf{w}^i = \mathbf{w}^\ell$, a wage scheme $\hat{\mathbf{w}}^i$ can be found above $\mathrm{IC}^{e^i}(v^*|c^i)$ and below $\mathrm{IP}(\mathbf{w}^i, e^i)$, sufficiently close to $\mathbf{w}^i$ not to attract any other c^j. If, as explained in step 2 below, distinct wage schemes need be found that Pareto-improve upon $(c^h, \mathbf{w}^h, h)$ and $(c^*(\mathbf{w}^h, \ell), \mathbf{w}^h, \ell)$, respectively, this can be done because the curve $\mathrm{IC}^h(v^*|c^h)$ is steeper than $\mathrm{IC}^\ell(v^*|c^*(\mathbf{w}^h, \ell))$ at $\mathbf{w}^h$.

More formally, $\widehat{W}$ is constructed as follows. Recall from Lemma 4 that for each $\mathbf{w}^i \neq \mathbf{w}^\ell$ there is a wage scheme, denoted by $\hat{\mathbf{w}}^i$, arbitrarily close to $\mathbf{w}^i$ that Pareto-improves upon $(c^i, \mathbf{w}^i, e^i)$.

1. By Lemma 6, one can find the wage schemes $\hat{\mathbf{w}}^i$ for all $i \in J = \{j\,|\,\mathbf{w}^j \neq \mathbf{w}^\ell, \mathbf{w}^h\} \subset \{1, \dots, I\}$ that have the "no-cross-selection" property for J in the sense that each c^i-type agent with $i \in J$ strictly prefers $\hat{\mathbf{w}}^i$ to $\hat{\mathbf{w}}^j$ for any other $j \in J$.

2. Let $K = \{j \mid \mathbf{w}^j = \mathbf{w}^h\}$. By Lemma 2, there are two possible equilibrium passages prompted by $\mathbf{w}^h$, which we denote by $(c^k, \mathbf{w}^k, e^k) = (c^h, \mathbf{w}^h, h)$ and $(c^{k'}, \mathbf{w}^{k'}, e^{k'}) = (c^*(\mathbf{w}^h, \ell), \mathbf{w}^h, \ell)$. Hence, $K \subset \{k, k'\}$. Since the interim indifference curve $\mathrm{IC}^h(v^*|c^k)$ is steeper than $\mathrm{IC}^\ell(v^*|c^{k'})$ at $\mathbf{w}^h$, one can find wage schemes $\hat{\mathbf{w}}^k$ and $\hat{\mathbf{w}}^{k'}$ that Pareto-improve upon $(c^k, \mathbf{w}^k, e^k)$ and $(c^{k'}, \mathbf{w}^{k'}, e^{k'})$, respectively, in such a way that the c^k-type strictly prefers $\mathbf{w}^k$ to $\mathbf{w}^{k'}$ and vice versa: find $\hat{\mathbf{w}}^{k'}$ above $\mathrm{IC}^\ell(v^*|c^{k'})$ but below $\mathrm{IC}^h(v^*|c^k)$ and $\mathrm{IP}(\mathbf{w}^h, \ell)$; then find $\hat{\mathbf{w}}^k$ sufficiently close to $\mathbf{w}^h$ to ensure that it is inferior to $\hat{\mathbf{w}}^{k'}$ by the $c^{k'}$-type. In conjunction with step 1, therefore, one can find wage schemes $\hat{\mathbf{w}}^i$ for all $i \in J \cup K$ that have the no-cross-selection property for $J \cup K$ in the sense defined earlier. (This step also applies when K is a singleton or empty set with the obvious modifications.)

3. If any $i \notin J \cup K$, then $(c^i, \mathbf{w}^i, e^i) = (c^\ell, \mathbf{w}^\ell, \ell)$. Take $\hat{\mathbf{w}}^i = \mathbf{w}^\ell$ for this i. Let $\widehat{W} = \{\hat{\mathbf{w}}^1, \dots, \hat{\mathbf{w}}^I\}$.

We now show that in any equilibrium with a finite support, $(c^\ell, \mathbf{w}^\ell, \ell)$ is the only passage that the agent may take, because otherwise the principal can surely increase her expected profit by offering an alternative wage menu such as $\widehat{W}$ constructed above.

THEOREM 1: *In any equilibrium of Γ with a finite support, the agent takes $(c^\ell, \mathbf{w}^\ell, \ell)$ with certainty where $\mathbf{w}^\ell = (\underline{w}, \underline{w})$ and $c^\ell = \underline{w}/2$.*

PROOF: First, consider the case in which

$$c^\ell \neq c^*(\mathbf{w}^h, \ell). \tag{15}$$

If $c^i \neq c^\ell$ for all i (i.e., c^ℓ is not an equilibrium type) or $c^i = c^\ell$ for some $i \in J \cup K$ (i.e., $(c^\ell, \mathbf{w}^\ell, \ell)$ is not the most profitable passage for the c^ℓ-type), it is obvious that if $\widehat{W}$ is offered, each c^i-type would adopt $\hat{\mathbf{w}}^i$ and exert e^i. Otherwise, i.e., if $(c^i, \mathbf{w}^i, e^i) = (c^\ell, \mathbf{w}^\ell, \ell)$ for some i, then the c^ℓ-type agent still strictly prefers $\mathbf{w}^\ell$ to $\hat{\mathbf{w}}^i$ for all $i \in J$ by Lemma 2; he would strictly prefer $\mathbf{w}^\ell$ to $\hat{\mathbf{w}}^j$ for $j \in K$, too, provided that he strictly prefers $\mathbf{w}^\ell$ to $\mathbf{w}^h$. If he is indifferent between $\mathbf{w}^\ell$ and $\mathbf{w}^h$, then $c^\ell = c^h$ due to (15) and if $(c^j, \mathbf{w}^j, e^j) = (c^*(\mathbf{w}^h, \ell), \mathbf{w}^h, \ell)$ for some $j \in K$, the wage scheme $\hat{\mathbf{w}}^j$ is selected in such a way that it is inferior to $\mathbf{w}^\ell$ for the c^ℓ-type as explained in step 2 above. Therefore, the wage menu $\widehat{W}$ constructed according to steps 1–3 above, would induce each c^i-type to adopt $\hat{\mathbf{w}}^i$ and exert e^i. Since each passage $(c^i, \hat{\mathbf{w}}^i, e^i)$ is more profitable than $(c^i, \mathbf{w}^i, e^i)$ if $\mathbf{w}^i \neq \mathbf{w}^\ell$, and is equally profitable if $\mathbf{w}^i = \mathbf{w}^\ell$, offering $\widehat{W}$ would surely increase the expected profit of the principal if $\mathbf{w}^i \neq \mathbf{w}^\ell$ for at least one i, failing interim rationality. This implies that c^ℓ is the only equilibrium type and $(c^\ell, \mathbf{w}^\ell, \ell)$ is the most profitable passage of the c^ℓ-type.

If (15) does not hold, i.e., if $c^\ell = c^*(\mathbf{w}^h, \ell)$,[10] then the argument above can face a problem: If $(c^j, \mathbf{w}^j, e^j) = (c^\ell, \mathbf{w}^\ell, \ell)$ and $(c^k, \mathbf{w}^k, e^k) = (c^h, \mathbf{w}^h, h)$ for some j and k, then the c^ℓ-type would adopt $\hat{\mathbf{w}}^k$ instead of $\hat{\mathbf{w}}^j = \mathbf{w}^\ell$, and exert ℓ. (Note that $(c^\ell, \mathbf{w}^\ell, \ell)$ is more profitable than $(c^*(\mathbf{w}^h, \ell), \mathbf{w}^h, \ell)$ for the c^ℓ-type; hence the latter passage is not relevant in step 2 above.) In this case, however, $c^h \neq c^\ell$ because $c^h \neq c^*(\mathbf{w}^h, \ell)$ and so the interim value of the outside option is less than v^* for the c^h-type by Lemma 5. Hence, modify step 2 above by selecting $\hat{\mathbf{w}}^k = \mathbf{w}^h - (\varepsilon, \varepsilon)$ for sufficiently small $\varepsilon > 0$ so that a c^h-type agent still prefers $\hat{\mathbf{w}}^k$ to the outside option. The alternative wage menu $\widehat{W}$ modified in this way, would still induce each c^i to choose $\hat{\mathbf{w}}^i$ and exert e^i, thereby increasing the expected profit of the principal if $\mathbf{w}^i \neq \mathbf{w}^\ell$ for at least one i.

Hence, we proved that c^ℓ is the only equilibrium type and $(c^\ell, \mathbf{w}^\ell, \ell)$ is the most profitable passage that the c^ℓ-type uses. If the c^ℓ-type uses another passage, it must be equally profitable (otherwise, offering the singleton wage menu $\{(\underline{w}, \underline{w})\}$ would increase the principal's profit) and, therefore, the principal can increase her expected profit by offering the wage scheme that Pareto-improves upon this other passage (such a wage scheme exists by Lemma 4). Since this would violate the interim rationality, we conclude that $(c^\ell, \mathbf{w}^\ell, \ell)$ is the only equilibrium passage. Clearly $\mathbf{w}^\ell = (w^\ell, w^\ell)$ is at least as good as the outside option, i.e., $w^\ell \geq \underline{w}$ must hold. If $w^\ell > \underline{w}$, however, the principal can increase her profit by offering the wage menu $\{(\underline{w}, \underline{w})\}$ instead of $\{\mathbf{w}^\ell\}$. Therefore, $\mathbf{w}^\ell = (\underline{w}, \underline{w})$. Then, $c^\ell = \underline{w}/2$ follows from optimal consumption smoothing. *Q.E.D.*

Finally, we establish that the inefficiency of Theorem 1 is a robust phenomenon by showing that generically all equilibria have finite supports if the agent's utility function u shows DARA, CARA, or IARA, i.e., if the absolute risk aversion of u is strictly decreasing, constant, or strictly increasing.

THEOREM 2: *Every equilibrium of Γ has a finite support for all u that shows CARA or IARA, and for generic u that is C^4 and shows DARA.*

The basic argument of the proof is similar to that of Theorem 1: we show that any equilibrium with an infinite support fails interim rationality because the principal can exploit either new opportunities of Pareto improvement due to a relaxed incentive constraint, or deteriorated reservation values of certain types of the agent. The technical details are much more complicated because we must examine the full spectrum of possible randomization, so we defer a formal proof to the Appendix.

[10]This equality holds only for knife edge cases of u. Hence, the argument in the previous paragraph alone establishes the result for generic u.

8. CONCLUDING REMARKS

A simple model of moral hazard has been investigated in which private financial decisions of the agent may create endogenous hidden information on the agent's interim preference at contracting dates. It has been shown that the minimal effort is the unique effort level that can be induced in equilibrium. Such inefficiency of equilibrium extends to multiperiod models where the parties revise and renew contracts each period.

This result stands in contrast with the largely positive findings of other studies on multiperiod contracting when no hidden information can arise at potential recontracting dates: The full-commitment optimum is generally achieved by a sequence of spot contracts (Malcomson and Spinnewyn (1988) and Fudenberg, Holmstrom, and Milgrom (1990)) or short-term contracts (Rey and Salanie (1990)).

In the study of contractual relationships, it seems realistic to consider privately accessible credit markets. However, the extreme inefficiency obtained as a result does not seem to be in good accordance with the reality. This evidently calls for further research in the area of dynamic contracting for a better understanding of reality from theoretical perspectives.

Dept. of Economics, University of Pittsburgh, 4S01 W. W. Posvar Hall, Pittsburgh, PA 15260, U.S.A.

Manuscript received January, 1998; final revision received February, 2003.

APPENDIX

Theorem 2 is proved in this Appendix in a few stages. Section A.1 formulates concepts and notations needed to formally discuss passage distributions with infinite support. Then, the proof for the cases of CARA and IARA follows straightforwardly. Section A.2 describes two kinds of deviation by the principal to be used in the proof for the case of DARA, and derives some related results. Section A.3 shows that any equilibrium with infinite support must be represented by a C^2 density function that satisfies a differential equation that is necessary for the principal's deviant offers to be unprofitable according to their first-order effects. Section A.4 concludes the proof by showing that for generic utility functions such a density function is subject to some profitable deviation due to higher-order effects. The C^4 property of u is needed for technical reasons.

A.1. *Preliminaries and the Proof for CARA and IARA*

Let $\Psi_e(w_g)$ denote the implicit function that represents the ex ante $\mathrm{IC}^e(v^*)$. For notational compactness we denote $\Psi(w_g) = \Psi_e(w_g)$ in the sequel when no confusion arises. Denoting

$$c^*_\Psi(w_g) \equiv c^*\big((w_g, \Psi(w_g)), e\big),$$

the optimal type for the wage scheme $(w_g, \Psi(w_g)) \in \mathrm{IC}^e(v^*)$, we have the identity for $\mathrm{IC}^e(v^*)$:

$$u(c^*_\Psi(w_g)) + p^e u(w_g - c^*_\Psi(w_g)) + (1-p^e)u\big(\Psi(w_g) - c^*_\Psi(w_g)\big) - d(e) = v^* \tag{16}$$

for all $w_g > 0$. For easy reference we reproduce the FOC (5) that $c^*_\Psi(w_g)$ satisfies:

$$u'(c^*_\Psi(w_g)) - p^e u'(w_g - c^*_\Psi(w_g)) - (1-p^e)u'\big(\Psi(w_g) - c^*_\Psi(w_g)\big) = 0. \tag{17}$$

By the Implicit Function theorem (see, e.g., Spivak (1965, p. 41)), Ψ is C^1 because u is assumed C^2; Ψ is C^3 in case u is C^4 as assumed in Theorem 2 for the case of DARA. (Below, we treat Ψ as C^3 but only the C^1 property will be used for the cases of CARA and IARA.) By differentiating (17) w.r.t. w_g and rearranging, we get the expression for the derivative of $c_\Psi^*(w_g)$ w.r.t. w_g:

$$c_\Psi^{*\prime}(w_g) = \frac{p^e u''(w_g - c_\Psi^*(w_g)) + (1-p^e)u''(\Psi(w_g) - c_\Psi^*(w_g))\Psi'(w_g)}{u''(c_\Psi^*(w_g)) + p^e u''(w_g - c_\Psi^*(w_g)) + (1-p^e)u''(\Psi(w_g) - c_\Psi^*(w_g))}. \tag{18}$$

By the Envelope theorem, differentiation of (16) w.r.t. w_g produces

$$p^e u'(w_g - c_\Psi^*(w_g)) + (1-p^e)u'\big(\Psi(w_g) - c_\Psi^*(w_g)\big)\Psi'(w_g) = 0. \tag{19}$$

Differentiating (19) w.r.t. w_g produces

$$\begin{aligned} &p^e u''(w_g - c_\Psi^*(w_g))(1 - c_\Psi^{*\prime}) \\ &\quad + (1-p^e)\big(u''\big(\Psi(w_g) - c_\Psi^*(w_g)\big)(\Psi'(w_g) - c_\Psi^{*\prime})\Psi'(w_g) \\ &\quad\quad + u'\big(\Psi(w_g) - c_\Psi^*(w_g)\big)\Psi''(w_g)\big) = 0. \end{aligned} \tag{20}$$

The first lemma says that the optimal type increases (decreases) as the wage scheme moves away from the certainty line along $\mathrm{IC}^e(v^*)$ if u shows IARA (DARA).

LEMMA A.1: *Fix $e \in \{h, \ell\}$, and consider $(w_g, \Psi(w_g)) \in \mathrm{IC}^e(v^*)$ below the certainty line.*
(a) $c_\Psi^{*\prime}(w_g) > 0$ *if u shows IARA*;
(b) $c_\Psi^{*\prime}(w_g) = 0$ *if u shows CARA*;
(c) $c_\Psi^{*\prime}(w_g) < 0$ *if u shows DARA*.

PROOF: By (19), the numerator of (18) is

$$p^e u'(w_g - c_\Psi^*(w_g))\left(\frac{u''(w_g - c_\Psi^*(w_g))}{u'(w_g - c_\Psi^*(w_g))} + \frac{u''(\Psi(w_g) - c_\Psi^*(w_g))}{u'(\Psi(w_g) - c_\Psi^*(w_g))}\right),$$

which is negative (0, and positive, resp.) if u shows IARA (CARA, and DARA, resp.) because $\Psi(w_g) < w_g$ by supposition. Since the denominator of (18) is negative, the assertions of the lemma follow. *Q.E.D.*

For each $\mathbf{x} = (x_g, \Psi(x_g)) \in \mathrm{IC}^e(v^*)$, let $\psi_{(\mathbf{x},e)}(w_g)$ denote the implicit function that represents the interim $\mathrm{IC}^e(v^*|c_\Psi^*(x_g))$. By the Implicit Function theorem as before, $\psi_{(\mathbf{x},e)}$ is C^1 because u is C^2, and is C^3 if u is C^4. Again we simplify notation as $\psi_\mathbf{x} = \psi_{(\mathbf{x},e)}$. The function $\psi_\mathbf{x}$ satisfies (16) with $c_\Psi^*(w_g)$ replaced with a constant $c_\mathbf{x} \equiv c_\Psi^*(x_g)$ and $\Psi(w_g)$ replaced with $\psi_\mathbf{x}(w_g)$ for all w_g. Differentiating this identity twice w.r.t. w_g, we get

$$p^e u'(w_g - c_\mathbf{x}) + (1-p^e)u'(\psi_\mathbf{x}(w_g) - c_\mathbf{x})\psi_\mathbf{x}'(w_g) = 0 \quad \text{and}$$

$$\begin{aligned} &p^e u''(w_g - c_\mathbf{x}) + (1-p^e)\big(u''(\psi_\mathbf{x}(w_g) - c_\mathbf{x})(\psi_\mathbf{x}'(w_g))^2 \\ &\quad + u'(\psi_\mathbf{x}(w_g) - c_\mathbf{x})\psi_\mathbf{x}''(w_g)\big) = 0. \end{aligned} \tag{21}$$

Since the indifference curves are strictly convex as asserted earlier, the first derivatives $\Psi'(w_g)$ and $\psi_\mathbf{x}'(w_g)$ are negative and the second derivatives $\Psi''(w_g)$ and $\psi_\mathbf{x}''(w_g)$ are positive (possibly 0 at some w_g but not over an interval). Since Ψ and $\psi_\mathbf{x}$ represent two curves that are tangent at $\mathbf{x}$, their values and their first derivatives coincide at $\mathbf{x}$:

$$\psi_\mathbf{x}(x_g) = \Psi(x_g) \quad \text{and} \quad \psi_\mathbf{x}'(x_g) = \Psi'(x_g). \tag{22}$$

From (20) and (21) evaluated at $w_g = x_g$ (note $c_{\mathbf{x}} = c^*_\Psi(x_g)$), therefore, we get

$$(23)\qquad \psi''_{\mathbf{x}}(x_g) - \Psi''(x_g) = -\frac{(p^e u''(x_g - c^*_\Psi(x_g)) + (1-p^e)u''(\Psi(x_g) - c^*_\Psi(x_g))\Psi'(x_g))c^{*\prime}_\Psi(x_g)}{(1-p^e)u'(\Psi(x_g) - c^*_\Psi(x_g))}.$$

LEMMA A.2: *If u shows DARA or IARA, for $\mathbf{x} = (x_g, \Psi(x_g)) \in \mathrm{IC}^e(v^*)$ below the certainty line,*

$$\psi''_{\mathbf{x}}(x_g) - \Psi''(x_g) > 0.$$

PROOF: Note from (18) that the expression in the big parentheses in the numerator of the right-hand side of (23) and $c^{*\prime}_\Psi(x_g)$ are of opposite signs unless the latter is 0. Hence, the lemma follows from Lemma A.1. *Q.E.D.*

Consider an equilibrium probability measure F on the passages that the agent takes. Recall F_C is the marginal distribution on the equilibrium types of the agent: $F_C(c)$ is a cumulative distribution function, i.e., an increasing right-continuous function taking values in $[0, 1]$. Obviously, F is concentrated on self-optimal passages with wage schemes on $\mathrm{IC}(v^*)$, where v^* is the equilibrium utility level of the agent. It proves useful to decompose F into two measures depending on $e = h, \ell$: F^e is the measure induced by F on the passages whose effort component is e. Since each $\mathbf{w} \in \mathrm{IC}^e(v^*)$ prompts a unique self-optimal passage with e as the effort component, we treat F^e as defined on the set of wage schemes $\mathrm{IC}^e(v^*) \cap \mathrm{IC}(v^*)$ as follows: for each e, $F^e(W) = F(\{(c, \mathbf{w}, e) | \mathbf{w} \in W\})$ for measurable $W \subset \mathrm{IC}^e(v^*) \cap \mathrm{IC}(v^*)$. The support of F^e, denoted by $\mathrm{supp}(F^e)$, is a compact subset of $\mathbb{R}^2$ for each e, because the wage schemes adopted in equilibrium are bounded. Note from Lemma 2 that $\mathbf{w}^h$ is the only wage scheme contained in the domains of both F^h and F^ℓ.

For two wage schemes $\mathbf{w}, \mathbf{w}' \in \mathbb{R}^2$, $\|\mathbf{w} - \mathbf{w}'\|$ denotes the usual norm in $\mathbb{R}^2$; we use $|\mathbf{w} - \mathbf{w}'|_g$ to denote the "horizontal distance" between $\mathbf{w}$ and $\mathbf{w}'$, i.e., $|\mathbf{w} - \mathbf{w}'|_g \equiv |\mathrm{proj}(\mathbf{w}) - \mathrm{proj}(\mathbf{w}')|$, where $\mathrm{proj}(\cdot)$ is the projection map of a vector in $\mathbb{R}^2$ onto its first coordinate. For two wage schemes $\mathbf{w} = (w_g, w_b)$ and $\mathbf{w}' = (w'_g, w'_b)$, we say $\mathbf{w}$ is *left* (*right*) of $\mathbf{w}'$ if $\mathrm{proj}(\mathbf{w}) = w_g < (>)\, w'_g = \mathrm{proj}(\mathbf{w}')$. A connected subset $I \subset \mathrm{IC}^e(v^*)$ is called *an arc.* Since $\mathrm{IC}^e(v^*)$ has a negative slope everywhere, an arc can be described by two end points: we use $(\mathbf{w}, \mathbf{w}')$ and $[\mathbf{w}, \mathbf{w}']$ to denote open and closed arcs, respectively, between $\mathbf{w}$ and $\mathbf{w}'$ where $\mathbf{w}$ is left of $\mathbf{w}'$; $(\mathbf{w}, \mathbf{w}']$ and $[\mathbf{w}, \mathbf{w}')$ denote half-open and half-closed arcs.

We say that a type c (a wage scheme $\mathbf{w} \in \mathrm{IC}^e(v^*)$, resp.) is *used* in the equilibrium if $c \in \mathrm{supp}(F_C)$ ($\mathbf{w} \in \mathrm{supp}(F^e)$, resp.). A used type c (wage scheme $\mathbf{w} \in \mathrm{IC}^e(v^*)$, resp.) is *isolated* if it is the only used type (wage scheme, resp.) in a small neighborhood of types (wage schemes on $\mathrm{IC}^e(v^*)$, resp.). Note that $\mathbf{w}^h$ may be isolated relative to $\mathrm{IC}^\ell(v^*)$ but not relative to $\mathrm{IC}^h(v^*)$, or vice versa. An isolated c or $\mathbf{w} \in \mathrm{IC}^e(v^*)$ necessarily has a point mass, i.e., $F_C(\{c\}) > 0$ or $F^e(\{\mathbf{w}\}) > 0$, but the converse is not true.

For expositional ease we present the proof as if $\mathrm{supp}(F^\ell)$ is contained in the lower half space determined by the certainty line. It is straightforward to extend the proof to the case that this is not so.

We now prove Theorem 2 when u shows CARA or IARA. In these cases, it is straightforward to verify from Lemma A.1 that $\mathrm{supp}(F^h) = \emptyset$ in equilibrium. To see this for the the case of CARA, note from Lemma A.1(b) that there are at most two used types, c^ℓ and c^h, and that the c^ℓ-type adopts $(\mathbf{w}, \ell)$ for some $\mathbf{w} \in \mathrm{IC}^\ell(v^*)$ and the c^h-type adopts $(\mathbf{w}, h)$ for some $\mathbf{w} \in \mathrm{IC}^h(v^*)$. Note also that the c^ℓ-type and c^h-type would be equally happy to adopt $(c^\ell, \mathbf{w}^\ell, \ell)$ and $(c^h, \mathbf{w}^h, h)$, respectively, which is more profitable for the principal. If $\emptyset \neq \mathrm{supp}(F^h) \neq \{\mathbf{w}^h\}$, therefore, the principal can increase her expected profit by offering a wage menu $\{\mathbf{w}^\ell, \mathbf{w}^h\}$ because then the c^ℓ-type and c^h-type would invariably adopt $(c^\ell, \mathbf{w}^\ell, \ell)$ and $(c^h, \mathbf{w}^h, h)$, respectively. If $\mathrm{supp}(F^h) = \{\mathbf{w}^h\}$, on the other hand, note that the c^h-type would strictly prefer $(\mathbf{w}^h - (\varepsilon, \varepsilon), h)$ to $(\mathbf{w}^\ell, \ell)$, hence to the outside option, for small $\varepsilon > 0$ because c^h is suboptimal for $(\mathbf{w}^\ell, \ell)$. So, the principal can increase her expected profit by offering a wage menu $\{\mathbf{w}^\ell, \mathbf{w}^h - (\varepsilon, \varepsilon)\}$ because then the c^ℓ-type and c^h-type would invariably adopt $(c^\ell, \mathbf{w}^\ell, \ell)$ and $(c^h, \mathbf{w}^h - (\varepsilon, \varepsilon), h)$, respectively. Hence, we have established that interim rationality would be violated if $\text{supp}(F^h) \neq \emptyset$ in the case of CARA.

In the case of IARA, by Lemma A.1(a) and the fact that $c^h > c^*(\mathbf{w}^h, \ell)$, no two used wage schemes are adopted by the same type. If $\text{supp}(F^h) \neq \emptyset$, let $\mathbf{w}^r$ denote the right-most used wage scheme, i.e., $\text{proj}(\mathbf{w}^r) \geq \text{proj}(\mathbf{w})$ for all $\mathbf{w} \in \text{supp}(F^h)$. Then, the type that adopts $\mathbf{w}^r$, say c^r, is bigger than any other used type by Lemma A.1(a). If $\mathbf{w}^r \neq \mathbf{w}^h$, for sufficiently small $\varepsilon > 0$ let $\mathbf{w}^\varepsilon$ denote the wage scheme on $\text{IC}^h(v^*) \cap \text{IC}(v^*)$ that is left of $\mathbf{w}^r$ by ε, i.e., $|\mathbf{w}^r - \mathbf{w}^\varepsilon|_g = \varepsilon$. Consider the wage menu W^ε obtained from the supposed equilibrium wage menu W by removing all wage schemes on the arc $(\mathbf{w}^\varepsilon, \mathbf{w}^r] \subset \text{IC}^h(v^*)$ and adding $\mathbf{w}^\varepsilon$. The principal can increase her expected profit by offering a wage menu W^ε for sufficiently small $\varepsilon > 0$ because then (i) the types that were to adopt $(\mathbf{w}, h)$ with $\mathbf{w} \in (\mathbf{w}^\varepsilon, \mathbf{w}^r]$ in the supposed equilibrium would switch to $(\mathbf{w}^\varepsilon, h)$ which is more profitable, and (ii) all other used types would adopt the same wage scheme-effort pair as in the supposed equilibrium.

If $\mathbf{w}^r = \mathbf{w}^h$, on the other hand, for sufficiently small $\varepsilon > 0$ let $\mathbf{w}^\varepsilon$ denote the wage scheme on $\text{IC}^\ell(v^*)$ left of $\mathbf{w}^r = \mathbf{w}^h$ by ε, i.e., $|\mathbf{w}^h - \mathbf{w}^\varepsilon|_g = \varepsilon$. Consider a wage menu W^ε obtained from the supposed equilibrium wage menu W by removing all wage schemes on the arc $(\mathbf{w}^\varepsilon, \mathbf{w}^h] \subset \text{IC}^\ell(v^*)$ and adding $\mathbf{w}^\varepsilon$. If W^ε is offered at the interim stage for sufficiently small $\varepsilon > 0$, (i) the c^h-type would adopt $(\mathbf{w}^\varepsilon, h)$ by continuity, (ii) the types that were to adopt $(\mathbf{w}, \ell)$ with $\mathbf{w} \in (\mathbf{w}^\varepsilon, \mathbf{w}^r]$ in the supposed equilibrium would switch to $(\mathbf{w}^\varepsilon, \ell)$, and (iii) all other used types would adopt the same wage scheme-effort pair as in the supposed equilibrium. Since the switches in (i) and (ii) above increase the expected profit of the principal, we have established that interim rationality would be violated if $\text{supp}(F^h) \neq \emptyset$ in the case of IARA as well.

Since $\text{supp}(F^h) = \emptyset$ as verified above, any used wage scheme is used with ℓ so $\Pi(\mathbf{w}^\ell, \ell) > \Pi(\mathbf{w}, e)$ for any used wage scheme-effort pair $(\mathbf{w}, e)$. Therefore, $\text{supp}(F^\ell) = \{\mathbf{w}^\ell\}$ because otherwise the principal can increase her expected profit by offering a singleton wage menu $\{\mathbf{w}^\ell\}$, for then the agent will adopt $(\mathbf{w}^\ell, \ell)$ with certainty. (Recall the tie-breaking assumption that the agent chooses to work for the principal when it is equivalent to the outside option.) Hence, we have proved Theorem 2 for the cases of CARA and IARA.

A.2. *Two Kinds of Deviation by the Principal*

We consider the case of DARA in the rest of Appendix. The argument used for the case of IARA above does not work for the case of DARA. In particular, when $\mathbf{w}^r \neq \mathbf{w}^h$ and W^ε is offered instead of W as explained above, the types that were to adopt $(\mathbf{w}, h)$ with $\mathbf{w}$ sufficiently close to $\mathbf{w}^r$ may adopt $(\mathbf{w}', \ell)$ for some $\mathbf{w}' \in W \cap \text{IC}^\ell(v^*)$ that prompts c^r or a nearby type. (Such $\mathbf{w}'$ may exist in the case of DARA but may not in the case of IARA by Lemma A.1.) Hence, we need a different argument for the case of DARA. In this section, we derive some preliminary results and describe two kinds of deviant wage menu offers to be used in this argument.

In this and the next sections we characterize the equilibrium with infinite support under the hypothesis that one exists, which is implicitly assumed in the statements of all lemmas. We start with an obvious observation that if $\text{supp}(F^h) \neq \emptyset$ in equilibrium with infinite support, then

$$\Pi(\mathbf{w}^h, h) > \Pi(\mathbf{w}^\ell, \ell). \tag{24}$$

To see this, note that assuming otherwise would mean that the agent adopts wage scheme-effort pairs that are strictly less profitable than $(\mathbf{w}^\ell, \ell)$ with positive probability in the supposed equilibrium. Then, the principal can increase her expected profit by offering the singleton wage menu $\{\mathbf{w}^\ell\}$ at the interim stage, so that the agent adopts $(\mathbf{w}^\ell, \ell)$ with certainty.

The next result is that wage schemes used in equilibrium are more profitable than other wage schemes on $\text{IC}(v^*)$ that prompt the same type.

LEMMA A.3: *If* $\hat{\mathbf{w}} \in \operatorname{supp}(F^e)$ *and* $\hat{c} \equiv c^*(\hat{\mathbf{w}}, e) = c^*(\hat{\mathbf{w}}', e')$ *for some* $\hat{\mathbf{w}}' \in \mathrm{IC}^{e'}(v^*) \cap \mathrm{IC}(v^*)$, *then* $\Pi(\hat{\mathbf{w}}, e) \geq \Pi(\hat{\mathbf{w}}', e')$.

PROOF: For each $c \in \operatorname{supp}(F_C)$ and $e \in \{h, \ell\}$, let $\mathbf{w}^{(e,c)}$ denote the wage scheme on $\mathrm{IC}^e(v^*) \cap \mathrm{IC}(v^*)$ that prompts c, if it exists. (By Lemma A.1(c), it is unique if it exists.) If $\mathbf{w}^{(e,c)}$ exists for both $e = h, \ell$, and $\Pi(\mathbf{w}^{(e,c)}, e) > \Pi(\mathbf{w}^{(\tilde{e},c)}, \tilde{e})$ where $e \neq \tilde{e}$, let $\mathbf{w}^c = \mathbf{w}^{(e,c)}$ and $\tilde{\mathbf{w}}^c = \mathbf{w}^{(\tilde{e},c)}$. Clearly,

$$F\big(\{(c, \tilde{\mathbf{w}}^c, \tilde{e}) | c \in \operatorname{supp}(F_C)\}\big) = 0 \tag{25}$$

for otherwise the principal can increase her profit by replacing $\{\tilde{\mathbf{w}}^c | c \in \operatorname{supp}(F_C)\}$ with $\{\mathbf{w}^c | c \in \operatorname{supp}(F_C)\}$ in the wage menu to offer.

To reach a contradiction, suppose contrary to the lemma that $\Pi(\hat{\mathbf{w}}', e') > \Pi(\hat{\mathbf{w}}, e)$. Then, (i) there does not exist an open arc $J \subset \mathrm{IC}^{e'}(v^*) \cap \mathrm{IC}(v^*)$, $J \ni \hat{\mathbf{w}}'$, such that $\hat{c}$ is contained in the interior of $\{c^*(\mathbf{w}, e') | \mathbf{w} \in J\}$, which implies $\hat{\mathbf{w}}' \in \{\mathbf{w}^\ell, \mathbf{w}^h\}$ by Lemma A.1(c), and (ii) there is an arc $[\hat{\mathbf{w}}, \hat{\mathbf{w}}^+] \subset \mathrm{IC}^e(v^*)$ such that $F^e([\hat{\mathbf{w}}, \hat{\mathbf{w}}^+]) = 0$. To see these, note that if (i) fails, so does (25) because wage schemes on $\mathrm{IC}^{e'}(v^*)$ near $\hat{\mathbf{w}}'$ are more profitable than those on $\mathrm{IC}^e(v^*)$ near $\hat{\mathbf{w}}$ by continuity and any open arc containing $\hat{\mathbf{w}}$ has a positive measure according to F^e because $\hat{\mathbf{w}} \in \operatorname{supp}(F^e)$; the same conclusion follows analogously if (i) holds but (ii) fails. From (ii), it further follows that (iii) $F^e((\hat{\mathbf{w}}^-, \hat{\mathbf{w}})) > 0$ for any nonempty open arc $(\hat{\mathbf{w}}^-, \hat{\mathbf{w}}) \subset \mathrm{IC}^e(v^*)$. Hence, we deduce from (25) and (i) that $\hat{\mathbf{w}}' = \mathbf{w}^\ell$ if $e' = \ell$ and $\hat{\mathbf{w}}' = \mathbf{w}^h$ if $e' = h$.

Fix an arc $[\hat{\mathbf{w}}, \hat{\mathbf{w}}^+] \subset \mathrm{IC}^e(v^*)$ with $F^e([\hat{\mathbf{w}}, \hat{\mathbf{w}}^+]) = 0$ as described in (ii) above. Pick $\hat{\mathbf{w}}^-$ described in (iii) sufficiently close to $\hat{\mathbf{w}}$ so that the $\hat{c}$-type prefers $\hat{\mathbf{w}}^-$ to $\hat{\mathbf{w}}^+$. Then so do all types prompted by wage schemes on $(\hat{\mathbf{w}}^-, \hat{\mathbf{w}}]$. If the principal removes $(\hat{\mathbf{w}}^-, \hat{\mathbf{w}}]$ from the wage menu and adds $\hat{\mathbf{w}}^-$, therefore, the types prompted by wage schemes on $(\hat{\mathbf{w}}^-, \hat{\mathbf{w}}]$ would adopt either $(\hat{\mathbf{w}}^-, e)$ or $(\hat{\mathbf{w}}', e')$. For all $\mathbf{w} \in (\hat{\mathbf{w}}^-, \hat{\mathbf{w}}]$, $\Pi(\hat{\mathbf{w}}^-, e) > \Pi(\mathbf{w}, e)$ because $\hat{\mathbf{w}}^-$ is to the left of $\mathbf{w}$ on the same $\mathrm{IC}^e(v^*)$ below the certainty line, and $\Pi(\hat{\mathbf{w}}', e') > \Pi(\mathbf{w}, e)$ by continuity if $\hat{\mathbf{w}}^-$ is sufficiently close to $\hat{\mathbf{w}}$. This means that interim rationality would be violated if $\Pi(\hat{\mathbf{w}}', e') > \Pi(\hat{\mathbf{w}}, e)$. *Q.E.D.*

A *path* (of wage schemes) is a continuous function $\kappa : [0, \bar{t}] \to \mathbb{R}^2_+$ for some $\bar{t} > 0$. Given a wage scheme $\hat{\mathbf{w}} \in \mathrm{IC}^e(v^*)$ below the certainty line, a path κ converges to $\hat{\mathbf{w}}$ from above (below, resp.) if $\kappa(0) = \hat{\mathbf{w}}$ and $\kappa(t)$ lies above (below, resp.) $\mathrm{IC}^e(v^*)$ for all $t \in (0, \bar{t}]$. Given a path κ that converges to $\hat{\mathbf{w}}$, for notational ease, let $\hat{\mathbf{w}}^t \equiv \kappa(t)$ for each $t \in [0, \bar{t}]$ and let $\{\hat{\mathbf{w}}^t\}$ denote the whole path κ. A path $\{\hat{\mathbf{w}}^t\}$ is C^n, $n = 1, 2, \ldots$, if the underlying function κ is C^n on $[0, \bar{t}]$: in this case let $D\hat{\mathbf{w}}^t \equiv D\kappa(t)$, the first derivative of $\kappa(t)$.

Given a path $\{\hat{\mathbf{w}}^t\}$ that converges to $\hat{\mathbf{w}}$ from above, due to Lemma A.1(c), for sufficiently small $t > 0$ one can find two unique wage schemes $\mathbf{x}^t = (x_g^t, \Psi(x_g^t))$ and $\mathbf{y}^t = (y_g^t, \Psi(y_g^t))$ on $\mathrm{IC}^e(v^*)$ below the certainly line, with $\mathbf{x}^t$ left of $\hat{\mathbf{w}}$ and $\mathbf{y}^t$ right of $\hat{\mathbf{w}}$, such that the $c^*(\mathbf{x}^t, e)$-type is indifferent between $(\mathbf{x}^t, e)$ and $(\hat{\mathbf{w}}^t, e)$ and the $c^*(\mathbf{y}^t, e)$-type is indifferent between $(\mathbf{y}^t, e)$ and $(\hat{\mathbf{w}}^t, e)$. Given a path $\{\hat{\mathbf{w}}^t\}$ that converges to $\hat{\mathbf{w}}$ from below, for sufficiently small $t > 0$ one can find two unique wage schemes $\mathbf{x}^t = (x_g^t, \Psi(x_g^t))$ and $\mathbf{y}^t = (y_g^t, \Psi(y_g^t))$ on $\mathrm{IC}^e(v^*)$ below the certainly line, with $\mathbf{x}^t$ left of $\hat{\mathbf{w}}$ and $\mathbf{y}^t$ right of $\hat{\mathbf{w}}$, such that the $c^*(\hat{\mathbf{w}}, e)$-type is indifferent between $(\mathbf{x}^t, e)$, $(\hat{\mathbf{w}}^t, e)$ and $(\mathbf{y}^t, e)$. We may assume, by reducing $\bar{t}$ if necessary, that such $\mathbf{x}^t$ and $\mathbf{y}^t$ uniquely exist for every wage scheme $\hat{\mathbf{w}}^t = (\hat{w}_g^t, \hat{w}_b^t)$ of a path: We refer to $\mathbf{x}^t$ and $\mathbf{y}^t$ as the *left* and *right contact points* of $\hat{\mathbf{w}}^t$, respectively. Let $\lambda(t) = \hat{w}_g^t - x_g^t$ and $\tilde{\lambda}(t) = y_g^t - \hat{w}_g^t$ denote the horizontal distances from $\hat{\mathbf{w}}^t$ to $\mathbf{x}^t$ and $\mathbf{y}^t$, respectively. It is clear from continuity of u and Lemma A.1(c) that

$$\lambda(t) \to 0 \quad \text{and} \quad \tilde{\lambda}(t) \to 0 \quad \text{as} \quad t \to 0.$$

LEMMA A.4: *If* $(\hat{c}, \hat{\mathbf{w}}, e) \in \operatorname{supp}(F)$ *and* $\hat{\mathbf{w}} \notin \{\mathbf{w}^h, \mathbf{w}^\ell\}$, *then* $F_C((\hat{c}, \hat{c} + \varepsilon)) > 0$ *for all* $\varepsilon > 0$.

PROOF: Suppose, to the contrary, that $F_C((\hat{c}, \hat{c} + \varepsilon)) = 0$ for some $\varepsilon > 0$. One can find a path $\{\hat{\mathbf{w}}^t\}$ above $\mathrm{IC}^e(v^* | \hat{c})$ and below the iso-profit line $\mathrm{IP}(\hat{\mathbf{w}}, e)$, that converges to $\hat{\mathbf{w}}$. For sufficiently small $t > 0$, both contact points, $\mathbf{x}^t$ and $\mathbf{y}^t$, of $\hat{\mathbf{w}}^t$ are sufficiently close to $\hat{\mathbf{w}}$ so that

$\mathbf{x}^t, \mathbf{y}^t \in \mathrm{IC}^e(v^*) \cap \mathrm{IC}(v^*)$ and $\hat{c} < c^*(\mathbf{x}^t, e) < \hat{c} + \varepsilon$. If $\hat{\mathbf{w}}^t$ is added to the supposed equilibrium wage menu for such t, the used types that would switch to $(\hat{\mathbf{w}}^t, e)$ from the supposed equilibrium passages, are exactly those that are prompted by wage schemes in the arc $[\mathbf{x}^t, \mathbf{y}^t] \subset \mathrm{IC}^e(v^*)$. Since we assumed $F_C((\hat{c}, \hat{c} + \varepsilon)) = 0$, in light of Lemma A.1(c), all the used types that would switch to $(\hat{\mathbf{w}}^t, e)$ are contained in the interval $[c^*(\mathbf{y}^t, e), \hat{c}]$. Note that the wage schemes on $\mathrm{IC}^e(v^*)$ that prompt these types are right of $\hat{\mathbf{w}}$, and the wage schemes on $\mathrm{IC}^{e'}(v^*)$, $e' \neq e$, that prompt these types are right of the wage scheme on $\mathrm{IC}^{e'}(v^*)$, say $\hat{\mathbf{w}}'$, that prompt $\hat{c}$. By Lemma A.3, this means that $(\hat{\mathbf{w}}, e)$ is more profitable than the wage schemes that these types would adopt in the supposed equilibrium. Since $\Pi(\hat{\mathbf{w}}^t, e) > \Pi(\hat{\mathbf{w}}, e)$, therefore, every used type that switches to $(\hat{\mathbf{w}}^t, e)$ would increase the profit of the principal. Note $F_C((c^*(\mathbf{y}^t, e), \hat{c}]) > 0$ because $\hat{c} \in \mathrm{supp}(F_C)$, yet $F_C((\hat{c}, \hat{c}+\varepsilon)) = 0$. Hence, adding $\hat{\mathbf{w}}^t$ to the wage menu would indeed increase the expected profit of the principal, violating the interim rationality. *Q.E.D.*

LEMMA A.5: *There is an arc* $(\mathbf{w}^h, \breve{\mathbf{w}}) \subset \mathrm{IC}^h(v^*)$ *such that* (i) $\Pi(\breve{\mathbf{w}}, h) > \Pi(\mathbf{w}^\ell, \ell)$, (ii) $\mathrm{supp}(F^h)$ *is dense on* $(\mathbf{w}^h, \breve{\mathbf{w}})$, *and* (iii) *for every* $\mathbf{w} \in (\mathbf{w}^h, \breve{\mathbf{w}})$, $c^*(\mathbf{w}, h)$*-type adopts only* $(\mathbf{w}, h)$ *in equilibrium.*

PROOF: The first step is to show $\mathbf{w}^h \in \mathrm{supp}(F^h)$. Let $\tilde{\mathbf{w}}$ denote the left-most wage scheme in $\mathrm{supp}(F^h)$, i.e., $\mathrm{proj}(\tilde{\mathbf{w}}) = \min\{\mathrm{proj}(\mathbf{w}) | \mathbf{w} \in \mathrm{supp}(F^h)\}$. Then, $\Pi(\tilde{\mathbf{w}}, h) > \Pi(\mathbf{w}^\ell, \ell)$ for otherwise, $(\mathbf{w}^\ell, \ell)$ is more profitable than any other wage scheme-effort pair adopted in the equilibrium; hence the principal could increase the expected profit by offering a singleton wage menu $\{\mathbf{w}^\ell\}$ in the interim stage so that the agent would adopt $(\mathbf{w}^\ell, \ell)$ with certainty. If $\tilde{\mathbf{w}} \neq \mathbf{w}^h$, by Lemma A.4 one can find an arbitrarily small arc $(\mathbf{w}', \tilde{\mathbf{w}}') \subset \mathrm{IC}^\ell(v^*)$ such that $c^*(\tilde{\mathbf{w}}', \ell) = c^*(\tilde{\mathbf{w}}, h)$ and $F^\ell((\mathbf{w}', \tilde{\mathbf{w}}')) > 0$. This, however, contradicts Lemma A.3 because $\Pi(\tilde{\mathbf{w}}', \ell) < \Pi(\mathbf{w}^\ell, \ell) < \Pi(\tilde{\mathbf{w}}, h)$. Hence, we conclude $\tilde{\mathbf{w}} = \mathbf{w}^h$ and consequently, $\mathbf{w}^h \in \mathrm{supp}(F^h)$.

As the next step we show that $\mathbf{w}^h$ is not isolated in $\mathrm{supp}(F^h)$ in the next three paragraphs. Suppose otherwise, i.e., there is an arc $(\mathbf{w}^h, \bar{\mathbf{w}}) \subset \mathrm{IC}^h(v^*)$ of wage schemes that are not used. Since $\Pi(\mathbf{w}^h, h) > \Pi(\mathbf{w}^\ell, \ell)$ by (24), Lemmas A.1 and A.3 imply that c^h is the only used type in a small neighborhood, say $(c', c'') \ni c^h$. If $c^h = c^\ell$ in this case, let $\hat{c}$ denote the maximum used type that is not an isolated point in $\mathrm{supp}(F_C)$: $\hat{c}$ exists because $\mathrm{supp}(F_C)$ is compact and assumed infinite. Then, there is a passage $(\hat{c}, \hat{\mathbf{w}}, e) \in \mathrm{supp}(F)$ such that $\hat{\mathbf{w}} \notin \{\mathbf{w}^\ell, \mathbf{w}^h\}$ and $F_C((\hat{c}, \hat{c}+\varepsilon)) > 0$ for sufficiently small $\varepsilon > 0$, contradicting Lemma A.4. Hence, $c^h \neq c^\ell$ in this case.

If $\mathbf{w}^h \notin \mathrm{supp}(F^\ell)$, i.e., $(\mathbf{w}^h, \ell)$ is not used, there is a small neighborhood of $\mathbf{w}^h$ in which $\mathbf{w}^h$ is the only used wage scheme. Since the wage scheme on $\mathrm{IC}^\ell(v^*)$ that prompts the c^h-type, if it exists, is not in $\mathrm{supp}(F^\ell)$ by Lemma A.3, this implies that the highest expected utility level the c^h-type may derive from any other used wage scheme is bounded away from v^*. Therefore, if the principal replaces $\mathbf{w}^h$ with $\mathbf{w}^h - (\varepsilon, \varepsilon)$ for sufficiently small $\varepsilon > 0$, by continuity the c^h-type would adopt $(\mathbf{w}^h - (\varepsilon, \varepsilon), h)$ over any other used wage scheme in the wage menu and over the outside option because $c^h \neq c^\ell$. This would increase the principal's expected profit, contradicting interim rationality.

If $\mathbf{w}^h \in \mathrm{supp}(F^\ell)$, i.e., $(\mathbf{w}^h, \ell)$ is used, the $c^*(\mathbf{w}^h, \ell)$-type may adopt at most two passages in the supposed equilibrium, namely, $(\mathbf{w}^h, \ell)$ and $(\mathbf{w}^\dagger, h)$ where $\mathbf{w}^\dagger$ is the wage scheme on $\mathrm{IC}^h(v^*)$ that prompts $c^*(\mathbf{w}^h, \ell)$. If $\mathbf{w}^\dagger$ is indeed used, then $\Pi(\mathbf{w}^h, \ell) = \Pi(\mathbf{w}^\dagger, h)$ by Lemma A.3. Furthermore, no wage scheme on $\mathrm{IC}^h(v^*)$ to the right of $\mathbf{w}^\dagger$ is used: otherwise, the principal could increase the profit by removing such wage schemes from the wage menu, for the the agent will then switch to other wage schemes in the menu, all of which are more profitable than the deleted ones because $\Pi(\mathbf{w}^h, \ell) = \Pi(\mathbf{w}^\dagger, h)$ and the wage scheme becomes more profitable as it moves toward the certainty line along $\mathrm{IC}^e(v^*)$. In light of these observations, we deduce that the principal can increase profit by removing all wage schemes that types in $[c^*(\mathbf{w}^h, \ell), c^*(\mathbf{w}^h, \ell) + \varepsilon)$ are supposed to adopt in equilibrium and adding the wage scheme on $\mathrm{IC}^\ell(v^*)$ that prompts $c^*(\mathbf{w}^h, \ell) + \varepsilon$, denoted by $\mathbf{w}^\varepsilon$, for sufficiently small $\varepsilon > 0$: Then, all used types in $[c^*(\mathbf{w}^h, \ell), c^*(\mathbf{w}^h, \ell) + \varepsilon)$ would adopt $(\mathbf{w}^\varepsilon, \ell)$ and the c^h-type would adopt $(\mathbf{w}^\varepsilon, h)$, and all these switches increase the expected profit. This establishes that $\mathbf{w}^h$ must not be isolated in $\mathrm{supp}(F^h)$.

Therefore, there is a used $\check{c} < c^h$ sufficiently close to c^h such that $c^\ell \notin (\check{c}, c^h)$ and the wage scheme on $\mathrm{IC}^h(v^*)$ that prompts $\check{c}$, say $\check{\mathbf{w}}$, is close enough to $\mathbf{w}^h$ to ensure $\Pi(\check{\mathbf{w}}, h) > \Pi(\mathbf{w}^\ell, \ell)$. This establishes the property (i) in the lemma directly; it also establishes the property (iii) in conjunction with Lemma A.3. To show property (ii), note that if $\mathrm{supp}(F^h)$ was not dense in $(\mathbf{w}^h, \check{\mathbf{w}})$, i.e., there was an open arc, say $(\mathbf{w}', \mathbf{w}'') \subset (\mathbf{w}^h, \check{\mathbf{w}})$, disjoint from $\mathrm{supp}(F^h)$, then $\hat{c} \equiv \sup\{c \in \mathrm{supp}(F_C) | c < c^*(\mathbf{w}'', h)\}$ would exist because $\{c \in \mathrm{supp}(F_C) | c < c^*(\mathbf{w}'', h)\}$ contains $\check{c}$, hence is nonempty. Since $\hat{c} \in \mathrm{supp}(F_C)$ because $\mathrm{supp}(F_C)$ is compact and $\hat{c} \neq c^\ell$ because $c^\ell \notin (\check{c}, c^h)$, this would contradict Lemma A.4. This completes the proof. *Q.E.D.*

Throughout the Appendix we use $(\mathbf{w}^h, \check{\mathbf{w}})$ to denote the arc on $\mathrm{IC}^h(v^*)$ characterized in Lemma A.5. We will define deviations of the principal relative to wage schemes on this arc, using paths converging to them. The next lemma states some useful relationships between $\lambda(t)$, $\tilde{\lambda}(t)$ and t for such paths, which we prove at the end of the Appendix for smooth flow of discussion.

LEMMA A.6: *Consider a path $\{\hat{\mathbf{w}}^t\}$ that converges to $\hat{\mathbf{w}} = (\hat{w}_g, \Psi(\hat{w}_g)) \in \mathrm{IC}^e(v^*)$ below the certainty line. If it converges to $\hat{\mathbf{w}}$ from above,*

$$\frac{\tilde{\lambda}(t)}{\lambda(t)} \to 1 \quad as \quad t \to 0.$$

If, in addition, the derivative $D\hat{\mathbf{w}}^t|_{t=0} = (a_g, a_b) \in \mathbb{R}^2$ exists and satisfies $(a_g, a_b) \cdot D_{\mathbf{w}} V(\hat{\mathbf{w}}, e) > 0$, where $D_{\mathbf{w}} V(\hat{\mathbf{w}}, e)$ is the gradient of the value function, defined in (7), *evaluated at $\hat{\mathbf{w}}$, then*

$$\text{(26)} \qquad \frac{t}{\lambda(t)^2} \to \frac{\psi''_{\hat{\mathbf{w}}}(\hat{w}_g) - \Psi''(\hat{w}_g)}{2(a_b - \Psi'(\hat{w}_g) a_g)} > 0 \quad as \quad t \to 0.$$

If $\{\hat{\mathbf{w}}^t\}$ is a C^2 path that converges to $\hat{\mathbf{w}}$ from below such that the derivative $D\hat{\mathbf{w}}^t|_{t=0} = (a_g, a_b)$ satisfies $(a_g, a_b) \cdot D_{\mathbf{w}} V(\hat{\mathbf{w}}, e) < 0$, then

$$\text{(27)} \qquad \frac{\tilde{\lambda}(t)}{\lambda(t)} \to 1 \quad and \quad \frac{t}{\lambda(t)^2} \to -\frac{\psi''_{\hat{\mathbf{w}}}(\hat{w}_g) - \Psi''(\hat{w}_g)}{2(a_b - \Psi'(\hat{w}_g) a_g)} > 0 \quad as \quad t \to 0.$$

The proof is provided at the end of the Appendix.

LEMMA A.7: *No wage scheme on the arc $(\mathbf{w}^h, \check{\mathbf{w}}) \subset \mathrm{IC}^h(v^*)$ has a point mass.*

PROOF: Suppose, to the contrary, that $\hat{\mathbf{w}} \in (\mathbf{w}^h, \check{\mathbf{w}})$ has a point mass $m > 0$. One can find a path $\{\hat{\mathbf{w}}^t\}$ on the interim indifference curve $\mathrm{IC}^h(v^* | c^*(\hat{\mathbf{w}}, h))$, left of $\hat{\mathbf{w}}$, that converges to $\hat{\mathbf{w}}$ from above. Let $\mathbf{x}^t$ and $\mathbf{y}^t$ denote the left and right contact points of $\hat{\mathbf{w}}^t$. Note that $\mathbf{y}^t = \hat{\mathbf{w}}$ for all t by construction. If the principal adds $\hat{\mathbf{w}}^t$ to the wage menu and removes $\mathbf{x}^t$ and $\mathbf{y}^t$ for small t, then the $c^*(\mathbf{w}, h)$-type would adopt $(\mathbf{w}^t, h)$ for every $\mathbf{w} \in [\mathbf{x}^t, \mathbf{y}^t]$. The switch by the $c^*(\hat{\mathbf{w}}, h)$-type generates a profit gain: Since $\tilde{\lambda}(t) = \hat{w}_g - \hat{w}_g^t$ and $\hat{w}_b - \hat{w}_b^t = \psi'_{\hat{\mathbf{w}}}(\hat{w}_g)\tilde{\lambda}(t) + o(\tilde{\lambda}(t))$ by Taylor's theorem (see, e.g., Buck (1978), p. 148) where $o(\tilde{\lambda}(t))$ is a function that is negligible relative to $\tilde{\lambda}(t)$ in the sense that $\lim_{\tilde{\lambda}(t) \to 0}(o(\tilde{\lambda}(t))/\tilde{\lambda}(t)) = 0$, this gain is

$$\big(p^h \tilde{\lambda}(t) + (1 - p^h)\big(\psi'_{\hat{\mathbf{w}}}(\hat{w}_g)\tilde{\lambda}(t) + o(\tilde{\lambda}(t))\big)\big)m.$$

Among other types that switch, the $c^*(\mathbf{x}^t, h)$-type incurs the largest profit loss which is calculated similarly. Hence, the loss is bounded below by

$$\big(-p^h \lambda(t) - (1 - p^h)\big(\psi'_{\mathbf{x}^t}(x_g^t)\lambda(t) + o(\lambda(t))\big)\big)\big(F^h([\mathbf{x}^t, \hat{\mathbf{w}}]) - m\big),$$

the maximum loss any switch may incur (i.e., the loss incurred by the $c^*(\mathbf{x}^t, h)$-type), multiplied by the measure of types that switch except $c^*(\hat{\mathbf{w}}, h)$. Hence, the net gain is at least as big as

$$\lambda(t)\bigg[p^h \frac{\tilde{\lambda}(t)}{\lambda(t)} m - p^h\big(F^h([\mathbf{x}^t, \hat{\mathbf{w}}]) - m\big) + (1-p^h)\bigg(\psi'_{\hat{\mathbf{w}}}(\hat{w}_g)\frac{\tilde{\lambda}(t)}{\lambda(t)} m - \psi'_{\mathbf{x}^t}(x^t_g)\big(F^h([\mathbf{x}^t, \hat{\mathbf{w}}]) - m\big)\bigg)\bigg]$$

as $t \to 0$. The value in the big bracket converges to $(p^h + (1-p^h)\psi'_{\hat{\mathbf{w}}}(\hat{w}_g))m > 0$ as $t \to 0$ because $F^h([\mathbf{x}^t, \hat{\mathbf{w}}]) \to F^h(\{\mathbf{w}^h\}) = m$ and $\tilde{\lambda}(t)/\lambda(t) \to 1$ by Lemma A.6. Hence the net gain is strictly positive for sufficiently small t, violating interim rationality. *Q.E.D.*

The rest of the Appendix evolves around the profitability of the two kinds of deviations by the principal described below. We introduce some notational conventions at this point. In light of the obvious one-to-one correspondence between $\mathrm{IC}^h(v^*)$ and $\mathrm{proj}(\mathrm{IC}^h(v^*)) \subset \mathbb{R}$, we now treat F^h as a cumulative distribution function defined on $\mathrm{proj}(\mathrm{IC}^h(v^*))$ in the obvious way:[11]

$$F^h(w_g) \equiv F^h\big(\{\mathbf{w} \in \mathrm{IC}^h(v^*) \mid \mathrm{proj}(\mathbf{w}) \le w_g\}\big).$$

Since the agent takes h in most of the discussions below, we let $p = p^h$ in the sequel. We use $o(\cdot)$ to denote a function that is "negligible relative to its argument" in the sense that the ratio of $o(\cdot)$ to its argument converges to 0 as the value of its argument tends to 0.

Given a wage scheme $\hat{\mathbf{w}} = (\hat{w}_g, \Psi(\hat{w}_g)) \in (\mathbf{w}^h, \check{\mathbf{w}}) \subset \mathrm{IC}^h(v^*)$ where Ψ is the implicit function for $\mathrm{IC}^h(v^*)$, consider an arc $I_\Delta(\hat{\mathbf{w}}) = \{(w_g, \Psi(w_g)) \mid \hat{w}_g \le w_g \le \hat{w}_g + \Delta\}$ on $\mathrm{IC}^h(v^*)$ of horizontal length $\Delta \ge 0$. Let $c^*_\Psi(I_\Delta(\hat{\mathbf{w}})) = \{c^*_\Psi(w_g) \mid \hat{w}_g \le w_g \le \hat{w}_g + \Delta\}$ be the set of types that wage schemes on $I_\Delta(\hat{\mathbf{w}})$ prompt. Consider $\mathrm{IC}^h(v^* + t)$ for small $t > 0$. Since $c^*(\cdot, h)$ monotonically decreases as the wage scheme moves away from the certainty line both along $\mathrm{IC}^h(v^*)$ and along $\mathrm{IC}^h(v^* + t)$ by Lemma A.1(c), for each given $\mathbf{w} = (w_g, w_b) \in I_\Delta(\hat{\mathbf{w}})$ there is a unique wage scheme on $\mathrm{IC}^h(v^* + t)$, denoted by $\mathbf{w}^t$, such that $c^*_\Psi(w_g) = c^*(\mathbf{w}^t, h)$. It is clear by definition that $\{\mathbf{w}^t\}$ is a path that converges to $\mathbf{w}$ from above, and that $\mathbf{w}^t = (w^t_g, w^t_b)$ satisfies the following two equations for all t (note that, for notational compactness, "$\mathbf{w}$" in $\{\mathbf{w}^t\}$ is used to denote the wage scheme on $(\mathbf{w}^h, \check{\mathbf{w}})$ to which the path $\{\mathbf{w}^t\}$ converges, i.e., $\lim_{t\to 0} \mathbf{w}^t = \mathbf{w} = (w_g, w_b)$):

$$u(c^*_\Psi(w_g)) + pu(w^t_g - c^*_\Psi(w_g)) + (1-p)u(w^t_b - c^*_\Psi(w_g)) - d(h) = v^* + t \quad \text{and} \tag{28}$$

$$u'(c^*_\Psi(w_g)) - pu'(w^t_g - c^*_\Psi(w_g)) - (1-p)u'(w^t_b - c^*_\Psi(w_g)) = 0.$$

The first equation simply means that $\mathbf{w}^t \in \mathrm{IC}^h(v^* + t)$ and the second one is the FOC at $\mathbf{w}^t$. Differentiation of these two identities w.r.t. t generate two equations that involve dw^t_g/dt and dw^t_b/dt. By solving these two equations simultaneously, we obtain the derivative $D\mathbf{w}^t = (dw^t_g/dt, dw^t_b/dt)$ as

$$\frac{dw^t_g}{dt} = u''(w^t_b - c^*_\Psi(w_g))\big[p\big(u''(w^t_g - c^*_\Psi(w_g))u'(w^t_b - c^*_\Psi(w_g)) - u'(w^t_g - c^*_\Psi(w_g))u''(w^t_b - c^*_\Psi(w_g))\big)\big]^{-1} \quad \text{and} \tag{29}$$

$$\frac{dw^t_b}{dt} = -u''(w^t_g - c^*_\Psi(w_g))\big[(1-p)\big(u''(w^t_g - c^*_\Psi(w_g))u'(w^t_b - c^*_\Psi(w_g)) - u'(w^t_g - c^*_\Psi(w_g))u''(w^t_b - c^*_\Psi(w_g))\big)\big]^{-1}.$$

[11] F^h is an "un-normalized" cdf in the sense that $F^h(+\infty) > 0$ may not be 1.

Note that $\mathbf{w}^t$ is C^2 because u is C^4. For each given $\mathbf{w} = (w_g, w_b) \in I_\Delta(\hat{\mathbf{w}})$, by Taylor's theorem

(30) $$\mathbf{w}^t = \mathbf{w} + (D\mathbf{w}^t|_{t=0})t + \begin{pmatrix} o_g(t) \\ o_b(t) \end{pmatrix},$$

where $o_g(t)$ and $o_b(t)$ are negligible functions relative to t.

The first kind of deviation is for the principal to offer $I^t_\Delta(\hat{\mathbf{w}}) \equiv \{\mathbf{w}^t | \mathbf{w} \in I_\Delta(\hat{\mathbf{w}})\} \subset \mathrm{IC}^h(v^* + t)$ for some fixed $t > 0$, in addition to the supposed equilibrium wage menu W, which we refer to as an *up-deviation at* $\hat{\mathbf{w}} \in \mathrm{IC}^h(v^*)$ *for* Δ *by* t. If such an up-deviation takes place at the interim stage, each type $c^*_\Psi(\mathbf{w})$ for $\mathbf{w} \in I_\Delta(\hat{\mathbf{w}})$ would switch from adopting $(\mathbf{w}, h)$ to $(\mathbf{w}^t, h)$, incurring a change in the principal's profit that is calculated as follows due to (30):

$$-\mathbf{p} \cdot (\mathbf{w}^t - \mathbf{w}) = -(\mathbf{p} \cdot D\mathbf{w}^t|_{t=0})t + o(t) = \xi(w_g)t + o(t),$$

where $\mathbf{p} = (p, 1 - p)$ and

(31) $$\begin{aligned}\xi(w_g) \equiv{} & \big(u''(\Psi(w_g) - c^*_\Psi(w_g)) - u''(w_g - c^*_\Psi(w_g))\big) \\ & \times \big[u''(w_g - c^*_\Psi(w_g))u'(\Psi(w_g) - c^*_\Psi(w_g)) \\ & - u'(w_g - c^*_\Psi(w_g))u''(\Psi(w_g) - c^*_\Psi(w_g))\big]^{-1} < 0.\end{aligned}$$

The inequality $\xi(w_g) < 0$ is straightforward from the definition of DARA.

In addition, such an up-deviation would induce types prompted by wage schemes slightly left of $\hat{\mathbf{w}}$ to adopt $(\hat{\mathbf{w}}^t, h)$ as well. Specifically, let $\mathbf{x}^t = (x^t_g, \Psi(x^t_g)) \in \mathrm{IC}^h(v^*)$ denote the left contact point of $\hat{\mathbf{w}}^t$; then the types prompted by wage schemes $\mathbf{w}$ on the arc $[\mathbf{x}^t, \hat{\mathbf{w}}] \subset \mathrm{IC}^h(v^*)$ switch to $(\hat{\mathbf{w}}^t, h)$, each of them incurring a profit change of $\mathbf{p} \cdot (\mathbf{w} - \hat{\mathbf{w}}^t)$. Similarly, types prompted by wage schemes slightly right of $\hat{\mathbf{w}}_\Delta \equiv (\hat{w}_g + \Delta, \Psi(\hat{w}_g + \Delta))$ would adopt $(\hat{\mathbf{w}}^t_\Delta, h)$. That is, let $\mathbf{y}^t = (y^t_g, \Psi(y^t_g)) \in \mathrm{IC}^h(v^*)$ denote the right contact point of $\hat{\mathbf{w}}^t_\Delta$; then the types prompted by wage schemes $\mathbf{w} \in [\hat{\mathbf{w}}_\Delta, \mathbf{y}^t] \subset \mathrm{IC}^h(v^*)$ switch to $(\hat{\mathbf{w}}^t_\Delta, h)$, incurring a profit change of $\mathbf{p} \cdot (\mathbf{w} - \hat{\mathbf{w}}^t_\Delta)$. We consider deviations by sufficiently small $t > 0$ such that no other used types switch to wage schemes on $I^t_\Delta(\hat{\mathbf{w}})$ from their equilibrium paths.

To sum up, an up-deviation at $\hat{\mathbf{w}} \in \mathrm{IC}^h(v^*)$ for Δ by t would induce types that were to adopt wage schemes on the arc $[\mathbf{x}^t, \mathbf{y}^t] \subset \mathrm{IC}^h(v^*)$ to switch to wage schemes on $I^t_\Delta(\hat{\mathbf{w}}) \subset \mathrm{IC}^h(v^* + t)$. This would incur a net profit change of

(32) $$\begin{aligned}N^{up}(\hat{\mathbf{w}}, \Delta, t) \equiv{} & \int_{[x^t_g, \hat{w}_g]} \mathbf{p} \cdot \big((w_g, \Psi(w_g)) - \hat{\mathbf{w}}^t\big)\, dF^h(w_g) \\ & + t \int_{[\hat{w}_g, \hat{w}_g + \Delta]} \xi(w_g)\, dF^h(w_g) \\ & + \int_{[\hat{w}_g + \Delta, y^t_g]} \mathbf{p} \cdot \big((w_g, \Psi(w_g)) - \hat{\mathbf{w}}^t_\Delta\big)\, dF^h(w_g) + o(t).\end{aligned}$$

We now describe a second kind of deviation. For a given wage scheme $\hat{\mathbf{w}} \in (\mathbf{w}^h, \check{\mathbf{w}})$ consider an arc $I_\Delta(\hat{\mathbf{w}})$ on $\mathrm{IC}^h(v^*)$ and $c^*_\Psi(I_\Delta(\hat{\mathbf{w}}))$ as defined earlier, but a lower ex ante indifference curve $\mathrm{IC}^h(v^* - s)$ for small $s > 0$. As before for each given $\mathbf{w} \in I_\Delta(\hat{\mathbf{w}})$ there is a unique wage scheme on $\mathrm{IC}^h(v^* - s)$, denoted by $\mathbf{w}^s$, such that $c^*_\Psi(\mathbf{w}) = c^*(\mathbf{w}^s, h)$. It is clear from definition that $\{\mathbf{w}^s\}$ is a path that converges to $\mathbf{w}$ from below, and that $\mathbf{w}^s = (w^s_g, w^s_b)$ satisfies the two equations in (28) with "$+t$" replaced by "$-s$" on the right-hand side of the first equation (and the superscript t replaced by s). It is straightforward to verify that, as a consequence, the derivative

$$D\mathbf{w}^s = \left(\frac{dw^s_g}{ds}, \frac{dw^s_b}{ds}\right)$$

is calculated as the formulae on the right-hand side of (29) but with opposite signs (and the superscript t replaced with s) and, therefore, $\mathbf{w}^s$ is C^2 because u is C^4.

The second kind of deviation is meant to induce each $c^*_\Psi(\mathbf{w})$-type agent to adopt $(\mathbf{w}^s, h)$ for $\mathbf{w} \in I_\Delta(\hat{\mathbf{w}})$. For this, however, not only do the wage schemes on $I_\Delta(\hat{\mathbf{w}})$ need to be replaced with $I^s_\Delta(\hat{\mathbf{w}}) \equiv \{\mathbf{w}^s | \mathbf{w} \in I_\Delta(\hat{\mathbf{w}})\} \subset \text{IC}^h(v^* - s)$ but other wage schemes in the equilibrium wage menu W that any type in $c^*_\Psi(I_\Delta(\hat{\mathbf{w}}))$ prefers to wage schemes in $I^s_\Delta(\hat{\mathbf{w}})$ need also to be removed. Specifically, let $\mathbf{x}^s \in \text{IC}^h(v^*)$ denote the left contact point of $\hat{\mathbf{w}}^s$ and let $\mathbf{y}^s \in \text{IC}^h(v^*)$ denote the right contact point of $\hat{\mathbf{w}}^s_\Delta$ where $\hat{\mathbf{w}}_\Delta = (\hat{w}_g + \Delta, \Psi(\hat{w}_g + \Delta))$. For sufficiently small $\Delta \geq 0$ and $s > 0$ such that $c^*_\Psi(I_\Delta(\hat{\mathbf{w}})) \not\ni c^\ell$, a *down-deviation at* $\hat{\mathbf{w}}$ *for* Δ *by* s refers to the following manipulation: the principal removes all wage schemes on the arc $(\mathbf{x}^s, \mathbf{y}^s) \subset \text{IC}^h(v^*)$ from the supposed equilibrium wage menu W and adds $I^s_\Delta(\hat{\mathbf{w}}) \subset \text{IC}^h(v^* - s)$ and $\mathbf{x}^s$ and $\mathbf{y}^s$, i.e., offers the wage menu $(W \setminus (\mathbf{x}^s, \mathbf{y}^s)) \cup I^s_\Delta(\hat{\mathbf{w}}) \cup \{\mathbf{x}^s, \mathbf{y}^s\}$.

If such a down-deviation takes place at the interim stage, each type $c^*_\Psi(\mathbf{w})$ for $\mathbf{w} \in I_\Delta(\hat{\mathbf{w}})$ would switch from adopting $(\mathbf{w}, h)$ to $(\mathbf{w}^s, h)$. By the same reasoning as before, for sufficiently small $s > 0$ the change in the principal's profit from such a switch is calculated as

$$-\mathbf{p} \cdot (\mathbf{w} - \mathbf{w}^s) = -\xi(w_g)s + o(s),$$

where $\xi(w_g)$ is as defined in (31) above. Since $\xi(w_g) < 0$ as argued above, this constitutes a profit gain for the principal. In addition, it would induce types prompted by wage schemes on the arc $(\mathbf{x}^s, \hat{\mathbf{w}}]$ to switch to $(\mathbf{x}^s, h)$, and induce types prompted by wage schemes on the arc $[\hat{\mathbf{w}}_\Delta, \mathbf{y}^s)$ to switch to $(\mathbf{y}^s, h)$.

To sum up, a down-deviation at $\hat{\mathbf{w}} \in \text{IC}^h(v^*)$ for Δ by $s > 0$ would induce types that were to adopt wage schemes on the arc $[\mathbf{x}^s, \mathbf{y}^s] \subset \text{IC}^h(v^*)$ to switch their wage schemes as explained above, which would incur a net profit change of

$$\begin{aligned} N_{dn}(\hat{\mathbf{w}}, \Delta, s) \equiv & \int_{[x^s_g, \hat{w}_g]} \mathbf{p} \cdot \big((w_g, \Psi(w_g)) - \mathbf{x}^s\big)\, dF^h(w_g) \\ & - s \int_{[\hat{w}_g, \hat{w}_g + \Delta]} \xi(w_g)\, dF^h(w_g) \\ & + \int_{[\hat{w}_g + \Delta, y^s_g]} \mathbf{p} \cdot \big((w_g, \Psi(w_g)) - \mathbf{y}^s\big)\, dF^h(w_g) + o(s). \end{aligned} \tag{33}$$

A.3. F^h *Should Have a* C^2 *Density Function*

Note that F^h is continuous on $\text{proj}(\mathbf{w}^h, \check{\mathbf{w}})$ by Lemma A.7. Since F^h is of bounded variation, by Kolmogorov and Fomin (1970, Corollary 1, p. 331),

(A) an associated density function, denoted by f, exists and is finite almost everywhere (a.e.) on $\text{proj}((\mathbf{w}^h, \check{\mathbf{w}}))$.

In this section we show that the density function f in fact exists and is C^2 everywhere on $\text{proj}((\mathbf{w}^h, \check{\mathbf{w}}))$. To do so, we re-express the net profit changes from up- and down-deviations derived in (32) and (33) above in terms of "right-density" $f^+(w_g)$ and "left-density" $f^-(w_g)$ defined below, which exist at least a.e. on $\text{proj}((\mathbf{w}^h, \check{\mathbf{w}}))$ by (A):

$$f^+(w_g) \equiv \lim_{\delta \downarrow 0} \frac{F^h(w_g + \delta) - F^h(w_g)}{\delta} \quad \text{and}$$

$$f^-(w_g) \equiv \lim_{\delta \downarrow 0} \frac{F^h(w_g) - F^h(w_g - \delta)}{\delta}$$

if they exist (i.e., are well defined and finite). Note that $f^+(w_g), f^-(w_g) \geq 0$ if they exist. Then, from the equilibrium requirement that the net profit changes must be nonpositive for all such deviations, we show that f must be C^2 everywhere and derive a first-order differential equation that f should satisfy. This process consists of several lemmas. We start with a technical result, which we prove at the end of the Appendix for smooth flow of discussion.

LEMMA A.8: *For $\hat{w}_g \in \text{proj}((\mathbf{w}^h, \check{\mathbf{w}}))$, if $f^+(\hat{w}_g)$ exists, then*

$$\int_{[\hat{w}_g, \hat{w}_g+\delta]} (w - \hat{w}_g)\, dF^h(w) = \frac{f^+(\hat{w}_g)}{2}\delta^2 + o(\delta^2), \quad \textit{and}$$

if $f^-(\hat{w}_g)$ exists, then

$$\int_{[\hat{w}_g-\delta, \hat{w}_g]} (w - \hat{w}_g)\, dF^h(w) = \frac{-f^-(\hat{w}_g)}{2}\delta^2 + o(\delta^2).$$

The proof is provided at the end of the Appendix.

Since $(w_g, \Psi(w_g)) - \hat{\mathbf{w}}^t = (w_g, \Psi(w_g)) - \hat{\mathbf{w}} + \hat{\mathbf{w}} - \hat{\mathbf{w}}^t$, the first integral of (32) is

$$\int_{[x_g^t, \hat{w}_g]} \mathbf{p} \cdot \big((w_g, \Psi(w_g)) - \hat{\mathbf{w}}\big)\, dF^h(w_g) + \mathbf{p} \cdot (\hat{\mathbf{w}} - \hat{\mathbf{w}}^t) \int_{[x_g^t, \hat{w}_g]} dF^h(w_g). \tag{34}$$

The latter term of (34) is negligible relative to t, i.e., an $o(t)$ function, because $\hat{\mathbf{w}} - \hat{\mathbf{w}}^t$ is first-order approximated by a linear function of t due to (30) and $\int_{[x_g^t, \hat{w}_g]} dF^h(w_g) = F^h(\hat{w}_g) - F^h(x_g^t)$ vanishes as $t \to 0$. By Taylor's theorem,

$$\begin{aligned} &\mathbf{p} \cdot \big((w_g, \Psi(w_g)) - \hat{\mathbf{w}}\big) \\ &\quad = p(w_g - \hat{w}_g) + (1-p)\bigg(\Psi'(\hat{w}_g)(w_g - \hat{w}_g) \\ &\qquad + \frac{\Psi''(\hat{w}_g)}{2}(w_g - \hat{w}_g)^2 + o\big((w_g - \hat{w}_g)^2\big)\bigg). \end{aligned} \tag{35}$$

If $f^-(\hat{w}_g)$ exists, $\int_{[x_g^t, \hat{w}_g]}(w_g - \hat{w}_g)^2\, dF^h$ is negligible relative to $(\hat{w}_g - x_g^t)^2$ by Lemma A.8; hence it is negligible relative to t, i.e., an $o(t)$ function, due to (26) because $\lambda(t) = \hat{w}_g - x_g^t + a_g t$, where $a_g = d\hat{w}_g^t/dt|_{t=0}$. From (34) and (35), therefore, the first integral of (32) is calculated as

$$\begin{aligned} &\int_{[x_g^t, \hat{w}_g]} \mathbf{p} \cdot \big((w_g, \Psi(w_g)) - \hat{\mathbf{w}}\big)\, dF^h + o(t) \\ &\quad = \big(p + (1-p)\Psi'(\hat{w}_g)\big) \int_{[x_g^t, \hat{w}_g]} (w_g - \hat{w}_g)\, dF^h + o(t) \\ &\quad = -\frac{f^-(\hat{w}_g)(p + (1-p)\Psi'(\hat{w}_g))(x_g^t - \hat{w}_g)^2}{2} + o(t) \\ &\quad = \frac{-tf^-(\hat{w}_g)(p + (1-p)\Psi'(\hat{w}_g))}{(1-p)u'(\Psi(\hat{w}_g) - c_\Psi^*(\hat{w}_g))(\psi_{\hat{\mathbf{w}}}''(\hat{w}_g) - \Psi''(\hat{w}_g))} + o(t). \end{aligned} \tag{36}$$

Here, the second equality follows from Lemma A.8 and the third one follows from (26) because $(x_g^t - \hat{w}_g)^2 = \lambda(t)^2 + o(t)$ by definition of $\lambda(t)$ and $a_b - \Psi'(\hat{w}_g)a_g$ is straightforwardly calculated to be the inverse of $(1-p)u'(\Psi(\hat{w}_g) - c_\Psi^*(\hat{w}_g))$ due to (19) and the fact that, in the current case, $a_g = d\hat{w}_g^t/dt|_{t=0}$ and $a_b = d\hat{w}_b^t/dt|_{t=0}$ as defined in (29).

Similarly, if $f^+(\hat{w}_g + \Delta)$ exists, the last integral of (32) is calculated as

$$\frac{tf^+(\hat{w}_g + \Delta)(p + (1-p)\Psi'(\hat{w}_g + \Delta))}{(1-p)u'(\Psi(\hat{w}_g + \Delta) - c^*)(\psi''_{\hat{\mathbf{w}}}(\hat{w}_g + \Delta) - \Psi''(\hat{w}_g + \Delta))} + o(t). \tag{37}$$

Consequently, the net change in profits from an up-deviation at $\hat{\mathbf{w}}$ for Δ by $t > 0$ is

$$N^{up}(\hat{\mathbf{w}}, \Delta, t) = t\biggl(-f^-(\hat{w}_g)\rho(\hat{w}_g) + \int_{[\hat{w}_g, \hat{w}_g+\Delta]} \xi(w_g)\, dF^h + f^+(\hat{w}_g + \Delta)\rho(\hat{w}_g + \Delta)\biggr) + o(t), \tag{38}$$

where $\rho(\cdot)$ represents the coefficient of $tf(\cdot)$ in (36) and (37), i.e.,

$$\rho(w_g) \equiv \frac{(p + (1-p)\Psi'(w_g))}{(1-p)u'(\Psi(w_g) - c^*)(\psi''_{\hat{\mathbf{w}}}(w_g) - \Psi''(w_g))} > 0. \tag{39}$$

That $\rho(w_g) > 0$ follows from Lemma A.2 because $p + (1-p)\Psi'(\hat{w}_g) > 0$ by (19).

Analogous calculation applies to a down-deviation at $\hat{\mathbf{w}}$ for Δ by s: The first integral of (33) is calculated as $sf^-(\hat{w}_g)\rho(\hat{w}_g) + o(s)$, and the last integral of (33) is $-sf^+(\hat{w}_g + \Delta)\rho(\hat{w}_g + \Delta) + o(s)$. (Recall that we verified that $\mathbf{w}^s$ is C^2; hence (27) applies.) Therefore, if $f^-(\hat{w}_g)$ and $f^+(\hat{w}_g + \Delta)$ exist,

$$N_{dn}(\hat{\mathbf{w}}, \Delta, s) = s\biggl(f^-(\hat{w}_g)\rho(\hat{w}_g) - \int_{[\hat{w}_g, \hat{w}_g+\Delta]} \xi(w_g)\, dF^h - f^+(\hat{w}_g + \Delta)\rho(\hat{w}_g + \Delta)\biggr) + o(s). \tag{40}$$

LEMMA A.9: *For all* $\hat{w}_g \in \text{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$ *and* $\Delta \geq 0$ *such that* $f^-(\hat{w}_g)$ *and* $f^+(\hat{w}_g + \Delta)$ *exist,*

$$f^-(\hat{w}_g)\rho(\hat{w}_g) - \int_{[\hat{w}_g, \hat{w}_g+\Delta]} \xi(w_g)\, dF^h - f^+(\hat{w}_g + \Delta)\rho(\hat{w}_g + \Delta) = 0 \quad \textit{and} \tag{41}$$

$$f^-(\hat{w}_g) = \lim_{\Delta \to 0} f^+(\hat{w}_g + \Delta). \tag{42}$$

PROOF: Let $\hat{\mathbf{w}} = (\hat{w}_g, \Psi(\hat{w}_g))$. Note that the coefficient of t in (38) is identical to that of s in (40), but with the opposite sign. If (41) fails for some $\hat{w}_g$ and Δ, therefore, either $N^{up}(\hat{\mathbf{w}}, \Delta, t)$ is strictly positive for sufficiently small t or $N_{dn}(\hat{\mathbf{w}}, \Delta, s)$ is strictly positive for sufficiently small s. This would mean that either an up- or a down-deviation at $\hat{\mathbf{w}}$ for Δ by t (or s) is profitable for the principal for sufficiently small t (or s), violating interim rationality. Hence, (41) must hold.

Since $\xi(\cdot)$ and $F^h(\cdot)$ are continuous in $\text{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$, $\int_{[\hat{w}_g, \hat{w}_g+\Delta]} \xi(w_g)\, dF^h \to 0$ as $\Delta \to 0$. Together with (41) this means $f^+(\hat{w}_g + \Delta)\rho(\hat{w}_g + \Delta) \to f^-(\hat{w}_g)\rho(\hat{w}_g)$ as $\Delta \to 0$. (One can find arbitrarily small $\Delta > 0$ such that $f^+(\hat{w}_g + \Delta)$ exists because f is defined a.e.) Since $\rho(\cdot) > 0$ is continuous, (42) follows. *Q.E.D.*

LEMMA A.10: *For every* $\hat{w}_g \in \text{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$, *if* $\{w'_n\}$ *and* $\{w''_n\}$ *are two sequences that converge to* $\hat{w}_g$ *such that* $f(w'_n)$ *and* $f(w''_n)$ *exist for all* n, *then* $\lim_{n\to\infty} f(w'_n) = \lim_{n\to\infty} f(w''_n) \neq \pm\infty$.

PROOF: Consider an arbitrarily sequence $\{w'_n\}$ that converges to a point $\hat{w}_g \in \text{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$ such that $f(w'_n)$ exists for all n. If $\{f(w'_n)\}$ either has a subsequence that diverges to ∞, or has two subsequences that converge to different limit points, then one can find $\theta > 0$ and w'_i and w'_j arbitrarily close to each other such that $|f(w'_i) - f(w'_j)| > \theta$. This contradicts (41) when $\min\{w'_i, w'_j\}$ is the "$\hat{w}_g$" of (41) and $\Delta = |w'_i - w'_j|$: since $\xi(\cdot)$ and $F^h(\cdot)$ are continuous, $\int_{[\hat{w}_g, \hat{w}_g+\Delta]} \xi(w_g)\, dF^h \to 0$

as $\varDelta \to 0$; hence (41) is violated for w_i' and w_j' sufficiently close to each other because $\rho(\cdot)$ is continuous. Therefore, we conclude that $\lim_{n\to\infty} f(w_n')$ exists and is finite.

For any other sequence $\{w_n''\}$ that converges to the same point $\hat{w}_g$, $\lim_{n\to\infty} f(w_n'')$ also exists and is finite by the same reason if $f(w_n'')$ exists for all n. If $\lim_{n\to\infty} f(w_n')$ and $\lim_{n\to\infty} f(w_n'')$ differ say by $\theta > 0$, then w_n' and w_n'' get arbitrarily close while $|f(w_n') - f(w_n'')| > \theta/2$ as $n \to \infty$, contradicting (41) by the same reasoning as above. This proves the lemma. *Q.E.D.*

LEMMA A.11: *The density f exists and is continuous everywhere on* $\text{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$.

PROOF: Fix arbitrary $\hat{w}_g \in \text{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$. Let $\hat{\mathbf{w}} = (\hat{w}_g, \varPsi(\hat{w}_g)) \in (\mathbf{w}^h, \breve{\mathbf{w}})$. First we show that

$$\nexists \{\delta_n\} \text{ such that } \quad \delta_n \to 0 \quad \text{and} \quad \frac{F^h(\hat{w}_g + \delta_n) - F^h(\hat{w}_g)}{\delta_n} \to \infty \quad \text{as} \quad n \to \infty. \tag{43}$$

We prove (43) for the case that $\delta_n \downarrow 0$. (The other case is proved analogously.) Suppose, to the contrary, such a sequence $\{\delta_n\}$ exists. For each n, consider the sequence of the ratio

$$\frac{F^h(\hat{w}_g + \delta_n/2^k) - F^h(\hat{w}_g + \delta_n/2^{k+1})}{\delta_n/2^k - \delta_n/2^{k+1}} \quad \text{for} \quad k = 0, 1, \dots. \tag{44}$$

For each n, there must exist a first integer $k(n)$ such that this ratio exceeds

$$\big(F^h(\hat{w}_g + \delta_n) - F^h(\hat{w}_g)\big)/(2\delta_n)$$

when $k = k(n)$, for assuming otherwise would imply

$$\frac{F^h(\hat{w}_g + \delta_n) - \lim_{w\downarrow\hat{w}_g} F^h(w)}{\lim_{w\downarrow\hat{w}_g}(\hat{w}_g + \delta_n - w)} \le \frac{F^h(\hat{w}_g + \delta_n) - F^h(\hat{w}_g)}{2\delta_n},$$

which is impossible because F^h is continuous at $\hat{w}_g$. Consider the sequence $\{\varepsilon_n\}$, where $\varepsilon_n = \delta_n/2^{k(n)}$ for all $n = 1, 2, \dots$. Clearly, $\varepsilon_n \to 0$ and

$$\frac{F^h(\hat{w}_g + \varepsilon_n) - F^h(\hat{w}_g + \varepsilon_n/2)}{\varepsilon_n/2} \to \infty \quad \text{as} \quad n \to \infty \tag{45}$$

because this ratio is the ratio in (44) when $k = k(n)$, which exceeds

$$\big(F^h(\hat{w}_g + \delta_n) - F^h(\hat{w}_g)\big)/(2\delta_n)$$

by construction, and this last ratio explodes to ∞ by our supposition on $\{\delta_n\}$.

Consider an up-deviation at $\hat{\mathbf{w}}_\varDelta^- \equiv (\hat{w}_g - \varDelta, \varPsi(\hat{w}_g - \varDelta)) \in (\mathbf{w}^h, \breve{\mathbf{w}})$ for $\varDelta$ by t_n such that $f(\hat{w}_g - \varDelta)$ exists and $\text{proj}(\mathbf{y}^{t_n}) = y_g^{t_n} = \hat{w}_g + \varepsilon_n$, where $\mathbf{y}^{t_n}$ is the right contact point of $\hat{\mathbf{w}}^{t_n}$. Since the first integral of (32), applied to this up-deviation, is $-t_n f(\hat{w}_g - \varDelta)\rho(\hat{w}_g - \varDelta)$ as per (36), the net change in profits from this up-deviation is

$$\begin{aligned} &N^{up}(\hat{\mathbf{w}}_\varDelta^-, \varDelta, t_n) \\ &\quad = -t_n\bigg[f(\hat{w}_g - \varDelta)\rho(\hat{w}_g - \varDelta) - \int_{\hat{w}_g-\varDelta}^{\hat{w}_g} \xi(w_g)\,dF^h \\ &\qquad - \frac{1}{t_n}\int_{[\hat{w}_g, \hat{w}_g+\varepsilon_n]} \mathbf{p}\cdot\big((w_g, \varPsi(w_g)) - \hat{\mathbf{w}}^{t_n}\big)\,dF^h + \frac{o(t_n)}{t_n}\bigg]. \end{aligned} \tag{46}$$

Note that

$$\begin{aligned}
&\frac{1}{t_n}\int_{[\hat{w}_g,\hat{w}_g+\varepsilon_n]} \mathbf{p}\cdot\big((w_g,\Psi(w_g))-\hat{\mathbf{w}}^{t_n}\big)\,dF^h \\
&> \frac{1}{t_n}\int_{[\hat{w}_g+\frac{\varepsilon_n}{2},\hat{w}_g+\varepsilon_n]} \mathbf{p}\cdot\big((w_g,\Psi(w_g))-\hat{\mathbf{w}}^{t_n}\big)\,dF^h \\
&> \frac{1}{t_n}\mathbf{p}\cdot\left(\left(\hat{w}_g+\frac{\varepsilon_n}{2},\Psi\left(\hat{w}_g+\frac{\varepsilon_n}{2}\right)\right)-\hat{\mathbf{w}}+\hat{\mathbf{w}}-\hat{\mathbf{w}}^{t_n}\right)\int_{[\hat{w}_g+\frac{\varepsilon_n}{2},\hat{w}_g+\varepsilon_n]} dF^h \\
&= \frac{\varepsilon_n}{2t_n}\big(p+(1-p)\Psi'(\hat{w}_g)\big)\int_{[\hat{w}_g+\frac{\varepsilon_n}{2},\hat{w}_g+\varepsilon_n]} dF^h + o(t_n) \\
&= \frac{\varepsilon_n^2}{2t_n}\big(p+(1-p)\Psi'(\hat{w}_g)\big)\frac{F^h(\hat{w}_g+\varepsilon_n)-F^h(\hat{w}_g+\frac{\varepsilon_n}{2})}{\varepsilon_n} + o(t_n) \\
&> \frac{\tilde{\lambda}(t_n)^2}{2t_n}\big(p+(1-p)\Psi'(\hat{w}_g)\big)\frac{F^h(\hat{w}_g+\varepsilon_n)-F^h(\hat{w}_g+\frac{\varepsilon_n}{2})}{\varepsilon_n} + o(t_n) \\
&\to \infty \quad \text{as} \quad n\to\infty.
\end{aligned}$$

Here, the first two inequalities are obvious, the first equality follows from Taylor approximation of $\Psi(\hat{w}_g+(\varepsilon_n/2))$ and the fact that $\mathbf{p}\cdot(\hat{\mathbf{w}}-\hat{\mathbf{w}}^{t_n})\int_{[\hat{w}_g+(\varepsilon_n/2),\hat{w}_g+\varepsilon_n]} dF^h$ is negligible relative to t_n, the next equality is a matter of definition, the next inequality holds because $\varepsilon_n = y_g^{t_n}-\hat{w}_g = \tilde{\lambda}(t_n)+\hat{w}_g^t-\hat{w}_g > \tilde{\lambda}(t_n)$, and the final convergence is due to (26) and (45). Hence, (46) is strictly positive for large n, violating interim rationality. This proves (43).

Next, we show that

$$\begin{aligned}
(47)\qquad &\nexists\{w_n'\},\{w_n''\} \text{ such that } \quad w_n'\to\hat{w}_g,\quad w_n''\to\hat{w}_g,\quad \text{and} \\
&\left|\frac{f(w_n'')-f(w_n')}{w_n''-w_n'}\right|\to\infty \quad \text{as} \quad n\to\infty.
\end{aligned}$$

Suppose, to the contrary, that such sequences $\{w_n'\}$ and $\{w_n''\}$ exist. First consider the case that the ratio

$$\frac{f(w_n'')-f(w_n')}{w_n''-w_n'}\to\infty.$$

By taking a subsequence if necessary, assume without loss of generality that $w_n' < w_n''$. Consider an up-deviation at $\mathbf{w}_n' \equiv (w_n',\Psi(w_n'))$ for $\Delta_n = w_n''-w_n'$ by t. The net gain from this deviation, divided by $t\Delta_n$, converges to

$$\begin{aligned}
(48)\qquad &\lim_{t\to 0}\frac{N^{up}(\mathbf{w}_n',\Delta_n,t)}{t\Delta_n} \\
&= -\frac{f(w_n')}{\Delta_n}\rho(w_n')+\xi(w_n')\frac{F^h(w_n'')-F^h(w_n')}{\Delta_n}+\frac{f(w_n'')}{\Delta_n}\rho(w_n'').
\end{aligned}$$

Since $\rho(w_n'') = \rho(w_n')+\rho'(w_n')\Delta_n+o(\Delta_n)$ by Taylor's theorem, we have

$$\begin{aligned}
&-\frac{f(w_n')}{\Delta_n}\rho(w_n')+\frac{f(w_n'')}{\Delta_n}\rho(w_n'') \\
&= \rho(w_n')\frac{f(w_n'')-f(w_n')}{\Delta_n}+f(w_n'')\left(\rho'(w_n')+\frac{o(\Delta_n)}{\Delta_n}\right)
\end{aligned}$$

which explodes to ∞ by our supposition on $(f(w_n'') - f(w_n'))/(w_n'' - w_n')$ because $\rho(\cdot) > 0$ by (39) and the sequence $\{f(w_n'')\}$ converges by Lemma A.10. Since $(F^h(w_n'') - F^h(w_n'))/\Delta_n$ is bounded above by (43), therefore, (48) becomes strictly positive for sufficiently large n, violating interim rationality. For the other case that the ratio

$$\frac{f(w_n'') - f(w_n')}{w_n'' - w_n'} \to -\infty,$$

an analogous argument for down-deviation at $\min\{w_n', w_n''\}$ for $\Delta_n = |w_n'' - w_n'|$ leads to the same conclusion. Hence, (47) is proved.

Thirdly, we show that $\forall\, \hat{w}_g \in \operatorname{proj}((\mathbf{w}^h, \check{\mathbf{w}}))$,

$$\text{(49)} \qquad \lim_{\delta\downarrow 0} \frac{F^h(\hat{w}_g + \delta) - F^h(\hat{w}_g)}{\delta} = \lim_{\delta\downarrow 0} \frac{F^h(\hat{w}_g) - F^h(\hat{w}_g - \delta)}{\delta} < \infty.$$

To prove this, first we show $\lim_{\delta\downarrow 0}(F^h(\hat{w}_g + \delta) - F^h(\hat{w}_g))/\delta$ exists. By (43), there is a sequence $\delta_n \downarrow 0$ such that this ratio converges to say $a \in [0, \infty)$. To reach a contradiction, suppose there is another sequence $\varepsilon_n \downarrow 0$ such that this ratio converges to another number b, i.e.,

$$b = \lim_{n\to\infty} \frac{F^h(\hat{w}_g + \varepsilon_n) - F^h(\hat{w}_g)}{\varepsilon_n}.$$

We may assume $b > a$. Let $\hat{f} \equiv \lim_{n\to\infty} f(w_n')$ for all sequences $\{w_n'\}$ that converge to $\hat{w}_g$ such that $f(w_n')$ exists for all n. The value $\hat{f}$ exists and is unique by Lemma A.10. Let

$$\bar{f}' \equiv \lim_{\delta\downarrow 0} \sup_{\{0<\tilde{\delta}\le\delta | f(\hat{w}_g+\tilde{\delta})\text{exists}\}} \left| \frac{f(\hat{w}_g + \tilde{\delta}) - \hat{f}}{\tilde{\delta}} \right|,$$

which is well defined because f exists a.e. Furthermore, $\bar{f}'$ is finite by (47).

First consider the case that

$$b > -\frac{\hat{f}\rho'(\hat{w}_g) + \bar{f}'\rho(\hat{w}_g)}{\xi(\hat{w}_g)}, \quad \text{i.e.,}$$

$$\text{(50)} \qquad b = -\frac{\hat{f}\rho'(\hat{w}_g) + \bar{f}'\rho(\hat{w}_g)}{\xi(\hat{w}_g)} + \beta \quad \text{for some } \beta > 0.$$

For any $\theta \in (0, b)$, one can find n large enough so that

$$\frac{F^h(\hat{w}_g + \varepsilon_n) - F^h(\hat{w}_g)}{\varepsilon_n} > b - \theta,$$

and furthermore, find $\Delta' > \varepsilon_n$ arbitrarily close to ε_n such that $f(\hat{w}_g + \Delta')$ exists,

$$\frac{F^h(\hat{w}_g + \Delta') - F^h(\hat{w}_g)}{\Delta'} > b - \theta$$

and $f(\hat{w}_g + \Delta') < \hat{f} + (\bar{f}' + \theta)\Delta'$, so that

$$\begin{aligned} & f(\hat{w}_g + \Delta')\rho(\hat{w}_g + \Delta') \\ & \quad < (\hat{f} + (\bar{f}' + \theta)\Delta')\big(\rho(\hat{w}_g) + \rho'(\hat{w}_g)\Delta' + o(\Delta')\big) \\ & \quad = \hat{f}\rho(\hat{w}_g) + \big(\hat{f}\rho'(\hat{w}_g) + (\bar{f}' + \theta)\rho(\hat{w}_g)\big)\Delta' + o(\Delta'). \end{aligned}$$

By taking large enough n (hence, small enough ε_n), therefore, one can ensure that

$$f(\hat{w}_g+\Delta')\rho(\hat{w}_g+\Delta')<\hat{f}\rho(\hat{w}_g)+\big(\hat{f}\rho'(\hat{w}_g)+(\bar{f}'+\theta)\rho(\hat{w}_g)\big)\Delta'+\eta\Delta' \tag{51}$$

for arbitrarily small $\eta>0$.

Note that one can also find sufficiently small $\Delta''>0$ such that $f(\hat{w}_g-\Delta'')$ exists and, by Lemma A.10, $f(\hat{w}_g-\Delta'')\rho(\hat{w}_g-\Delta'')$ is arbitrarily close to $\hat{f}\rho(\hat{w}_g)$, in particular,

$$|f(\hat{w}_g-\Delta'')\rho(\hat{w}_g-\Delta'')-\hat{f}\rho(\hat{w}_g)|<\eta\Delta'.$$

Together with (51), this implies that

$$\begin{aligned}&f(\hat{w}_g-\Delta'')\rho(\hat{w}_g-\Delta'')-f(\hat{w}_g+\Delta')\rho(\hat{w}_g+\Delta')\\&\quad>-\big(\hat{f}\rho'(\hat{w}_g)+(\bar{f}'+\theta)\rho(\hat{w}_g)\big)\Delta'-2\eta\Delta'.\end{aligned} \tag{52}$$

In addition, since $\xi(w_g)=\xi(\hat{w}_g-\Delta'')+\xi'(\hat{w}_g-\Delta'')(w_g-\hat{w}_g+\Delta'')+o(w_g-\hat{w}_g+\Delta'')$ by Taylor's theorem, denoting $\Delta=\Delta'+\Delta''$ we get

$$\begin{aligned}&\int_{[\hat{w}_g-\Delta'',\hat{w}_g+\Delta']}\xi(w_g)\,dF^h\\&\quad=\xi(\hat{w}_g-\Delta'')\big(F^h(\hat{w}_g+\Delta')-F^h(\hat{w}_g-\Delta'')\big)+o(\Delta)\\&\quad\le\xi(\hat{w}_g-\Delta'')\big(F^h(\hat{w}_g+\Delta')-F^h(\hat{w}_g)\big)+o(\Delta)\\&\quad<\xi(\hat{w}_g-\Delta'')(b-\theta)\Delta'+o(\Delta).\end{aligned}$$

Here, the equality follows because $\int_{[\hat{w}_g-\Delta'',\hat{w}_g+\Delta']}(w_g-\hat{w}_g+\Delta'')\,dF^h$ is an $o(\Delta)$ function by Lemma A.8, the weak inequality follows because F is increasing and $\xi(\cdot)<0$ by (31), and the final inequality follows from the choice of Δ' explained above and $\xi(\cdot)<0$. Hence, Δ' and Δ'' can be chosen sufficiently small so that $\int_{[\hat{w}_g-\Delta'',\hat{w}_g+\Delta']}\xi(w_g)\,dF^h<\xi(\hat{w}_g)(b-\theta)\Delta'+\eta\Delta$. Together with (52) and $\xi(\hat{w}_g)b=-(\hat{f}\rho'(\hat{w}_g)+\bar{f}'\rho(\hat{w}_g))+\beta\xi(\hat{w}_g)$ by (50), this has the following implication on the profit change induced by a down-deviation at $\hat{\mathbf{w}}^-_{\Delta''}\equiv(\hat{w}_g-\Delta'',\Psi(\hat{w}_g-\Delta''))$ for $\Delta=\Delta'+\Delta''$ by s: according to (40),

$$\begin{aligned}&\lim_{s\downarrow 0}\frac{N_{dn}(\hat{\mathbf{w}}^-_{\Delta''},\Delta,s)}{s}\\&\quad=f(\hat{w}_g-\Delta'')\rho(\hat{w}_g-\Delta'')-\int_{[\hat{w}_g-\Delta'',\hat{w}_g+\Delta']}\xi(w_g)\,dF^h-f(\hat{w}_g+\Delta')\rho(\hat{w}_g+\Delta')\\&\quad>-\xi(\hat{w}_g)(b-\theta)\Delta'-\big(\hat{f}\rho'(\hat{w}_g)+(\bar{f}'+\theta)\rho(\hat{w}_g)\big)\Delta'-3\eta\Delta'-\eta\Delta''\\&\quad=\big(-\xi(\hat{w}_g)b-\hat{f}\rho'(\hat{w}_g)-\bar{f}'\rho(\hat{w}_g)\big)\Delta'+\theta(\xi(\hat{w}_g)-\rho(\hat{w}_g))\Delta'-3\eta\Delta'-\eta\Delta''\\&\quad=\big(\hat{f}\rho'(\hat{w}_g)+\bar{f}'\rho(\hat{w}_g)-\beta\xi(\hat{w}_g)-\hat{f}\rho'(\hat{w}_g)-\bar{f}'\rho(\hat{w}_g)\big)\Delta'\\&\qquad+\theta(\xi(\hat{w}_g)-\rho(\hat{w}_g))\Delta'-3\eta\Delta'-\eta\Delta''\\&\quad=\left(-\beta\xi(\hat{w}_g)+\theta(\xi(\hat{w}_g)-\rho(\hat{w}_g))-3\eta-\eta\frac{\Delta''}{\Delta'}\right)\Delta',\end{aligned}$$

which is strictly positive for sufficiently small θ, η, and $\Delta''<\Delta'$, because $\beta>0$ and $\xi(\cdot)<0$. (One can always choose Δ'' smaller than Δ' because the values of θ, Δ', and η place no lower bound on Δ'' other than 0.) This violates interim rationality.

For the case in which $a < -(\hat{f}\rho'(\hat{w}_g) + \bar{f}'\rho(\hat{w}_g))/\xi(\hat{w}_g)$, an analogous argument on the up-deviation at $\hat{\mathbf{w}}^-_{\Delta''}$ for $\Delta = \Delta' + \Delta''$ leads to the same conclusion. Since this case and the case of (50) cover all possibilities, we established that

$$f^+(\hat{w}_g) = \lim_{\delta\downarrow 0} \frac{F^h(\hat{w}_g + \delta) - F^h(\hat{w}_g)}{\delta}$$

exists and is finite. That

$$f^-(\hat{w}_g) = \lim_{\delta\downarrow 0} \frac{F^h(\hat{w}_g) - F^h(\hat{w}_g - \delta)}{\delta}$$

exists and is finite, can be established analogously. Then, (49) follows from (41) when $\Delta = 0$, which proves that the density f exists everywhere on $\text{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$.

It remains to show that f is continuous everywhere on $\text{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$. To reach a contradiction, suppose f is not continuous at $\hat{w}_g$. Then, by Lemma A.10, $f(\hat{w}_g) > \lim_{w\to\hat{w}_g} f(w)$ or $f(\hat{w}_g) < \lim_{w\to\hat{w}_g} f(w)$, contradicting (42). This completes the proof. *Q.E.D.*

LEMMA A.12: *The density f is differentiable everywhere on* $\text{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$.

PROOF: We need to show for each $\hat{w}_g \in \text{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$ that both $\lim_{\delta\downarrow 0}(f(\hat{w}_g + \delta) - f(\hat{w}_g))/\delta$ and $\lim_{\delta\downarrow 0}(f(\hat{w}_g) - f(\hat{w}_g - \delta))/\delta$ exist (and are finite) and are equal. Consider $\lim_{\delta\downarrow 0}(f(\hat{w}_g + \delta) - f(\hat{w}_g))/\delta$ first. By (47), this ratio is bounded above; hence there is a sequence $\delta_n \downarrow 0$ such that this ratio converges to say $a \in \mathbb{R}$. Then,

$$f(\hat{w}_g + \delta_n)\rho(\hat{w}_g + \delta_n) = f(\hat{w}_g)\rho(\hat{w}_g) + \big(a\rho(\hat{w}_g) + f(\hat{w}_g)\rho'(\hat{w}_g)\big)\delta_n + o(\delta_n).$$

Furthermore, since f and ξ are continuous everywhere on $\text{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$,

$$\int_{[\hat{w}_g, \hat{w}_g + \delta_n]} \xi(w_g)\, dF^h = \xi(\hat{w}_g) f(\hat{w}_g)\delta_n + o(\delta_n).$$

Therefore, equation (41) for $\Delta = \delta_n$ is

$$\big(\xi(\hat{w}_g) f(\hat{w}_g) + a\rho(\hat{w}_g) + f(\hat{w}_g)\rho'(\hat{w}_g)\big)\delta_n + o(\delta_n) = 0. \tag{53}$$

To show that $\lim_{\delta\downarrow 0}(f(\hat{w}_g + \delta) - f(\hat{w}_g))/\delta$ exists, it suffices to show that $\lim_{n\to\infty}(f(\hat{w}_g + \varepsilon_n) - f(\hat{w}_g))/\varepsilon_n = a$ for every sequence $\varepsilon_n \downarrow 0$. To reach a contradiction, suppose there is a sequence $\varepsilon_n \downarrow 0$ such that this limit is $b \neq a$. We may assume $b > a$. (Due to (47), one can ensure by taking a subsequence if necessary that b exists and is finite.) By the same reasoning as above, equation (41) for $\Delta = \varepsilon_n$ is

$$\big(\xi(\hat{w}_g) f(\hat{w}_g) + b\rho(\hat{w}_g) + f(\hat{w}_g)\rho'(\hat{w}_g)\big)\varepsilon_n + o(\varepsilon_n) = 0. \tag{54}$$

For (53) and (54) to hold simultaneously for sufficiently large n, the coefficient of δ_n in (53) and that of ε_n in (54) must coincide, i.e., $a = b$ must hold. This establishes that $f'^+(\hat{w}_g) \equiv \lim_{\delta\downarrow 0}(f(\hat{w}_g + \delta) - f(\hat{w}_g))/\delta$ exists and is finite. That $f'^-(\hat{w}_g) \equiv \lim_{\delta\downarrow 0}(f(\hat{w}_g) - f(\hat{w}_g - \delta))/\delta$ exists and is finite, can be established analogously.

Finally, suppose $f'^+(\hat{w}_g) = a \neq b = f'^-(\hat{w}_g)$. Then, (41) for $\hat{w}_g$ for δ_n is the same as (53). On the other hand, since

$$\int_{[\hat{w}_g - \delta_n, \hat{w}_g]} \xi(w_g)\, dF^h = \xi(\hat{w}_g) f(\hat{w}_g)\delta_n + o(\delta_n)$$

and

$$f(\hat{w}_g - \delta_n)\rho(\hat{w}_g - \delta_n) = f(\hat{w}_g)\rho(\hat{w}_g) - \big(b\rho(\hat{w}_g) + f(\hat{w}_g)\rho'(\hat{w}_g)\big)\delta_n + o(\delta_n),$$

the equation (41) for $\hat{w}'_g = \hat{w}_g - \delta_n$ and $\Delta' = \delta_n$ is

$$(55)\qquad -\big(\xi(\hat{w}_g)f(\hat{w}_g) + b\rho(\hat{w}_g) + f(\hat{w}_g)\rho'(\hat{w}_g)\big)\delta_n + o(\delta_n) = 0.$$

For (53) and (55) to hold simultaneously for sufficiently large n, it is again necessary that $a = b$, i.e., $f'^{+}(\hat{w}_g) = f'^{-}(\hat{w}_g) \neq \pm\infty$ must hold. *Q.E.D.*

LEMMA A.13: *The density f is C^2 everywhere on* $\text{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$ *and*

$$(56)\qquad f'(\hat{w}_g)\rho(\hat{w}_g) + f(\hat{w}_g)(\rho'(\hat{w}_g) + \xi(\hat{w}_g)) = 0 \quad \textit{for all } \hat{w}_g \in \text{proj}((\mathbf{w}^h, \breve{\mathbf{w}})).$$

PROOF: Given an arbitrary point $\hat{w}_g \in \text{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$, $f(w)$ is first-order approximated as[12]

$$f(w) = f(\hat{w}_g) + f'(\hat{w}_g)(w - \hat{w}_g) + o(w - \hat{w}_g)$$

for w near $\hat{w}_g$. The left-hand side of the equation (41) is then calculated as

$$\begin{aligned} & f(\hat{w}_g)\rho(\hat{w}_g) - \big(f(\hat{w}_g) + f'(\hat{w}_g)\Delta + o(\Delta)\big)\big(\rho(\hat{w}_g) + \rho'(\hat{w}_g)\Delta + o(\Delta)\big) \\ & \quad - \int_{\hat{w}_g}^{\hat{w}_g+\Delta} \big(\xi(\hat{w}_g) + \xi'(\hat{w}_g)(w - \hat{w}_g) + o(w - \hat{w}_g)\big) \\ & \qquad\qquad \times \big(f(\hat{w}_g) + f'(\hat{w}_g)(w - \hat{w}_g) + o(w - \hat{w}_g)\big)\, dw \\ & \quad = -\big(f'(\hat{w}_g)\rho(\hat{w}_g) + f(\hat{w}_g)\rho'(\hat{w}_g) + f(\hat{w}_g)\xi(\hat{w}_g)\big)\Delta + o(\Delta). \end{aligned}$$

For this to vanish for all $\Delta > 0$ as required by (41), (56) is necessary. Since f is differentiable by Lemma A.12 and ρ, ρ' and ξ are C^1 because u is C^4, it follows from (56) that f' is differentiable. This is possible only if f' is continuous, i.e., f is C^1. Hence, it further follows from (56) that f' is also C^1. *Q.E.D.*

A.4. *Conclusion of the Proof*

We showed above that any equilibrium passage distribution must have a C^2 density function f that satisfies (56) on an arc $(\mathbf{w}^h, \breve{\mathbf{w}}) \subset \text{IC}^h(v^*)$. The condition (56) is obtained to ensure that the first-order effect is nil for all small up- and down-deviations, which is necessary for them to be unprofitable for the principal. In this final section we conclude the proof by showing that for generic u there exists some up- or down-deviations that are profitable for the principal due to higher-order effects; hence such a density would violate interim rationality and fail to be an equilibrium.

For use in intermediate steps, we first describe a hypothetical situation that the agent of various types, instead of following the supposed equilibrium path represented by a density f that satisfies (56), adopts wage schemes on slightly higher ex ante indifference curves.

Let $\Psi(w_g)$ be the implicit function for the ex ante indifference curve $\text{IC}^h(v^*)$. Given $\tau > 0$, let $\Psi_\tau(w_g)$ be the implicit function for $\text{IC}^h(v^* + \tau)$; let $c^*_{\Psi_\tau}(w_g) = c^*(w_g, \Psi_\tau(w_g))$ be the optimal type for the wage scheme $(w_g, \Psi_\tau(w_g)) \in \text{IC}^h(v^* + \tau)$ and the effort level h. Define a function $\kappa^\tau_g : \text{proj}((\mathbf{w}^h, \breve{\mathbf{w}})) \to \mathbb{R}_+$ such that $(w_g, \Psi(w_g)) \in \text{IC}^h(v^*)$ and $(\kappa^\tau_g(w_g), \Psi_\tau(\kappa^\tau_g(w_g))) \in \text{IC}^h(v^* + \tau)$ prompt the same type. Denoting $w^\tau_g = \kappa^\tau_g(w_g)$ for notational compactness, $c^*_\Psi(w_g) = c^*_{\Psi_\tau}(w^\tau_g)$. Define a density $f_\tau : \{w^\tau_g | w_g \in \text{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))\} \to \mathbb{R}_+$ as $f_\tau(w^\tau_g) = f(w_g)$, that is, f_τ is the density that

[12]Since f' is not shown to be continuous, yet, we assert this approximation as an implication of the definition of $f'(\hat{w}_g)$ rather than the Taylor's theorem.

would result if the agent, having obtained the type according to f, adopts the wage scheme on $\mathrm{IC}^h(v^*+\tau)$ that prompts the same type. Let $f'_\tau(\cdot)$ be the derivative of f_τ.

By construction of f_τ, for each $\hat{w}_g \in \mathrm{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$ we have

$$f'_\tau(\hat{w}_g^\tau) \equiv \lim_{\varepsilon\to 0} \frac{f_\tau(\hat{w}_g^\tau+\varepsilon) - f_\tau(\hat{w}_g^\tau)}{\varepsilon} = \lim_{\varepsilon\to 0} \frac{f(\hat{w}_g+\tilde{\varepsilon}) - f(\hat{w}_g)}{\varepsilon},$$

where $\tilde{\varepsilon}$ is such that $c^*_\Psi(\hat{w}_g+\tilde{\varepsilon}) = c^*_{\Psi_\tau}(\hat{w}_g^\tau+\varepsilon)$. Since

$$c^*_\Psi(\hat{w}_g+\tilde{\varepsilon}) = c^*_\Psi(\hat{w}_g) + c^{*\prime}_\Psi(\hat{w}_g)\tilde{\varepsilon} + o(\tilde{\varepsilon}) \quad \text{and}$$
$$c^*_{\Psi_\tau}(\hat{w}_g^\tau+\varepsilon) = c^*_{\Psi_\tau}(\hat{w}_g^\tau) + c^{*\prime}_{\Psi_\tau}(\hat{w}_g^\tau)\varepsilon + o(\varepsilon)$$

by Taylor's theorem and $c^*_\Psi(\hat{w}_g) = c^*_{\Psi_\tau}(\hat{w}_g^\tau)$ by construction, it further follows that

$$(57) \qquad f'_\tau(\hat{w}_g^\tau) = \lim_{\varepsilon\to 0}\left(\frac{f(\hat{w}_g+\tilde{\varepsilon}) - f(\hat{w}_g)}{\tilde{\varepsilon}}\right)\frac{\tilde{\varepsilon}}{\varepsilon} = f'(\hat{w}_g)\frac{c^{*\prime}_{\Psi_\tau}(\hat{w}_g^\tau)}{c^{*\prime}_\Psi(\hat{w}_g)}.$$

Define the functions $\xi_\tau(w_g)$ and $\rho_\tau(w_g)$ for $(w_g, \Psi_\tau(w_g)) \in \mathrm{IC}^h(v^*+\tau)$ analogously to (31) and (39), respectively. Then, for $\hat{w}_g \in \mathrm{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$ we have

$$\begin{aligned}(58) \qquad & \frac{d}{d\tau}\big(f'_\tau(\hat{w}_g^\tau)\rho_\tau(\hat{w}_g^\tau) + f_\tau(\hat{w}_g^\tau)(\rho'_\tau(\hat{w}_g^\tau) + \xi_\tau(\hat{w}_g^\tau))\big)\big|_{\tau=0} \\ &= \frac{d}{d\tau}\left(f'(\hat{w}_g)\frac{c^{*\prime}_{\Psi_\tau}(\hat{w}_g^\tau)}{c^{*\prime}_\Psi(\hat{w}_g)}\rho_\tau(\hat{w}_g^\tau) + f_\tau(\hat{w}_g^\tau)(\rho'_\tau(\hat{w}_g^\tau) + \xi_\tau(\hat{w}_g^\tau))\right)\bigg|_{\tau=0} \\ &= f(\hat{w}_g)\frac{d}{d\tau}\left(\frac{\rho'(\hat{w}_g)+\xi(\hat{w}_g)}{\rho(\hat{w}_g)c^{*\prime}_\Psi(\hat{w}_g)}c^{*\prime}_{\Psi_\tau}(\hat{w}_g^\tau)\rho_\tau(\hat{w}_g^\tau) + (\rho'_\tau(\hat{w}_g^\tau) + \xi_\tau(\hat{w}_g^\tau))\right)\bigg|_{\tau=0},\end{aligned}$$

where the first equality is due to (57) and the second equality is due to (56) and $f(\hat{w}_g) = f_\tau(\hat{w}_g^\tau)$.

As indicated earlier, we will show below that a density f that satisfies (56) on the arc $(\mathbf{w}^h, \breve{\mathbf{w}}) \subset \mathrm{IC}^h(v^*)$ fails to be an equilibrium for generic u. Specifically, we first show that such a density would necessarily violate interim rationality as long as the following two conditions hold:

$$(59) \qquad \frac{d}{d\tau}\left(\frac{\rho'(\hat{w}_g)+\xi(\hat{w}_g)}{\rho(\hat{w}_g)c^{*\prime}_\Psi(\hat{w}_g)}c^{*\prime}_{\Psi_\tau}(\hat{w}_g^\tau)\rho_\tau(\hat{w}_g^\tau) + (\rho'_\tau(\hat{w}_g^\tau) + \xi_\tau(\hat{w}_g^\tau))\right)\bigg|_{\tau=0} \neq 0 \quad \text{at} \quad \hat{w}_g = w_g^h,$$

and

$$\begin{aligned}(60) \qquad B(\hat{w}_g) \equiv{}& \frac{(1-p)\Psi''(\hat{w}_g)}{3} - \frac{2(p+(1-p)\Psi'(\hat{w}_g))(\rho'(\hat{w}_g)+\xi(\hat{w}_g))}{3\rho(\hat{w}_g)} \\ & - (\mathbf{p}\cdot D\hat{\mathbf{w}}^t|_{t=0})\frac{\psi''_{\breve{\mathbf{w}}}(\hat{w}_g) - \Psi''(\hat{w}_g)}{(a_b - \Psi'(\hat{w}_g)a_g)} \neq 0 \quad \text{at} \quad \hat{w}_g = w_g^h,\end{aligned}$$

where $(a_g, a_b) = D\hat{\mathbf{w}}^t|_{t=0}$, the derivative of the path of wage schemes on $\mathrm{IC}^h(v^*+t)$ that prompt the same type as $\hat{\mathbf{w}} \in \mathrm{IC}^h(v^*)$ evaluated at $t=0$. Then, we complete the proof by showing that these two conditions indeed hold for generic u, i.e., for all u in an open and dense subset of C^4 utility functions that show DARA.

In this Appendix we provide an analysis for the case that the derivative on the left-hand side of (59) is strictly positive at $\hat{w}_g = w_g^h$, which we assume hereafter. (The analysis for the other case is analogous.) Consider a density function f that satisfies (56) on $\mathrm{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$. Then, since the left-hand side of (59) is continuous in $\hat{w}_g$, for $\hat{w}_g \in \mathrm{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$ sufficiently close to w_g^h the derivative in (58) is strictly positive as long as $f(\hat{w}_g) > 0$ which is indeed the case for $\hat{w}_g$ in an open dense subset of $\text{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$ by Lemma A.5. Together with (56), this implies that

$$f'_\tau(\hat{w}^\tau_g)\rho_\tau(\hat{w}^\tau_g) + f_\tau(\hat{w}^\tau_g)(\rho'_\tau(\hat{w}^\tau_g) + \xi_\tau(\hat{w}^\tau_g)) > 0$$

for generic $\hat{w}_g$ in a small interval $(w^h_g, w') \subset \text{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$ and sufficiently small $\tau > 0$. Consequently, for $\hat{w}_g \in (w^h_g, w')$ and sufficiently small $\Delta > 0$ and $\tau > 0$, we have

$$\text{(61)} \qquad A(\hat{w}^\tau_g, \Delta) \equiv \int_{\hat{w}^\tau_g}^{\hat{w}^\tau_g+\Delta} f'_\tau(w_g)\rho_\tau(w_g) + f_\tau(w_g)(\rho'_\tau(w_g) + \xi_\tau(w_g))\, dw_g$$
$$= f_\tau(\hat{w}^\tau_g + \Delta)\rho_\tau(\hat{w}^\tau_g + \Delta) - f_\tau(\hat{w}^\tau_g)\rho_\tau(\hat{w}^\tau_g) + \int_{\hat{w}^\tau_g}^{\hat{w}^\tau_g+\Delta} f_\tau(w_g)\xi_\tau(w_g)\, dw_g > 0.$$

In Section A.2 we defined an up-deviation for F^h on $\text{IC}^h(v^*)$, which we now extend in the obvious way for the distribution represented by the density f_τ on $\text{IC}^h(v^* + \tau)$. For $\hat{\mathbf{w}} = (\hat{w}_g, \Psi(\hat{w}_g)) \in (\mathbf{w}^h, \breve{\mathbf{w}}) \subset \text{IC}^h(v^*)$ recall that $\hat{\mathbf{w}}^\tau = (\hat{w}^\tau_g, \Psi_\tau(\hat{w}^\tau_g)) \in \text{IC}^h(v^* + \tau)$ such that $c^*(\hat{\mathbf{w}}, h) = c^*(\hat{\mathbf{w}}^\tau, h)$. Consider an up-deviation at $\hat{\mathbf{w}}^\tau$ for $\Delta > 0$ by t. Let $N^{up}_\tau(\hat{\mathbf{w}}^\tau, \Delta, t)$ denote the net change in expected profit from this up-deviation. From (38) applied to f_τ on $\text{IC}^h(v^* + \tau)$, $N^{up}_\tau(\hat{\mathbf{w}}^\tau, \Delta, t) = A(\hat{w}^\tau_g, \Delta)t + o(t)$ and, therefore,

$$\frac{N^{up}_\tau(\hat{\mathbf{w}}^\tau, \Delta, t)}{t^2} = \frac{1}{t}\left(A(\hat{w}^\tau_g, \Delta) - \frac{o(t)}{t}\right) \to \infty \quad \text{as} \quad t \to 0$$

because $A(\hat{w}^\tau_g, \Delta) > 0$ by (61). Hence, it follows from continuity of $N^{up}_\tau(\hat{\mathbf{w}}^\tau, \Delta, t)$ that this limit is nonnegative when $\Delta = 0$, i.e.,

$$\text{(62)} \qquad \lim_{t \downarrow 0} \frac{N^{up}_\tau(\hat{\mathbf{w}}^\tau, 0, t)}{t^2} \geq 0$$

for all $\hat{w}_g \in (w^h_g, w') \subset \text{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$ and sufficiently small $\tau > 0$.

We now calculate $N^{up}_\tau(\hat{\mathbf{w}}^\tau, 0, t)$ more precisely, i.e., beyond the first-order effects. Recall that we decomposed the first integral of (32) into two integrals as in (34). Using Taylor approximations of Ψ_τ and f_τ near $\hat{w}^\tau_g$, the first integral of (34) for the current up-deviation is

$$\begin{aligned}
\text{(63)} \qquad & \int_{x^t_g}^{\hat{w}^\tau_g} \big[\big(p(w - \hat{w}^\tau_g) + (1-p)\Psi'_\tau(\hat{w}^\tau_g)(w - \hat{w}^\tau_g) \\
& \quad + (1-p)\Psi''_\tau(\hat{w}^\tau_g)(w - \hat{w}^\tau_g)^2/2\big)\big(f_\tau(\hat{w}^\tau_g) + f'_\tau(\hat{w}^\tau_g)(w - \hat{w}^\tau_g)\big)\big]\, dw \\
& \quad + o((x^t_g - \hat{w}^\tau_g)^3) \\
&= -\frac{f_\tau(\hat{w}^\tau_g)(p + (1-p)\Psi'_\tau(\hat{w}^\tau_g))}{2}(\hat{w}^\tau_g - x^t_g)^2 \\
& \quad + \frac{f_\tau(\hat{w}^\tau_g)(1-p)\Psi''_\tau(\hat{w}^\tau_g) + 2(p + (1-p)\Psi'_\tau(\hat{w}^\tau_g))f'_\tau(\hat{w}^\tau_g)}{6}(\hat{w}^\tau_g - x^t_g)^3 \\
& \quad + o((\hat{w}^\tau_g - x^t_g)^3) \\
&= -f_\tau(\hat{w}^\tau_g)\rho_\tau(\hat{w}^\tau_g)t \\
& \quad + \frac{f_\tau(\hat{w}^\tau_g)(1-p)\Psi''_\tau(\hat{w}^\tau_g) + 2(p + (1-p)\Psi'_\tau(\hat{w}^\tau_g))f'_\tau(\hat{w}^\tau_g)}{6}(\hat{w}^\tau_g - x^t_g)^3 \\
& \quad + o((\hat{w}^\tau_g - x^t_g)^3),
\end{aligned}$$

where $\mathbf{x}^t = (x_g^t, \Psi_\tau(x_g^t))$ denotes the left contact point of $\hat{\mathbf{w}}^\tau$ on $\mathrm{IC}^h(v^* + \tau)$, and the second inequality follows from (36) and (39). The second integral of (34) is

$$-\big((\mathbf{p}\cdot D\hat{\mathbf{w}}^t|_{t=\tau})t + o(t)\big)\int_{x_g^t}^{\hat{w}_g^\tau}\big(f_\tau(\hat{w}_g^\tau) + f_\tau'(\hat{w}_g^\tau)(w - \hat{w}_g^\tau)\big)\,dw + o((\hat{w}_g^\tau - x_g^t)^3) \tag{64}$$

$$= -(\mathbf{p}\cdot D\hat{\mathbf{w}}^t|_{t=\tau})f_\tau(\hat{w}_g^\tau)t(\hat{w}_g^\tau - x_g^t) + o((\hat{w}_g^\tau - x_g^t)^3)$$

$$= -(\mathbf{p}\cdot D\hat{\mathbf{w}}^t|_{t=\tau})f_\tau(\hat{w}_g^\tau)\frac{\psi''_{\hat{\mathbf{w}}^\tau}(\hat{w}_g^\tau) - \Psi''_\tau(\hat{w}_g^\tau)}{2(a_b^\tau - \Psi'_\tau(\hat{w}_g)a_g^\tau)}(\hat{w}_g^\tau - x_g^t)^3 + o((\hat{w}_g^\tau - x_g^t)^3),$$

where $(a_g^\tau, a_b^\tau) = D\hat{\mathbf{w}}^t|_{t=\tau}$ and the last equality follows from (26) because $\lambda_\tau(t) = \hat{w}_g^\tau - x_g^t + a_g^\tau t$.

Similarly, with $\Delta = 0$, the last integral of (32) is the sum of the following two integrals:

$$\int_{\hat{w}_g^\tau}^{y_g^t}\Big[\big(p(w - \hat{w}_g^\tau) + (1-p)\Psi_\tau'(\hat{w}_g^\tau)(w - \hat{w}_g^\tau) + (1-p)\Psi_\tau''(\hat{w}_g^\tau)(w - \hat{w}_g^\tau)^2/2\big)\big(f_\tau(\hat{w}_g^\tau) + f_\tau'(\hat{w}_g^\tau)(w - \hat{w}_g^\tau)\big)\Big]\,dw + o((y_g^t - \hat{w}_g^\tau)^3) \tag{65}$$

$$= f_\tau(\hat{w}_g^\tau)\rho_\tau(\hat{w}_g^\tau)t + \frac{f_\tau(\hat{w}_g^\tau)(1-p)\Psi_\tau''(\hat{w}_g^\tau) + 2(p + (1-p)\Psi_\tau'(\hat{w}_g^\tau))f_\tau'(\hat{w}_g^\tau)}{6}(y_g^t - \hat{w}_g^\tau)^3 + o((y_g^t - \hat{w}_g^\tau)^3),$$

where $\mathbf{y}^t = (y_g^t, \Psi_\tau(y_g^t))$ denotes the right contact point of $\hat{\mathbf{w}}^\tau$ on $\mathrm{IC}^h(v^* + \tau)$, and

$$-\big((\mathbf{p}\cdot D\hat{\mathbf{w}}^t|_{t=\tau})t + o(t)\big)\int_{\hat{w}_g^\tau}^{y_g^t}\big(f_\tau(\hat{w}_g^\tau) + f_\tau'(\hat{w}_g^\tau)(w - \hat{w}_g^\tau)\big)\,dw + o((y_g^t - \hat{w}_g^\tau)^3) \tag{66}$$

$$= -(\mathbf{p}\cdot D\hat{\mathbf{w}}^t|_{t=\tau})f_\tau(\hat{w}_g^\tau)\frac{\psi''_{\hat{\mathbf{w}}^\tau}(\hat{w}_g^\tau) - \Psi''_\tau(\hat{w}_g^\tau)}{2(a_b^\tau - \Psi'_\tau(\hat{w}_g)a_g^\tau)}(y_g^t - \hat{w}_g^\tau)^3 + o((y_g^t - \hat{w}_g^\tau)^3).$$

Therefore, $N_\tau^{up}(\hat{\mathbf{w}}^\tau, 0, t)$ is the sum of (63)–(66). (The second integral of (32) is nil.) Since $\hat{w}_g^\tau - x_g^t = \lambda_\tau(t) - a_g^\tau t$ and $y_g^t - \hat{w}_g^\tau = \tilde{\lambda}_\tau(t) + a_g^\tau t$, from Lemma A.6 this sum is calculated as

$$N_\tau^{up}(\hat{\mathbf{w}}^\tau, 0, t) = \Bigg[\frac{f_\tau(\hat{w}_g^\tau)(1-p)\Psi_\tau''(\hat{w}_g^\tau) + 2(p + (1-p)\Psi_\tau'(\hat{w}_g^\tau))f_\tau'(\hat{w}_g^\tau)}{3} - (\mathbf{p}\cdot D\hat{\mathbf{w}}^t|_{t=\tau})f_\tau(\hat{w}_g^\tau)\frac{\psi''_{\hat{\mathbf{w}}^\tau}(\hat{w}_g^\tau) - \Psi''_\tau(\hat{w}_g^\tau)}{(a_b^\tau - \Psi'_\tau(\hat{w}_g)a_g^\tau)}\Bigg]\lambda_\tau(t)^3 + o(\lambda_\tau(t)^3). \tag{67}$$

Note here that, when $\tau = 0$, the coefficient of $\lambda_\tau(t)^3$, i.e., the formula in the big bracket of (67), is $f(\hat{w}_g)B(\hat{w}_g)$ by (56) where $B(\hat{w}_g)$ is defined in (60). Suppose $B(w_g^h) \neq 0$ as per the condition (60). If $B(w_g^h) > 0$, then $B(\hat{w}_g) > 0$ by continuity for $\hat{w}_g$ sufficiently close to w_g^h and consequently, $f(\hat{w}_g)B(\hat{w}_g) > 0$ if $f(\hat{w}_g) > 0$ which is the case for generic $\hat{w}_g$ in $\mathrm{proj}((\mathbf{w}^h, \breve{\mathbf{w}}))$ by Lemma A.5. In light of (67), this would mean that $N^{up}(\hat{\mathbf{w}}, 0, t) > 0$ for sufficiently small $t > 0$, i.e., an up-deviation at $\hat{\mathbf{w}} = (\hat{w}_g, \Psi(\hat{w}_g)) \in (\mathbf{w}^h, \breve{\mathbf{w}})$ would be profitable for the principal for small $t > 0$. This would contradict interim rationality of the passage distribution represented by f.

If $B(w_g^h) < 0$, by the same token as above, $f(\hat{w}_g)B(\hat{w}_g) < 0$ for generic $\hat{w}_g$ in a small interval $(w_g^h, w_g') \subset \text{proj}((\mathbf{w}^h, \check{\mathbf{w}}))$, hence for such $\hat{w}_g$ the coefficient of $\lambda_\tau(t)^3$ in (67) would be strictly negative for sufficiently small $\tau > 0$. This is impossible because it would contradict (62) since $\lambda_\tau(t)^3/t^2 \to \infty$ as $t \to 0$ according to (26). Therefore, we established that a density f that satisfies (56) cannot be supported as an equilibrium if (60) holds, provided that the condition (59) also holds. (As mentioned before, we provided the analysis for the case that the derivative in (59) is strictly positive; an analogous analysis applies for the other case.)

Finally, we complete the proof by showing that both (59) and (60) hold for u in an open and dense subset of C^4 utility functions that show DARA. First we show this for (60). It is obvious from continuity of B that if $B(w_g^h) \neq 0$ for some u, then the same is true for utility functions in a small neighborhood of u. This means that the set of utility functions for which (60) holds is open. It remains to show that this set is dense, for which it suffices to show the following: if $B(w_g^h) = 0$ for some u, one can find another utility function arbitrarily close to u, for which $B(w_g^h) \neq 0$.

To show this, first note that by Taylor's theorem u can be represented as

$$u(c) = u(c_g) + u'(c_g)(c - c_g) + \frac{u''(c_g)}{2}(c - c_g)^2 + \frac{u'''(c_g)}{3!}(c - c_g)^3 + R(c),$$

where $c_g \equiv w_g^h - c^h$ and $R(c)$ is a function negligible relative to $(c - c_g)^3$. For $a > 0$ let $\Omega_a(c)$ be a C^∞ function such that $\Omega_a(c) = a$ for $c \in [c_g - a, c_g + a]$ and $\Omega_a(c) = 0$ for $c \notin [c_g - 2a, c_g + 2a]$, constructed in the manner explained in Spivak (1979, p. 43). Construct $u_a(c)$ by

$$u_a(c) = u(c_g) + u'(c_g)(c - c_g) + \frac{u''(c_g)}{2}(c - c_g)^2 + \frac{u'''(c_g) + \Omega_a(c)}{3!}(c - c_g)^3 + R(c).$$

It is clear from construction of Ω_a in Spivak (1979) that $\Omega_a'(c)$ and $\Omega_a''(c)$ are bounded and, therefore, u_a is strictly concave and exhibits DARA for sufficiently small $a > 0$. Note further that the derivatives $\Omega_a(c)$ of all order vanish for $c \in [c_g - a, c_g + a]$ and $c \notin [c_g - 2a, c_g + 2a]$. For sufficiently small $a > 0$, therefore:

(i) $u(c) = u_a(c)$, $u'(c) = u_a'(c)$, $u''(c) = u_a''(c)$, and $u'''(c) = u_a'''(c)$ at $c = c^h$, $w_b^h - c^h$, $c^*(\mathbf{w}^h, \ell)$, $w_g^h - c^*(\mathbf{w}^h, \ell)$, and $w_b^h - c^*(\mathbf{w}^h, \ell)$; and

(ii) $u(c_g) = u_a(c_g)$, $u'(c_g) = u_a'(c_g)$, $u''(c_g) = u_a''(c_g)$, but $u_a'''(c_g) = u'''(c_g) + a$.

Since u and u_a agree up to the second derivatives at $c = c^h$, $w_g^h - c^h$ $(= c_g)$, and $w_b^h - c^h$, it follows from the FOC (17) that c^h is the optimal type for $(\mathbf{w}^h, h)$ under u_a, i.e., when u_a is the agent's utility function. Similarly, $c^*(\mathbf{w}^h, \ell)$ is the optimal type for $(\mathbf{w}^h, \ell)$ under u_a. Furthermore, the expected payoffs from the passages $(c^h, \mathbf{w}^h, h)$ and $(c^*(\mathbf{w}^h, \ell), \mathbf{w}^h, \ell)$ are commonly v^* under u_a. So the intersection of $\text{IC}^h(v^*)$ and $\text{IC}^\ell(v^*)$ under u, $\mathbf{w}^h$, remains to be the intersection point under u_a.

If $B(w_g^h) = 0$ under u, however, $B(w_g^h) \neq 0$ under u_a for sufficiently small $a > 0$, as we wanted to show. To see this, we only need to focus on the terms of $B(w_g^h)$ that involve the third derivative of the utility function evaluated at c_g (because u and u_a agree up to the second derivatives at all relevant points and the third derivatives agree at all relevant points except c_g), which only appear in $\rho'(w_g^h)$ and are calculated as:

$$-\frac{p(p + (1-p)\Psi'(w_g^h))(2 - 3c_\Psi^{*\prime}(w_g^h) + c_\Psi^{*\prime}(w_g^h)^2)c_\Psi^{*\prime}(w_g^h)}{((1-p)u'(\Psi(w_g^h) - c^h)(\psi_{\mathbf{w}^h}''(w_g^h) - \Psi''(w_g^h)))^2} u'''(c_g). \tag{68}$$

Since

$$p + (1-p)\Psi'(w_g^h) = p\left(1 - \frac{u'(w_g^h - c^h)}{u'(w_b^h - c^h)}\right) > 0$$

and $c_\Psi^{*\prime}(w_g^h) < 0$ by Lemma A.1(c), the coefficient of $u'''(c^h)$ does not vanish. In addition, according to (18), (19), and (23), this coefficient is expressed only in terms of the values of the utility function and its first and second derivatives at $c = c^h$, c_g, and $w_b^h - c^h$. Hence, this coefficient assumes the same value under u and u_a and, therefore, (68) differs under u and u_a because $u'''(c_g) \neq u_a'''(c_g)$. Since $B(w_g^h)$ differs under u and u_a only in the terms involving the third derivative evaluated at c_g as argued above, if $B(w_g^h)$ vanishes under u, it does not under u_a. This establishes that the set of utility functions for which $B(w_g^h) \neq 0$ is dense, hence is a generic set (together with the earlier assertion that it is open).

Proving that (59) holds for generic u is similar but involves a modification of fourth derivative. Again, it is clear from continuity that the set of utility functions for which (59) holds is an open set. To show it is dense, let u be a function under which (59) fails, and construct u_a by

$$u_a(c) = u(c) + \frac{\Omega_a(c)}{4!}(c - c_g)^4.$$

By the same reasoning as before, the intersection of $\mathrm{IC}^h(v^*)$ and $\mathrm{IC}^\ell(v^*)$ under u, $\mathbf{w}^h$, remains the intersection point under u_a. However, (59) holds under u_a for sufficiently small $a > 0$. To see this, we only need to focus on the terms on the left-hand side of (59) that involve the fourth derivative of the utility function evaluated at c_g by an analogous reason as before, which only appear in

$$\left.\frac{d}{d\tau}\rho_\tau'(\hat{w}_g^\tau)\right|_{\tau=0}$$

and are calculated as:

$$-\frac{a_g p\big(p + (1-p)\Psi'(w_g^h)\big)\big(2 - 3c_\Psi^{*\prime}(w_g^h) + c_\Psi^{*\prime}(w_g^h)^2\big)c_\Psi^{*\prime}(w_g^h)(1 - c_\Psi^{*\prime}(w_g^h))}{((1-p)u'(\Psi(w_g^h) - c^h)(\psi_{\mathbf{w}^h}''(w_g^h) - \Psi''(w_g^h)))^2}u^{(4)}(c_g), \tag{69}$$

where $u^{(4)}(\cdot)$ is the fourth derivative of the utility function. Since

$$p + (1-p)\Psi'(w_g^h) = p\left(1 - \frac{u'(w_g^h - c^h)}{u'(w_b^h - c^h)}\right) > 0$$

and $c_\Psi^{*\prime}(w_g^h) < 0$ by Lemma A.1(c), the coefficient of $u^{(4)}(c_g)$ does not vanish. In addition, this coefficient assumes the same value under u and u_a by the same reason as before. Hence, (69) differs under u and u_a because $u^{(4)}(c_g) \neq u_a^{(4)}(c_g)$. Since the value of the left-hand side of (59) differs under u and under u_a only in the terms that involve the fourth derivative evaluated at c_g as argued above, if the left-hand side of (59) vanishes under u, it does not under u_a. This establishes that the set of utility functions for which (59) holds is dense, hence is a generic set (together with the earlier assertion that it is open). Now the proof of Theorem 2 is complete because the set of utility functions that satisfy (59) and (60) is a generic set as the intersection of two generic sets.

PROOF OF LEMMA A.6: First consider the case that a path $\{\mathbf{w}^t\}$ converges to $\hat{\mathbf{w}}$ from above. Let $\mathbf{x}^t$ and $\mathbf{y}^t$ denote the left and right contact points of $\mathbf{w}^t$. Applying Taylor's theorem to Ψ and $\psi_{\mathbf{x}^t}$ around x_g^t,

$$\Psi(w_g) = \Psi(x_g^t) + \Psi'(x_g^t)(w_g - x_g^t) + \Psi''(x_g^t)(w_g - x_g^t)^2/2 + o((w_g - x_g^t)^2) \quad \text{and}$$

$$\psi_{\mathbf{x}^t}(w_g) = \psi_{\mathbf{x}^t}(x_g^t) + \psi_{\mathbf{x}^t}'(x_g^t)(w_g - x_g^t) + \psi_{\mathbf{x}^t}''(x_g^t)(w_g - x_g^t)^2/2 + o((w_g - x_g^t)^2),$$

where $o((w_g - x_g^t)^2)$ denotes a function that is negligible relative to $(w_g - x_g^t)^2$. Due to (22),

$$\psi_{\mathbf{x}^t}(w_g) - \Psi(w_g) = \big(\psi_{\mathbf{x}^t}''(x_g^t) - \Psi''(x_g^t)\big)(w_g - x_g^t)^2/2 + o((w_g - x_g^t)^2).$$

Applying Taylor's theorem analogously to Ψ and $\psi_{\mathbf{y}^t}$ around y_g^t,

$$\psi_{\mathbf{y}^t}(w_g) - \Psi(w_g) = \big(\psi_{\mathbf{y}^t}''(y_g^t) - \Psi''(y_g^t)\big)(w_g - y_g^t)^2/2 + o((w_g - y_g^t)^2).$$

Since $\psi_{\mathbf{x}^t}(w_g) - \Psi(w_g) = \psi_{\mathbf{y}^t}(w_g) - \Psi(w_g)$ at $w_g = w_g^t$ by construction,

$$\frac{(\psi''_{\mathbf{x}^t}(x_g^t) - \Psi''(x_g^t))\lambda(t)^2}{2} + o(\lambda(t)^2) = \frac{(\psi''_{\mathbf{y}^t}(y_g^t) - \Psi''(y_g^t))\tilde{\lambda}(t)^2}{2} + o(\tilde{\lambda}(t)^2). \tag{70}$$

Since $\mathbf{x}^t, \mathbf{y}^t \to \hat{\mathbf{w}}$ as $t \to 0$, dividing both sides by $\lambda(t)^2$ leads us to conclude that

$$\left(\frac{\tilde{\lambda}(t)}{\lambda(t)}\right)^2 \to \lim_{t\to 0}\left(\frac{\psi''_{\mathbf{x}^t}(x_g^t) - \Psi''(x_g^t)}{\psi''_{\mathbf{y}^t}(y_g^t) - \Psi''(y_g^t)}\right) = 1 \quad \text{as} \quad t \to 0.$$

If the derivative $D\mathbf{w}^t|_{t=0} = (a_g, a_b) \in \mathbb{R}^2$ exists, $\mathbf{w}^t = (w_g^t, w_b^t)$ is first-order approximated by $w_z^t = \hat{w}_z + a_z t + o_z(t)$ for $z = g, b$, where $o_z(t)$ is a function negligible relative to t. Since $\psi_{\mathbf{x}^t}(w_g^t) = w_b^t = \hat{w}_b + a_b t + o_b(t)$ and $\Psi(w_g^t) = \Psi(\hat{w}_g) + \Psi'(\hat{w}_g)a_g t + o(t)$ by Taylor's theorem, $\psi_{\mathbf{x}^t}(w_g^t) - \Psi(w_g^t) = (a_b - \Psi'(\hat{w}_g)a_g)t + o(t)$. Since this is also equal to the value of (70), we deduce

$$\frac{t}{\lambda(t)^2} = \left(\frac{\psi''_{\mathbf{x}^t}(x_g^t) - \Psi''(x_g^t)}{2(a_b - \Psi'(\hat{w}_g)a_g)}\right) + \frac{o(\lambda(t)^2)}{\lambda(t)^2} + \frac{o(t)}{t}\frac{t}{\lambda(t)^2} \to \left(\frac{\psi''_{\mathbf{x}^t}(x_g^t) - \Psi''(x_g^t)}{2(a_b - \Psi'(\hat{w}_g)a_g)}\right),$$

which is a positive number due to Lemma A.2 because

$$a_b - \Psi'(\hat{w}_g)a_g = \frac{(a_g, a_b)\cdot D_{\mathbf{w}}V(\hat{\mathbf{w}}, e)}{(1-p^e)u'(\hat{w}_b - c^*_\Psi(\hat{w}_g))}$$

by (7) and (19), and $(a_g, a_b)\cdot D_{\mathbf{w}}V(\hat{\mathbf{w}}, e) > 0$ by supposition.

Next, consider the case that $\{\mathbf{w}^t\}$ is a C^2 path that converges to $\hat{\mathbf{w}}$ from below such that the derivative $D\mathbf{w}^t|_{t=0} = (a_g, a_b)$ satisfies $(a_g, a_b)\cdot D_{\mathbf{w}}V(\hat{\mathbf{w}}, e) < 0$. Note that the utility level that the $\hat{c} \equiv c^*(\hat{\mathbf{w}}, e)$-type agent derives from $\mathbf{w}^t$ is lower than v^*, which we denote by $\hat{v}^t < v^*$. Let $\varphi_{(\hat{c},t)}(w_g)$ denote the implicit function that represents $\mathrm{IC}^e(\hat{v}^t|\hat{c})$, the interim indifference curve of the $\hat{c}$-type that goes through $\mathbf{w}^t$, i.e.,

$$u(\hat{c}) + p^e u(w_g - \hat{c}) + (1-p^e)u(\varphi_{(\hat{c},t)}(w_g) - \hat{c}) - d(e) = \hat{v}^t. \tag{71}$$

By Taylor's theorem applied to Ψ and to $\varphi_{(\hat{c},t)}$ around w_g^t,

$$\begin{aligned}\Psi(w_g) &= \Psi(w_g^t) + \Psi'(w_g^t)(w_g - w_g^t) \\ &\quad + \Psi''(w_g^t)(w_g - w_g^t)^2/2 + o((w_g - w_g^t)^2) \quad \text{and} \\ \varphi_{(\hat{c},t)}(w_g) &= \varphi_{(\hat{c},t)}(w_g^t) + \varphi'_{(\hat{c},t)}(w_g^t)(w_g - w_g^t) \\ &\quad + \varphi''_{(\hat{c},t)}(w_g^t)(w_g - w_g^t)^2/2 + o((w_g - w_g^t)^2).\end{aligned}$$

By construction, $\varphi_{(\hat{c},t)}(w_g) = \Psi(w_g)$ at $w_g = x_g^t$ and y_g^t:

$$\begin{aligned}\Psi(w_g^t) - \varphi_{(\hat{c},t)}(w_g^t) &= \lambda(t)\big(\Psi'(w_g^t) - \varphi'_{(\hat{c},t)}(w_g^t)\big) \\ &\quad - \lambda(t)^2\big(\Psi''(w_g^t) - \varphi''_{(\hat{c},t)}(w_g^t)\big)/2 + o(\lambda(t)^2)\end{aligned} \tag{72}$$

and

$$\begin{aligned}\Psi(w_g^t) - \varphi_{(\hat{c},t)}(w_g^t) &= -\tilde{\lambda}(t)\big(\Psi'(w_g^t) - \varphi'_{(\hat{c},t)}(w_g^t)\big) \\ &\quad - \tilde{\lambda}(t)^2\big(\Psi''(w_g^t) - \varphi''_{(\hat{c},t)}(w_g^t)\big)/2 + o(\tilde{\lambda}(t)^2).\end{aligned} \tag{73}$$

Note also that

$$\Psi(w_g^t) - \varphi_{(\hat{c},t)}(w_g^t) = \Psi(w_g^t) - w_g^t = (\Psi'(\hat{w}_g)a_g - a_b)t + o(t). \tag{74}$$

We now show that $(\Psi'(w_g^t) - \varphi'_{(\hat{c},t)}(w_g^t))$ is first-order approximated by a linear function of t. By differentiating (71) with respect to w_g and evaluating at $w_g = w_g^t$, we get

$$\varphi'_{(\hat{c},t)}(w_g^t) = -\frac{pu'(w_g^t - \hat{c})}{(1-p)u'(w_b^t - \hat{c})}$$

and, therefore,

$$\begin{aligned}
&\frac{d}{dt}(\varphi'_{(\hat{c},t)}(w_g^t))\bigg|_{t=0}\\
&= \frac{p^e(1-p^e)\big(u''(w_g^t - \hat{c})\frac{dw_g^t}{dt}u'(w_b^t - \hat{c}) - u'(w_g^t - \hat{c})u''(w_b^t - \hat{c})\frac{dw_b^t}{dt}\big)}{((1-p^e)u'(w_b^t - \hat{c}))^2}\Bigg|_{t=0}\\
&= \frac{p^e\big(u''(\hat{w}_g - \hat{c})u'(\hat{w}_b - \hat{c})\frac{dw_g^t}{dt}\big|_{t=0} - u'(\hat{w}_g - \hat{c})u''(\hat{w}_b - \hat{c})\frac{dw_b^t}{dt}\big|_{t=0}\big)}{(1-p^e)u'(\hat{w}_b - \hat{c})^2},
\end{aligned}$$

which is a finite number because $\{\mathbf{w}^t\}$ is C^2. Since

$$\frac{d}{dt}(\Psi'(w_g^t))\bigg|_{t=0} = \frac{d}{dt}(\Psi'(\hat{w}_g + a_g t))\bigg|_{t=0} = \Psi''(\hat{w}_g)a_g,$$

it follows that

$$\frac{d}{dt}\big(\Psi'(w_g^t) - \varphi'_{(\hat{c},t)}(w_g^t)\big)\bigg|_{t=0}$$

is a constant, say θ. So, by Taylor's theorem $\Psi'(w_g^t) - \varphi'_{(\hat{c},t)}(w_g^t) = \theta t + o(t)$.

Finally, dividing the right-hand side of (72) and (74) by t, we get

$$\lambda(t)\theta + \frac{\lambda(t)^2}{2t}\big(\varphi''_{(\hat{c},t)}(w_g^t) - \Psi''(w_g^t)\big) + \frac{o(\lambda(t)^2)}{t} = \Psi'(\hat{w}_g)a_g - a_b + \frac{o(t)}{t}. \tag{75}$$

Since the right-hand side of (75) converges to $\Psi'(\hat{w}_g)a_g - a_b$ as $t \to 0$, so must the left-hand side. This is possible only if $o(\lambda(t)^2)/t \to 0$ as $t \to 0$, for otherwise the second term of the left-hand side would explode. Since $\lambda(t) \to 0$ and $\varphi''_{(\hat{c},t)}(w_g^t) \to \psi''_{(\hat{\mathbf{w}},e)}(\hat{w}_g)$ as $t \to 0$,

$$\frac{t}{\lambda(t)^2} \to -\frac{\psi''_{\hat{\mathbf{w}}}(\hat{w}_g) - \Psi''(\hat{w}_g)}{2(a_b - \Psi'(\hat{w}_g)a_g)} \quad \text{as} \quad t \to 0.$$

Since $a_b - \Psi'(\hat{w}_g)a_g < 0$ by the supposition that $(a_g, a_b) \cdot D_{\mathbf{w}}V(\hat{\mathbf{w}}, e) < 0$, this proves the second convergence of (27). It is easily verified that $t/\tilde{\lambda}(t)^2$ converges to the same limit by dividing the right-hand side of (73) and (74) by t. Hence, it follows that $\tilde{\lambda}(t)/\lambda(t) \to 1$, proving the first convergence. *Q.E.D.*

PROOF OF LEMMA A.8: We prove the claim for $f^+(\hat{w}_g)$ in the lemma. (The claim for $f^-(\hat{w}_g)$ is proved analogously.) By resetting the origin at $\hat{w}_g$, the proof amounts to showing

$$\lim_{\delta\downarrow 0}\frac{1}{\delta^2}\int_{[0,\delta]} w\,dF^h(w) = \frac{a}{2} \quad \text{if} \quad f^+(0) = \lim_{\delta\downarrow 0}\frac{F^h(\delta) - F^h(0)}{\delta} = a \in \mathbb{R}_+. \tag{76}$$

If $a = 0$, then $\lim_{\delta\downarrow 0}\frac{1}{\delta}\int_{[0,\delta]} dF^h(w) = 0$ by definition. Hence, the assertion of (76) follows because $0 \le \lim_{\delta\downarrow 0}\frac{1}{\delta^2}\int_{[0,\delta]} w dF^h(w) \le \lim_{\delta\downarrow 0}\frac{\delta}{\delta^2}\int_{[0,\delta]} dF^h(w) = 0$.

Assume $a > 0$ below. From the supposition of (76), $\forall\,\tau > 0$, $\exists\,\delta(\tau) > 0$ such that

$$a - \tau < \frac{F^h(w) - F^h(0)}{w} < a + \tau \quad \forall\, w \in (0, \delta(\tau)). \tag{77}$$

Consider n equal partitions of the interval $[0, \delta]$ by points $k\delta/n$, $k = 0, 1, \ldots, n$, for $\delta \leq \delta(\tau)$. Since

$$F^h\left(\frac{k\delta}{n}\right) - F^h(0) = F^h\left(\frac{k\delta}{n}\right) - F^h\left(\frac{(k-1)\delta}{n}\right) + F^h\left(\frac{(k-1)\delta}{n}\right) - F^h(0),$$

it follows that

$$\text{(78)} \qquad \frac{F^h(\frac{k\delta}{n}) - F^h(\frac{(k-1)\delta}{n})}{F^h(\frac{(k-1)\delta}{n}) - F^h(0)} = \frac{(F^h(\frac{k\delta}{n}) - F^h(0))/(\frac{k\delta}{n})}{(F^h(\frac{(k-1)\delta}{n}) - F^h(0))/(\frac{(k-1)\delta}{n}\frac{k}{k-1})} - 1$$
$$\in \left(\frac{(a-\tau)k}{(a+\tau)(k-1)} - 1, \frac{(a+\tau)k}{(a-\tau)(k-1)} - 1\right)$$

for $0 < \tau < a$ because

$$a - \tau < \frac{F^h(\frac{k\delta}{n}) - F^h(0)}{\frac{k\delta}{n}}, \frac{F^h(\frac{(k-1)\delta}{n}) - F^h(0)}{\frac{(k-1)\delta}{n}} < a + \tau$$

by (77).

Since the interval on the right-hand side of (78) converges to a singleton set $\{1/(k-1)\}$ as $\tau \downarrow 0$, for any $\eta > 0$, there is $\tau(\eta) > 0$ such that the ratio on the left-hand side of (78) is within η of $1/(k-1)$ if $\tau < \tau(\eta)$. In particular, there is a sufficiently small $\tau_n < 1/n^2$ such that the ratio on the left-hand side of (78) is within $1/n^2$ of $1/(k-1)$ if $\tau < \tau_n$ and $\delta < \delta(\tau)$. Multiplying the left-hand side of (78) by

$$\frac{F^h(\frac{(k-1)\delta}{n}) - F^h(0)}{\frac{(k-1)\delta}{n}}$$

which is within $1/n^2$ of a by (77), therefore, we get

$$\text{(79)} \qquad \left(a - \frac{1}{n^2}\right)\left(\frac{1}{k-1} - \frac{1}{n^2}\right) < \frac{n}{(k-1)\delta}\left(F^h\left(\frac{k\delta}{n}\right) - F^h\left(\frac{(k-1)\delta}{n}\right)\right)$$
$$< \left(a + \frac{1}{n^2}\right)\left(\frac{1}{k-1} + \frac{1}{n^2}\right)$$

for $\delta < \delta(\tau_n)$. Since this holds for every $n = 1, 2, \ldots,$

$$\begin{aligned}
\lim_{\delta\downarrow 0} \frac{1}{\delta^2}\int_{[0,\delta]} w\,dF^h(w) &\leq \lim_{n\to\infty} \sum_{k=1}^{n} \frac{k}{\delta n}\left(F^h\left(\frac{k\delta}{n}\right) - F^h\left(\frac{(k-1)\delta}{n}\right)\right) \\
&\leq \lim_{n\to\infty} \sum_{k=1}^{n} \frac{k}{n^2}\left(a + \frac{1}{n^2}\right)\left(1 + \frac{k-1}{n^2}\right) \\
&= \lim_{n\to\infty} \left(a + \frac{1}{n^2}\right) \sum_{k=1}^{n}\left(\frac{k}{n^2} + \frac{k^2 - k}{n^4}\right) \\
&= \lim_{n\to\infty} \left(a + \frac{1}{n^2}\right)\left(\frac{n(n+1)}{2n^2} + \frac{n(n+1)(2n+1)}{6n^4} - \frac{n(n+1)}{2n^4}\right) \\
&= \frac{a}{2}.
\end{aligned}$$

Obtaining the reverse inequality analogously from the first inequality of (79), we deduce that

$$\lim_{\delta\downarrow 0} \frac{1}{\delta^2}\int_{[0,\delta]} w\,dF^h(w) = \frac{a}{2}. \qquad \textit{Q.E.D.}$$

REFERENCES

ARNOTT, R., AND J. STIGLITZ (1991): "Equilibrium in Competitive Insurance Markets with Moral Hazard," NBER Working Paper No. 3588.

BISIN, A., AND D. GUAITOLI (2000): "Moral Hazard and Non-Exclusive Contracts," *RAND Journal of Economics*, forthcoming.

BISIN, A., AND A. A. RAMPINI (2002): "Exclusive Contracts and the Institution of Bankruptcy," Mimeo, New York University and Northwestern University.

BIZER, D. S., AND P. M. DEMARZO (1992): "Sequential Banking," *Journal of Political Economy*, 100, 41–61.

——— (1999): "Optimal Incentive Contracts when Agents Can Save, Borrow, and Default," *Journal of Financial Intermediation*, 8, 241–269.

BUCK, R. C. (1978): *Advanced Calculus* (Third Ed.). New York: McGraw-Hill.

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

FUDENBERG, D., B. HOLMSTROM, AND P. MILGROM (1990): "Short-Term Contracts and Long-Term Agency Relationships," *Journal of Economic Theory*, 51, 1–31.

FUDENBERG, D., AND J. TIROLE (1990): "Moral Hazard and Renegotiation in Agency Contracts," *Econometrica*, 58, 1279–1319.

KOLMOGOROV, A., AND S. FOMIN (1970): *Introductory Real Analysis*. New York: Dover Publications.

KAHN, C. M., AND D. MOOKHERJEE (1998): "Competition and Incentives with Nonexclusive Contracts," *RAND Journal of Economics*, 29, 443–465.

MA, C.-T. A. (1991): "Adverse Selection in Dynamic Moral Hazard," *Quarterly Journal of Economics*, 106, 255–275.

——— (1994): "Renegotiation and Optimality in Agency Contracts," *Review of Economic Studies*, 61, 109–129.

MALCOMSON, J., AND F. SPINNEWYN (1988): "The Multiperiod Principal-Agent Problem," *Review of Economic Studies*, 55, 391–408.

MATTHEWS, S. A. (1995): "Renegotiation of Sales Contracts," *Econometrica*, 63, 567–589.

PARK, I.-U. (2001): "Repeated Moral Hazard with Renegotiation and Free Access to Credit Markets," Mimeo, University of Pittsburgh.

PAULY, M. V. (1974): "Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection," *Quarterly Journal of Economics*, 88, 44–62.

REY, P., AND B. SALANIE (1990): "Long-term, Short-term and Renegotiation: On the Value of Commitment in Contracting," *Econometrica*, 58, 597–619.

ROCHET, J.-C., AND L. STOLE (2003): "The Economics of Multidimensional Screening," in *Advances in Economics and Econometrics*: *Theory and Applications*. *Eighth Wold Congress*, *Volume I*, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky. Cambridge, U.K.: Cambridge University Press, 150–197.

SPIVAK, M. (1965): *Calculus on Manifolds.* New York: W. A. Benjamin, Inc.

——— (1979): *A Comprehensive Introduction to Differential Geometry*, *Volume I* (Second Ed.). Wilmington, DE: Publish or Perish, Inc.