WORKING PAPERS

2023-06

01 May 2023

The Principal-Agent Model under Smooth Ambiguity

My Dam and Yacine Chitour

CEPS
Centre for Economics at Paris-Saclay

école normale supérieure paris–saclay
université evry val-d'essonne
université PARIS-SACLAY
université PARIS-SACLAY
Maison des Sciences de l'Homme PARIS-SACLAY

The principal-agent model under smooth ambiguity

My Dam* Yacine Chitour† François Pannequin‡

May 15, 2019

Abstract

We characterize the symmetric information benchmark for the principal-agent model under ambiguity in the sense of Klibanoff et al. (2005). We recast the problem in an optimal control framework and proved the existence of an optimal wage function under a standard set of assumptions. When the principal is risk-averse, we show that optimal wage is robust to ambiguity in the sense that it is nondecreasing in outcomes regardless of the attitudes of the contracting parties towards ambiguity and the number of ambiguous states under consideration. When the principal is risk-neutral and there are only two ambiguous states, we prove that unambiguous optimal wage is robust to ambiguity if ambiguity has an one-sided structure, namely that if ambiguity "contaminates" either the lower or the higher range of outcomes, but not both.

Keywords: principal-agent, ambiguity aversion, optimal contract.
JEL classification: D86, D82.

*UPSaclay, UEVE, EPEE. Email: thitramy.dam@univ-evry.fr
†Université Paris-Sud. Email: yacine.chitour@l2s.centralesupelec.fr
‡UPSaclay, ENS Paris-Saclay. Email: panneq@ens-paris-saclay.fr

Contents

Chapter nomenclature

Notation	Meaning	Reference page(s)
$\mathcal{I}$	Index set of (second-order) states	4
p_i	Prior on state $i \in \mathcal{I}$	4
$\tilde{x}$	Outcome random variable	4
e	Effort level	4
I_e	Domain of effort	4
$f_i(x\|e)$	State-i effort-e conditional density	4
$F_i(x\|e)$	State-i effort-e conditional density	4
I_x	Common support of the conditional densities	4
$c(\cdot)$	Cost of exerting effort	6
$w(\cdot)$	Wage (the control function)	6
$\ell_{ij}(\cdot\|e)$	Effort-conditional likelihood ratio	5
$v(\cdot)$	Principal's utility function	6
$\phi_P(\cdot)$	Principal's second-order utility (welfare) function	6
$u(\cdot)$	Agent's utility function	6
$\phi_A(\cdot)$	Agent's second-order utility (welfare) function	6

Table 1: Notations used in the paper

1 Assumptions and statement of the problem

Consider a principal-agent model where decision makers (DMs), the agent and the principal in this case, face ambiguity in the distribution of the states. Consequently, the distributions of outcomes are state-conditional. We assume that the state space is finite, and the DMs have common priors over the distribution of the states. Our objective is to determine an optimal wage contract under symmetric information.

Notation 1. *Throughout this paper, the subscripts A and P refer to the agent and the principal, respectively.*

Notation 2. *The n-dimensional Euclidean space whose elements have all non-negative coordinates is denoted by $\mathbb{R}^n_+$. The n-dimensional Euclidean space whose elements have all strictly positive coordinates are denoted by $\mathbb{R}^n_{++}$*

Assumption 1. *Let the state space be $\mathcal{I} = \{1, 2, \dots, n\}$, where $n < +\infty$. Let p_i denotes the prior belief of both DMs regarding the likelihood of state i occurring. Assume that $p_i \in (0, 1)$ for all $i \in \mathcal{I}$ and $\sum_{i=1}^n p_i = 1$*

Let $e \in I_e = [\underline{e}, \bar{e}] \subset \mathbb{R}_+$ be the level of effort/action to be implemented by the agent. The principal has to determine the desirable level of effort she demands of the agent. The effort exerted by the agent is assumed to be verifiable and legally enforceable. In other words, it is a valid contracting variable. Conditional on e, the outcome is assumed to be a continuous random variable $\tilde{x}$ whose state-conditional distributions have common support $I_x = [0, \bar{x}]$. In particular, the following assumption holds.

Assumption 2. *For each $e \in I_e$ and $i \in \mathcal{I}$, let $F_i(\cdot|e) : I_x \to [0, 1]$ be the conditional cumulative distribution function (cdf) of $\tilde{x}$ defined by*

$$F_i(x|e) = \Pr(\tilde{x} \le x \mid e), \quad i \in \mathcal{I}.$$

Assume that all the cdfs are C^2 on their common support I_x. Let $f_i(\cdot|e) : I_x \to \mathbb{R}_{++}$ be the conditional probability density function (pdf) associated with $F_i(\cdot|e)$ defined by

$$f_i(x|e) = \frac{\partial F_i(x|e)}{\partial x}, \quad i \in \mathcal{I}.$$

Then $\int_{I_x} f_i(x|e)dx = 1$ for each $i \in \mathcal{I}$, and $f(x|e) = (f_i(x|e))_{i\in\mathcal{I}} \in \mathbb{R}^n_{++}$ for all $x \in I_x$.

We assume that ambiguous states can be ranked according to the likelihood ratio (LR), as next defined.

Assumption 3. *For two distinct indexes i, j in $\mathcal{I}$, let $\ell_{ij}(\cdot|e) : I_x \to \mathbb{R}_+$ be the effort-conditional likelihood ratio defined by $\ell_{ij}(x|e) = \frac{f_i(x|e)}{f_j(x|e)}$. Then state i is said to dominate state j in the sense of likelihood ratio dominance (LRD) if $\frac{\partial \ell_{ij}(x|e)}{\partial x} = \ell'_{ij}(x|e) \geq 0$ for a.e. $x \in I_x$, with strict inequality in some subset of positive measure of I_x.*[1]

Example 1. *Suppose $n = 2$, $\bar{x} = 1$ (so that any outcome is viewed as a fraction of the maximum outcome), and the outcome distribution follows a truncated exponential distribution with an ambiguous parameter.*[2] *In particular:*

$$f_i(x|e) = \begin{cases} \alpha_i(e)\beta_i(e)\exp\{-\beta_i(e)x\} & x \in [0,1],\ \beta_i(e) > 0,\ i \in \{1,2\} \\ 0 & o.w. \end{cases},$$

where $\alpha_i(e) = \frac{\exp(\beta_i(e))}{\exp(\beta_i(e))-1} > 1$. The likelihood ratio $\ell_{12}(x)$ is increasing if and only if

$$\begin{aligned} [\beta_2(e) - \beta_1(e)]\,\alpha_1(e)\alpha_2(e)\exp\{-(\beta_i(e)+\beta_j(e))x\} &\geq 0 \\ \iff \beta_2(e) - \beta_1(e) &\geq 0. \end{aligned}$$

Thus an amelioration in the LRD sense is equivalent to a reduction in the parameter of the exponential distribution. In other words, the more favorable state (state 1) is associated with a smaller parameter. At this point, we have not explicitly specified how effort changes this parameter. In general, this relationship also state-conditional. One might hypothesize that the higher the level of effort, the smaller the gap $\delta(e) = \beta_2(e) - \beta_1(e)$. Intuitively, this gap represents the severity, or the consequence of ambiguity. Letting $\delta(e)$ decrease in e means believing that high efforts can mitigate the severity of ambiguity.

Next, we hypothesize that in any given state i, raising efforts *improves* the outcome distribution in following sense.

Assumption 4. *Let the function $\ell_i(\cdot) : I_x \to \mathbb{R}_+$ be the state conditional likelihood ratio defined by $\ell_i(x) := \frac{f_i(x|e_2)}{f_i(x|e_1)}$, $e_1 < e_2 \in I_e$. Then $\ell'_i(x) \geq 0$ a.e. $x \in I_x$, with strict inequality in some subset of positive measure of I_x.*

[1] Note that LRD is a special case of FSD. Thus $F_i(\cdot)$ dominates $F_j(\cdot)$ in the sense of LRD implies $F_i(x) \leq F_j(x)$ for all $x \in I_x$, with strict inequality on some subset of I_x of positive measure. See Wolfstetter (1999) for further discussion.

[2] Recall that if X is distributed as an exponential distribution of parameter β, its density is:

$$f(x) = \beta\exp\{-\beta x\}, \quad x \geq 0,\ \beta > 0.$$

For our purpose, we need to "redistribute" the mass over a bounded interval I_x, instead of the whole $\mathbb{R}_+$. This conditioning is achieved by dividing the original density by the cumulative mass contained in this interval, which in this example is $F(1) = \int_0^1 f(x)dx = 1 - \exp(-\beta)$.

Assumption 5. *The wage is a measurable function $w : I_x \to \mathbb{R}_+$ satisfying $w(x) \in [0, x]$ for a.e. $x \in I_x$.*[3]

Assumption 6. *The cost of effort is a C^2 function $c : I_e \to \mathbb{R}_+$ satisfying $c(0) = 0$, $c' > 0$, $c'' \geq 0$.*

We model the DMs' attitude towards risk by the von Neumann-Morgenstern utility functions. Recall that the utility function being concave, linear, or convex corresponds to a risk-averse, risk-neutral, or risk-seeking DM, respectively. Typically, the agent is assumed to be risk-averse and the principal risk-neutral. For greater generality, we allow for the possibility of the principal being risk-averse.

Assumption 7. *The agent has utility function $u : \mathbb{R}_+ \to \mathbb{R}$, which is at least C^2, satisfying $u(0) = 0$, $u' > 0$, $u'' < 0$, and the Inada condition $\lim_{w \to 0} u'(w) = +\infty$. Similarly, the principal also has a C^2 utility function $v : \mathbb{R}_+ \to \mathbb{R}$ satisfying $v(0) = 0$, $v' > 0$ and $v'' \leq 0$. If $v'' < 0$, then $\lim_{w \to 0} v'(w) = +\infty$.*

To capture the phenomenon known as "ambiguity aversion" postulated by Ellsberg (1961), we follow the smooth model of Klibanoff et al. (2005). We refer to this model as KMM (2005) from now on. According to KMM (2005), attitudes towards ambiguity can be modeled by a functional ϕ_J, which is referred to as the welfare functional throughout this paper. In particular, the welfare functional being concave, linear, or convex corresponds to a DM who is ambiguity-averse, ambiguity-neutral or ambiguity-seeking, respectively. The DMs are assumed to be either ambiguity-averse or ambiguity-neutral.

Assumption 8. *Let the welfare functional be $\phi_J : \mathcal{V} \to \mathbb{R}$, where $\mathcal{V}$ is the range of J's utility function, for $J \in \{A, P\}$. For each $J \in \{A, P\}$, assume that ϕ_J is at least C^2 on its domain, satisfying $\phi_J' > 0$ and $\phi_J'' \leq 0$.*

Example 2. *Following the empirical work of Chakravarty and Roy (2009) and more recently of Berger and Bosetti (2016), we can let $\phi_J(u) = u^{1-\sigma}$, $u \geq 0$ where $\sigma \in [0, 1)$ represents the degree of relative ambiguity aversion (RAA) and $\sigma = 0$ corresponds to J being neutral to ambiguity.*

The agent extracts satisfaction from wage and dissatisfaction from exerting efforts (there is no utility coming from work other than that from the payment). The principal, on the other hand, cares only about profits (outcomes net of compensation to the agent). Assuming that the cost of effort and the welfare of the agent are separable, the principal's problem in absence of moral hazard is:

[3] This assumption deviates from the mainstream of the existing literature that considers a global, rather than a point-wise constraint on $w(\cdot)$. We shall also consider the implication of this assumption in a separate later section.

$$
\begin{aligned}
&\max_{w(.),e} \sum_{i=1}^{n} p_i \phi_P \Big(\int_0^{\bar{x}} (x - w(x))\, f_i(x|e) dx \Big) \\
&\qquad s.t. \\
&w(x) \in [0, x] \quad \forall x \in I_x, \\
&e \in I_e, \\
&\sum_{i=1}^{n} p_i \phi_A \left(\int_0^{\bar{x}} u(w(x)) f_i(x|e) dx \right) - c(e) \geq \bar{U},
\end{aligned} \tag{1}
$$

where $\bar{U} \in \mathbb{R}$ is the reservation welfare of the agent (representing her outside option), $\bar{U} \geq \phi_A(0) \equiv \bar{\phi}_A \in \mathbb{R}$.

2 Formulation of the optimal control problem

In this section, we reformulate the optimization problem of the principal-agent model in the form of an optimal control problem (OCP) following Trélat (2008). To this end let the state vector $X = (z, y, e) \in \mathcal{X} = \mathbb{R}^n_+ \times \mathbb{R}^n_+ \times I_e$ be defined as follows:

$$
\dot{X} = \begin{pmatrix} \dot{z} \\ \dot{y} \\ \dot{e} \end{pmatrix} = \begin{pmatrix} v\,(x - w(x))\, f(x|e) \\ u\,(w(x))\, f(x|e) \\ 0 \end{pmatrix}, \quad X(0) = \begin{pmatrix} z(0) = 0 \\ y(0) = 0, \\ e(0) = e \in I_e \end{pmatrix}, \tag{2}
$$

where $f(x|e) = (f_i(x|e))_{i \in \mathcal{I}} \in \mathbb{R}^n_{++}$ and $I_e = [\underline{e}, \bar{e}] \subset \mathbb{R}_+$.

Let $M(I_x) = \{w : I_x \to I_x \text{ measurable}\}$ be the set of measurable controls, and $\mathcal{U}$ be the set of admissible controls defined by [4]

$$
\mathcal{U} = \{w \in M(I_x) \mid w(x) \in [0, x] \text{ a.e. } x \in I_x\}.
$$

Lemma 1. *The set $\mathcal{U}$ is compact with respect to the weak-$\star$ topology.*

Proof. We want to show that every sequence w^k in $\mathcal{U}$ weak-$\star$ converges to $\bar{w}$ in $\mathcal{U}$, up to a subsequence. Arguing by contradiction, we suppose that $\bar{w} \notin \mathcal{U}$ i.e., there exists a measurable set $J \subset I_x$ of positive measure such that $\bar{w}(x) > x$ for all $x \in J$. Let $\chi_J : I_x \to I_x$ be a characteristic function defined by

$$
\chi_J(x) = \begin{cases} 1 & x \in J \\ 0 & x \notin J \end{cases}.
$$

[4]Specifically, let σ_I be a $\sigma-algebra$ on I_x, then(I_x, σ_I) is a measurable space. The function $w : I_x \to I_x$ is called measurable if, for all I in σ_I, the preimage of I under w is also in σ_I where the preimage of I under w is the set $\text{preim}_w(I) = \{x \in I_x | w(x) \in I\}$.

By assumption of weak-$\star$ convergence, we have, as k tends to infinity:

$$\begin{aligned}
\int_{I_x} \chi_J(x) w^k(x) dx \quad &\longrightarrow \quad \int_{I_x} \chi_J(x) \bar{w}(x) dx \\
\iff \int_{I_x} \chi_J(x) \left(w^k(x) - x\right) dx \quad &\longrightarrow \quad \int_{I_x} \chi_J(x) \left(\bar{w}(x) - x\right) dx \\
\iff \int_J \left(w^k(x) - x\right) dx \quad &\longrightarrow \quad \int_J \left(\bar{w}(x) - x\right) dx, \qquad (3)
\end{aligned}$$

which yields a contradiction since the RHS of (3) is strictly positive by hypothesis, while the LHS is negative by construction. Thus there exists no such set J, implying that $\bar{w}$ is in $\mathcal{U}$, completing the proof. ■

We now proceed to define the OCP. To this end, let the cost functional be:

$$g(w, e) = -\sum_{i=1}^{n} p_i \phi_P(z_i(\bar{x})), \qquad (4)$$

which is just minus the welfare functional of the principal, and the net welfare functional of the agent be:

$$h(w, e) = \sum_{i=1}^{n} p_i \phi_A \left(y_i(\bar{x})\right) - c(e) - \bar{U}. \qquad (5)$$

Under the new notation, the equivalent statement of the original optimization problem (1) is

$$\begin{aligned}
&\min_{w(\cdot), e} g(w, e) \\
&\quad s.t. \qquad\qquad (6) \\
&h(w, e) \geq 0.
\end{aligned}$$

We shall refer to this problem as the OCP in the sequel.

Lemma 2. *The constraint is active at an optimum.*

Proof. If it is not the case and $w(\cdot)$ is optimal, then since ϕ_A is continuous, there exists $\epsilon > 0$ such that

$$\begin{aligned}
\sum_{i=1}^{n} p_i \phi_A \left(\int_0^{\bar{x}} u(w(x)) f_i(x|e) dx - \epsilon \right) - c(e) \quad &> \quad \bar{U} \\
\iff \sum_{i=1}^{n} p_i \phi_A \left(\int_{I_x} \left(u(w(x)) - \epsilon\right) f_i(x|e) dx \right) - c(e) \quad &> \quad \bar{U}.
\end{aligned}$$

Since all the prior densities f_i are strictly positive and $u(\cdot)$ is continuous, there exist a subset $K \subset I_x$ of positive measure and some sufficiently small number $\epsilon_K > 0$ satisfying $w(x) - \epsilon_K \geq 0$ for all $x \in K$ such that:

$$\sum_{i=1}^{n} p_i \phi_A \left(\int_K u(w(x)) - \epsilon_K) f_i(x|e) dx + \int_{I_x \setminus K} u(w(x)) f_i(x|e) dx \right) \geq c(e) + \bar{U}.$$

Define the function $\tilde{w} : I_x \to I_x$ by $\tilde{w}(x) = \begin{cases} w(x) - \epsilon_K & x \in K \\ w(x) & x \in I_x \setminus K \end{cases}$. Then $\tilde{w}(\cdot)$ is both admissible and satisfies the constraint by construction. In addition, since $v(\cdot)$ is also continuous and strictly increasing, and all the prior densities are strictly positive, we must have

$$\int_K v(\tilde{w}(x)) f_i(x|e) dx > \int_K v(w(x)) f_i(x|e) dx, \quad \forall i \in \mathcal{I}. \tag{7}$$

Finally, since $p_i > 0$ for all $i \in \mathcal{I}$ and $\phi_P(\cdot)$ is strictly increasing:

$$\sum_{i=1}^{n} p_i \phi_P \left(\int_{I_x} v(x - \tilde{w}(x)) f_i(x|e) dx \right) > \sum_{i=1}^{n} p_i \phi_P \left(\int_0^{\bar{x}} (x - w(x)) f_i(x|e) dx \right),$$

implying that $w(\cdot)$ is not optimal, a contradiction. Hence if $w(\cdot)$ is optimal, we must have:

$$\sum_{i=1}^{n} p_i \phi_A \left(\int_0^{\bar{x}} u(w(x)) f_i(x|e) dx \right) - c(e) = \bar{U}.$$

■

3 Existence of optimal wage

Remark 1. *We discuss briefly the two extreme cases where one of the state variables might have zero value.*

Recall that $y_i(\bar{x}) = \int_0^{\bar{x}} u(w(x)) f_i(x|e) dx$ *where* $f_i > 0$ *for all* $i \in \mathcal{I}$*. Thus, if there exists* $i_* \in \mathcal{I}$ *such that* $y_{i_*}(\bar{x}) = 0$ *then* $u(w(x)) = 0$ *for a.e.* $x \in I_x$*. But this implies* $y_i(\bar{x}) = 0$ *for all* $i \in \mathcal{I}$*. Since* u *is strictly increasing and* $u(0) = 0$*, we must have in this case* $w(x) = 0$ *for a.e.* $x \in I_x$*, which is obviously the wage schedule that costs the least to the principal. But this wage satisfies the participation constraint if and only if* $\bar{\phi}_A - c(e) \geq \bar{U} \iff c(e) = 0$ *since* $\bar{U} \geq \bar{\phi}_A$ *by assumption. In sum,* $y_i(\bar{x}) = 0$ *for some* $i \in \mathcal{I}$ *if and only if* $c(e) = 0 \iff e = 0$*, which is not an economically interesting case. If* $e > 0$*, uniformly zero wage does not satisfy the participation constraint and thus is not admissible.*

By the same reasoning, $z_i(\bar{x}) = 0$ *for some* $i \in \mathcal{I}$ *if and only if* $w(x) = x$ *for a.e.* $x \in I_x$*, which in turn implies* $z_i(\bar{x}) = 0$ *for all* $i \in \mathcal{I}$*. This is the most* *expensive wage to implement for the principal. If this wage satisfies the participation constraint with strict inequality, then by the same argument made under the proof of Lemma 2, it is not optimal. On the other hand, if this wage satisfies the participation constraint with equality, then it is the only admissible candidate for a solution, and thus is trivially optimal.*

Both of these cases are discussed for technical reasons but are not very interesting economically. For this reason, we shall assume in the sequel that $y(\bar{x}) \in \mathbb{R}^n_{++}$ and $z(\bar{x}) \in \mathbb{R}^n_{++}$.

Let $\mathcal{M}_0$ and $\mathcal{M}_1$ be measurable subsets of the state space $\mathcal{X}$ defined as

$$\begin{aligned} \mathcal{M}_0 &= \{0\} \times \{0\} \times I_e, \quad (8)\\ \mathcal{M}_1 &= \mathbb{R}^n_{++} \times S_{y,e}, \quad (9) \end{aligned}$$

where $S_{y,e} = \{y \in \mathbb{R}^n_{++} \times I_e \mid -h(w,e) \leq 0\}$.

Our objective is to find a trajectory $X(\cdot)$ defined on I_x which solves (2) and corresponds to an admissible control $w \in \mathcal{U}$ satisfying

$$X(0) \in \mathcal{M}_0,\ X(\bar{x}) \in \mathcal{M}_1,$$

such that the cost functional is minimized over all possible trajectories $X(\cdot)$ linking $\mathcal{M}_0$ to $\mathcal{M}_1$.

Proposition 1. *The OCP admits an optimal control.*

Proof. Let $\delta = \inf_{(w,e)\in\mathcal{U}\times I_e} g(w,e)$. Consider a sequence of trajectories $\{X^k(\cdot)\}_{k\in\mathbb{N}}$ associated with the sequence of admissible controls $\{w^k(\cdot)\}_{k\in\mathbb{N}}$ defined by

$$X^k(x) = \begin{pmatrix} z^k(x) \\ y^k(x) \\ e^k \end{pmatrix} = \begin{pmatrix} \left(z_i^k(x)\right)_{i\in\mathcal{I}} \\ \left(y_i^k(x)\right)_{i\in\mathcal{I}} \\ e^k \end{pmatrix}, \quad x \in I_x, \quad (10)$$

such that $g(w^k, e^k) \longrightarrow \delta$ as $k \longrightarrow \infty$, where

$$\begin{aligned} z_i^k(x) &= \int_0^x v(t - w^k(t)) f_i(t|e^k)dt, \quad \forall i \in \mathcal{I}, \\ y_i^k(x) &= \int_0^x u(w^k(t)) f_i(t|e^k)dt, \quad \forall i \in \mathcal{I}. \end{aligned}$$

By the weak-$\star$ compactness of $\mathcal{U}$, the sequence $\{w^k(\cdot)\}_{k\in\mathbb{N}}$ weak-$\star$ converges to $\bar{w}(\cdot) \in \mathcal{U}$ up to some subsequence, i.e. $w^k(\cdot) \rightharpoonup \bar{w}(\cdot)$. By the compactness of I_e, the sequence $\{e^k\}_{k\in\mathbb{N}}$ converges to $e_* \in I_e$, up to some subsequence. Denote the limiting trajectory as:

$$\bar{X}(x) = \begin{pmatrix} \bar{z}(x) \\ \bar{y}(x) \\ e_* \end{pmatrix} = \begin{pmatrix} (\bar{z}_i(x))_{i\in\mathcal{I}} \\ (\bar{y}_i(x))_{i\in\mathcal{I}} \\ e_* \end{pmatrix}, \quad x \in I_x \quad (11)$$

where

$$\bar{z}_i(x) = \int_0^x v(t - \bar{w}(t)) f_i(t|e_*) dt, \quad i \in \mathcal{I},$$
$$\bar{y}_i(x) = \int_0^x u(\bar{w}(t)) f_i(t|e_*) dt, \quad i \in \mathcal{I}.$$

We first show that this limiting trajectory also brings the system from $\mathcal{M}_0$ to $\mathcal{M}_1$, which is completed by Lemma 3. Second, we show that the control associated with this limiting trajectory is optimal, which is the result of Lemma 4.

Lemma 3. *We have* $h(\bar{w}, e_*) \geq 0$.

Proof. By construction, $h(w^k, e^k) \equiv \sum_{i=1}^n p_i \phi_A(y_i^k(\bar{x})) - c(e^k) - \bar{U} \geq 0$ for all $k \in \mathbb{N}$. For $i \in \mathcal{I}$ and $k \geq 0$, let us write

$$y_i^k(\bar{x}) = \int_0^{\bar{x}} u(w^k(t)) f_i(t|e_*) dt + \Delta_i^k, \tag{12}$$

where $\Delta_i^k = \int_0^{\bar{x}} u(w^k(t)) \left(f_i(t|e^k) - f_i(t|e_*)\right) dt$. It is immediate to see that Δ_i^k tends to zero as k tends to infinity since u is bounded and f_i defined on the compact set $I_x \times I_e$ is uniformly continuous.

From Lee and Markus (1967), we have, for a convex function Γ_i:

$$\int_{I_x} \Gamma_i(\bar{w}(x)) dx \leq \liminf \int_{I_x} \Gamma(w^k(x)) dx. \tag{13}$$

Let $\Gamma_i(w^k(x)) \equiv -u(w^k(x)) f_i(x|e_*)$. Then Γ_i is convex with respect to w^k since u is concave and f_i is positive and does not depend on w^k [5]. We can rewrite (13) as:

$$\begin{aligned} -\bar{y}_i(\bar{x}) &\leq \liminf(-y_i^k(\bar{x}) - \Delta_i^k) \\ \iff \bar{y}_i(\bar{x}) &\geq \limsup y_i^k(\bar{x}). \end{aligned} \tag{14}$$

By the continuity of $\bar{y}_i$, for all $\epsilon > 0$, there exists a sufficiently large positive integer k such that $\bar{y}_i(\bar{x}) \geq y_i^k(\bar{x}) - \epsilon$. Since ϕ_A is increasing,

$$\phi_A(\bar{y}_i(\bar{x})) \geq \phi_A(y_i^k(\bar{x}) - \epsilon). \tag{15}$$

By the first fundamental theorem of calculus,

$$\phi_A\left(y_i^k(\bar{x})\right) - \phi_A\left(y_i^k(\bar{x}) - \epsilon\right) = \int_a^b \phi_A'(\zeta) d\zeta,$$

[5] In particular, let w_1, w_2 be arbitrary functions in $\mathcal{U}$. For any $\alpha \in [0,1]$, we have by the concavity of u that $u(\alpha w_1(x) + (1-\alpha) w_2(x)) \geq \alpha u(w_1(x)) + (1-\alpha) u(w_2(x))$. Multiplying both sides of the inequality by $f_i(x|e^k)$ which is positive and does not depend on w, we see that $u(w(x)) f_i(x|e^k)$ is concave.

where $b = y_i^k(\bar{x})$ and $a = b - \epsilon$. Let $M \in \mathbb{R}_+$ be an upper bound of ϕ'_A over $[a, b]$. Then $\phi_A\left(y_i^k(\bar{x})\right) - \phi_A\left(y_i^k(\bar{x}) - \epsilon\right) \leq M(b - a) = M\epsilon$, implying that $\phi_A\left(y_i^k(\bar{x}) - \epsilon\right) \geq \phi_A\left(y_i^k(\bar{x})\right) - M\epsilon$, which together with (15) implies

$$\begin{aligned} \phi_A(\bar{y}_i(\bar{x})) &\geq \phi_A\left(y_i^k(\bar{x})\right) - M\epsilon \\ \implies \sum_{i=1}^n p_i\phi_A(\bar{y}_i(\bar{x})) &\geq \sum_{i=1}^n p_i\phi_A\left(y_i^k(\bar{x})\right) - M\epsilon. \end{aligned} \tag{16}$$

Since $\phi_A(y_i^k(x)) \geq c(e^k) + \bar{U} \geq c(e_*) + \bar{U} - \epsilon$ for k large enough, from (16) we have $\sum_{i=1}^n p_i\phi_A(\bar{y}_i(\bar{x})) \geq c(e_*) + \bar{U} - (M + 1)\epsilon$. Finally, since ϵ was arbitrary, letting $\epsilon \to 0$ yields

$$\sum_{i=1}^n p_i\phi_A(\bar{y}_i(\bar{x})) \geq c(e_*) + \bar{U},$$

or equivalently, $h(\bar{w}, e_*) \geq 0$. ■

Next, we show that the sequence of welfare functional associated with $\left\{w^k(\cdot)\right\}_k$ converges to the minimal cost.

Lemma 4. *Let $\left\{g(w^k, e^k)\right\}_{k\in\mathbb{N}}$ be the sequence of cost functional defined by*

$$g(w^k, e^k) = -\sum_{i=1}^n p_i\phi_P\left(z_i^k(\bar{x})\right), \quad k \in \mathbb{N},$$

and let the cost at the limiting control $\bar{w}$ be

$$g(\bar{w}, e_*) = -\sum_{i=1}^n p_i\phi_P\left(\bar{z}_i(\bar{x})\right).$$

Then $\bar{w}$ is optimal, i.e., $g(\bar{w}, e_*) \leq \delta$.

Proof. For $i \in \mathcal{I}$ and $k \geq 0$, we write

$$z_i^k(\bar{x}) = \int_0^{\bar{x}} v(t - w^k(t))f_i(t|e_*)dt + \nu_i^k, \tag{17}$$

where $\nu_i^k = \int_0^{\bar{x}} v(t - w^k(t))\left(f_i(t|e^k) - f_i(t|e_*)\right)dt$. It is immediate to see that ν_i^k tends to zero as k tends to infinity since v is bounded and f_i defined on the compact set $I_x \times I_e$ is uniformly continuous.

We then invoke the same argument as in the proof of Lemma 3 for the convex function $\Gamma_i \equiv v(x - w(x))f_i(x|e)$, for each $i \in \mathcal{I}$. Let $M \in \mathbb{R}_+$ be an upper bound for ϕ'_P, then for sufficiently large $k \in \mathbb{N}$ and sufficiently small $\epsilon > 0$,

$$g(\bar{w}, e_*) \leq -\sum_{i=1}^n p_i\phi_P\left(z_i^k(\bar{x})\right) + M\epsilon,$$

which is the analogy of (16) in this case. Again, letting $\epsilon \to 0$ yields:

$$g(\bar{w}, e_*) \leq -\sum_{i=1}^{n} p_i \phi_P \left(z_i^k(\bar{x}) \right), \tag{18}$$

implying that $g(\bar{w}, e_*)$ is a lower bound for $-\sum_{i=1}^{n} p_i \phi_P \left(z_i^k(\bar{x}) \right)$. Thus $g(\bar{w}, e_*) \leq \delta$ by the definition of the infimum. ■

To sum up, we have proved that the limiting trajectory satisfies the constraint, hence it brings the system from $\mathcal{M}_0$ to $\mathcal{M}_1$. Moreover, the cost achieved by this trajectory is the minimum cost. Thus the limiting control $\bar{w}$ is an optimal control. ■

4 Characterization of the optimal wage

We employ the Pontryagin Maximum Principle (PMP) to characterize the necessary conditions that must be satisfied by a solution to the OCP, which has been shown to exist in Proposition 1. With a slight modification from Trélat (2008), the statement of the PMP applied to this problem is the following.

Theorem 5. ***[Pontryagin Maximum Principle]*** *Suppose (X, w) is a solution to the OCP. There exists an absolutely continuous vector-valued function $\lambda : I_x \to \mathbb{R}^{2n+1}$ and a real number $\lambda_0 \in \{0, 1\}$ with $(\lambda, \lambda_0) \neq 0 \in \mathbb{R}^{2n+2}$ such that:*

1. *λ satisfies the canonical equations*

$$\begin{aligned} \dot{X}(x) &= \nabla_\lambda H \left(X(x), w(x), \lambda(x), \lambda_0, x \right), & (19) \\ \dot{\lambda}(x) &= -\nabla_X H(X(x), w(x), \lambda(x), \lambda_0, x), & (20) \end{aligned}$$

for almost every $x \in I_x$, where the real-valued function $H : \mathbb{R}^{2n+1} \times \mathbb{R} \times \mathbb{R}^{2n+1} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, called the Hamiltonian associated with the OCP is defined by:

$$H(X, \lambda, \lambda_0, \omega, x) = v(x - \omega) \langle \lambda_z, f(x|e) \rangle + u(\omega) \langle \lambda_y, f(x|e) \rangle, \tag{21}$$

where $\lambda = (\lambda_z, \lambda_y, \lambda_e)^T \in \mathbb{R}^{2n+1}$ is called the adjoint vector whose components $\lambda_z \in \mathbb{R}^n$, $\lambda_y \in \mathbb{R}^n$ and $\lambda_e \in \mathbb{R}$ themselves are the adjoint vectors associated with the state variables z, y and e respectively.

2. *The maximum condition*

$$H \left(X(x), w(x), \lambda(x), \lambda_0, x \right) = \max_{\omega \in [0,x]} H \left(X(x), \omega, \lambda(x), \lambda_0, x \right) \tag{22}$$

is satisfied for almost every $x \in I_x$.

3. *The transversality conditions (TCs)*

$$\lambda(0) \in N_{\mathcal{M}_0}(X(0)), \tag{23}$$
$$-\lambda_0 \nabla_X g(\bar{x}, X(\bar{x})) - \lambda(\bar{x}) \in N_{\mathcal{M}_1}(X(\bar{x})) \tag{24}$$

are satisfied, where $\mathcal{M}_0$ *and* $\mathcal{M}_1$ *are respectively defined by* (8) *and* (9)*, and* $N_{\mathcal{M}_i}(X(x))$ *denotes the normal cone to* $\mathcal{M}_i$ *at* $X(x)$*,* $i \in \{0, 1\}$*.*

We start by computing the normal cones to $\mathcal{M}_0$ and $\mathcal{M}_1$.

First, consider $N_{\mathcal{M}_0}(X(0))$. Recall that $X(0) = (0, 0, e)$. Let $\xi = (\xi_z, \xi_y, \xi_e) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$. Take an element $M_0 \in \mathcal{M}_0$, then $M_0 = (0, 0, a) \in \mathbb{R}^n \times \mathbb{R}^n \times I_e$. The normal cone to $\mathcal{M}_0$ at $X(0)$ can be written as:

$$N_{\mathcal{M}_0}(X(0)) = \left\{\xi \in \mathbb{R}^{2n+1} \mid \langle \xi, M_0 - X(0)\rangle \leq 0,\ \forall M_0 \in \mathcal{M}_0\right\}$$
$$\Longrightarrow N_{\mathcal{M}_0}(X(0)) = \left\{\xi \in \mathbb{R}^{2n+1} \mid \xi_e(a - e) \leq 0,\ \forall a \in I_e\right\}.$$

One of the following scenarios can occur.

- If $e \in (\underline{e}, \bar{e})$, then $\xi_e = 0$ since $\xi_e(a - e) \leq 0$ must be satisfied for any $a \in I_e$. Thus the normal cone is:

$$N_{\mathcal{M}_0}(X(0)) = \mathbb{R}^n \times \mathbb{R}^n \times \{0\}. \tag{25}$$

- If $e = \underline{e}$ then $\xi_e(a - \underline{e}) \leq 0$ if and only if $\xi_e \leq 0$ since $a \geq \underline{e}$ for all $a \in I_e$. Thus the normal cone is

$$N_{\mathcal{M}_0}(X(0)) = \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_-. \tag{26}$$

- If $e = \bar{e}$ then $\xi_e(a - \bar{e}) \leq 0$ if and only if $\xi_e \geq 0$ since $a \leq \bar{e}$ for all $a \in I_e$. Thus the normal cone in this case is

$$N_{\mathcal{M}_0}(X(0)) = \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+. \tag{27}$$

Next, consider $N_{\mathcal{M}_1}(X(\bar{x}))$. Recall that $X(\bar{x}) = (z(\bar{x}), y(\bar{x}), e)$. Thanks to Lemma 2, we know from Clarke (1990) that the normal cone $N_{\mathcal{M}_1}(X(\bar{x}))$ can be written as:

$$N_{\mathcal{M}_1}(X(\bar{x})) = -\mu_h \nabla_X h(w, e) + \mu_{\underline{e}} \nabla_X(\underline{e} - e) + \mu_{\bar{e}} \nabla_X(e - \bar{e}), \tag{28}$$

for some $\mu_h \geq 0$ and $\mu_e = (\mu_{\underline{e}}, \mu_{\bar{e}}) \in \mathbb{R}^2_+$ satisfying the complementary slackness conditions:

$$\mu_{\underline{e}}(e - \underline{e}) \equiv \mu_{\underline{e}}(e - \underline{e}) = 0, \quad \mu_{\underline{e}} \geq 0, \tag{29}$$
$$\mu_{\bar{e}}(e - \bar{e}) \equiv \mu_{\bar{e}}(e - \bar{e}) = 0, \quad \mu_{\bar{e}} \geq 0. \tag{30}$$

Simplification of (28) yields:

$$N_{\mathcal{M}_1}(X(\bar{x})) = \begin{pmatrix} 0 \\ -\mu_h \nabla_y h(w, e) \\ -\mu_h \nabla_e h(w, e) - \mu_{\underline{e}} + \mu_{\bar{e}} \end{pmatrix}. \tag{31}$$

Lemma 6. *If $e \in (\underline{e}, \bar{e})$, then $\lambda_e(0) = 0$. If $e = \underline{e}$, then $\lambda_e(0) \leq 0$. If $e = \bar{e}$, then $\lambda_e(0) \geq 0$.*

Proof. The proof follows directly from condition 23 applied to different forms of $N_{\mathcal{M}_0}(X(0))$ depending on where e takes values as previously computed. In particular:

- If $e \in (\underline{e}, \bar{e})$, then the normal cone takes the form 25, implying $\lambda_e(0) = 0$;
- If $e = \underline{e}$, then the normal cone takes the form 26, implying $\lambda_e(0) \leq 0$;
- If $e = \bar{e}$, the normal cone takes the form 27, implying $\lambda_e(0) \geq 0$.

■

Proposition 2. *The adjoint vector to e satisfies*

$$\lambda_e(0) = 2\left(\left\langle \lambda_z, \frac{\partial z(\bar{x})}{\partial e} \right\rangle + \left\langle \lambda_y, \frac{\partial y(\bar{x})}{\partial e} \right\rangle\right) - \mu_h c'(e) + \mu_{\underline{e}} - \mu_{\bar{e}}. \tag{32}$$

Moreover, when e is interior,

$$2\left(\left\langle \lambda_z, \frac{\partial z(\bar{x})}{\partial e} \right\rangle + \left\langle \lambda_y, \frac{\partial y(\bar{x})}{\partial e} \right\rangle\right) = \mu_h c'(e). \tag{33}$$

Proof. Note that (20) implies

$$\dot{\lambda}(x) = \begin{pmatrix} \dot{\lambda}_z(x) \\ \dot{\lambda}_y(x) \\ \dot{\lambda}_e(x) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -v(x - w(x)) \left\langle \lambda_z, \frac{\partial f(x|e)}{\partial e} \right\rangle - u(w(x)) \left\langle \lambda_y, \frac{\partial f(x|e)}{\partial e} \right\rangle \end{pmatrix}, \tag{34}$$

for a.e. $x \in I_x$. Hence $\lambda_z(x) = \lambda_z = cons$ and $\lambda_y(x) = \lambda_z = cons$ for all $x \in I_x$. Morever, in view of (31), condition (24) is equivalent to:

$$\begin{pmatrix} \lambda_z(\bar{x}) \\ \lambda_y(\bar{x}) \\ \lambda_e(\bar{x}) \end{pmatrix} = \begin{pmatrix} -\lambda_0 \nabla_z g(w, e) \\ \mu_h \nabla_y h(w, e) \\ -\lambda_0 \nabla_e g(w, e) + \mu_h \nabla_e h(w, e) - \mu_{\bar{e}} + \mu_{\underline{e}} \end{pmatrix}. \tag{35}$$

Thus

$$\lambda_z = \lambda_0 \left(p_i \phi'_P(z_i(\bar{x}))\right)_{i \in \mathcal{I}}, \tag{36}$$

$$\lambda_y = \mu_h \left(p_i \phi'_A(y_i(\bar{x}))\right)_{i \in \mathcal{I}}, \tag{37}$$

$$\lambda_e(\bar{x}) = \lambda_0 \sum_{i=1}^{n} p_i \phi'_P(z_i(\bar{x})) \frac{\partial z_i(\bar{x})}{\partial e} + \mu_h \left(\sum_{i=1}^{n} p_i \phi'_A(y_i(\bar{x})) \frac{\partial y_i(\bar{x})}{\partial e} - c'(e)\right) - \mu_{\bar{e}} + \mu_{\underline{e}}. \tag{38}$$

Substituting (36) and (37) into (38) yields a more compact expression for $\lambda_e(\bar{x})$. In particular,

$$\lambda_e(\bar{x}) = \left\langle \lambda_z, \frac{\partial z(\bar{x})}{\partial e} \right\rangle + \left\langle \lambda_y, \frac{\partial y(\bar{x})}{\partial e} \right\rangle - \mu_h c'(e) - \mu_{\bar{e}} + \mu_{\underline{e}}. \tag{39}$$

To ease on notation, define

$$\begin{aligned} Z(e) &= \int_{I_x} v(x - w(x)) \langle \lambda_z, f(x|e) \rangle \, dx, \\ Y(e) &= \int_{I_x} u(w(x)) \langle \lambda_y, f(x|e) \rangle \, dx. \end{aligned}$$

Then from (34),

$$\int_{I_x} \dot{\lambda}_e(x) dx = - (Z(e) + Y(e)) \, .$$

Observe that:

$$Y(e) = \frac{\partial}{\partial e} \left(\int_{I_x} \langle \lambda_y, \dot{z}(x) \rangle \, dx \right) = \left\langle \lambda_y, \frac{\partial}{\partial e} \left(\int_{I_x} \dot{z}(x) dx \right) \right\rangle$$

$$\implies Y(e) = \left\langle \lambda_y, \frac{\partial y(\bar{x})}{\partial e} \right\rangle. \tag{40}$$

Analogously,

$$Z(e) = \left\langle \lambda_z, \frac{\partial z(\bar{x})}{\partial e} \right\rangle. \tag{41}$$

Plugging (39), (40) and (41) into $\lambda_e(0) = \lambda_e(\bar{x}) - \int_{I_x} \dot{\lambda}_e(x)$, we arrive at (32).

To show the remaining part of the proposition, note that if $e \in (\underline{e}, \bar{e})$, then $\mu_{\underline{e}} = \mu_{\bar{e}} = 0$ by the complementary slackness conditions (29) and (30). Moreover, in this case $\lambda_e(0) = 0$ by Lemma 6. Plugging these additional pieces of information into (32) yields (33). ■

Lemma 7. *The non-triviality condition* $(\lambda_0, \mu_h) \neq 0 \in \mathbb{R}^2_+$ *holds.*

Proof. Suppose by contradiction that $\lambda_0 = \mu_h = 0$, then from (37) and (36) we have $\lambda_z = \lambda_y = 0$, which via (34) implies that $\dot{\lambda}_e(x) = 0$. Hence $\lambda_e(x) = \lambda_e = cons$ for all $x \in I_x$. In view of (32) we have $\lambda_e = \mu_{\underline{e}} - \mu_{\bar{e}}$. Consider the following cases.

- If $e \in (\underline{e}, \bar{e})$, then $\lambda_e = 0$, violating $(\lambda, \lambda_0) \neq 0$.
- If $e = \underline{e}$, then the complementary slackness condition (30) implies $\lambda_e = \mu_{\underline{e}}$. If $\mu_{\underline{e}} = 0$, then again $(\lambda, \lambda_0) \neq 0$ is violated. If $\mu_{\underline{e}} > 0$, then Lemma 6 is contradicted.

- If $e = \bar{e}$, then the complementary slackness condition (29) implies $\lambda_e = -\mu_{\bar{e}} \leq 0$. Thus the only admissible value for $\mu_{\bar{e}}$ in this case is zero, violating $(\lambda, \lambda_0) \neq 0$.

We see that regardless of the value of e, a contradiction follows if $\lambda_0 = \mu_h = 0$. Hence we always have $(\lambda_0, \mu_h) \neq 0$. ∎

Lemma 8. *The adjoint vectors λ_z and λ_y satisfies $(\lambda_z, \lambda_y) \in \mathbb{R}^{2n}_{++}$.*

Proof. In light of Lemma 7, we only need to consider the following cases.

1. If $\mu_h = 0$, $\lambda_0 = 1$, then

$$\begin{aligned} \lambda_z &= (p_i \phi'_P(z_i(\bar{x})))_{i \in \mathcal{I}}, \\ \lambda_y &= 0. \end{aligned}$$

   By Assumption 1 and Assumption 8, we have $\lambda_z \in \mathbb{R}^n_{++}$. But this implies $\frac{\partial H}{\partial \omega} = -v'(x-\omega) \langle \lambda_z, f(x|e) \rangle < 0$ for all $\omega \in [0, x]$, since $v' > 0$ by Assumption 7 and $f(x|e) \in \mathbb{R}^n_{++}$ by Assumption 2. Thus in this case $w(x) = 0$ for all $x \in I_x$, which is ruled out in light of Remark 1.

2. If $\mu_h > 0$, $\lambda_0 = 0$, then

$$\begin{aligned} \lambda_z &= 0, \\ \lambda_y &= \mu_h (p_i \phi'_A(y_i(\bar{x})))_{i \in \mathcal{I}}, \end{aligned}$$

   implying $\frac{\partial H}{\partial \omega} = u'(\omega) \langle \lambda_y, f(x|e) \rangle > 0$ for all $\omega \in [0, x]$ by the same argument as in the previous case. Thus $w(x) = x$ for all $x \in I_x$. This case is also ruled out in light of Remark 1.

3. If $\mu_h > 0$, $\lambda_0 = 1$, then

$$\begin{aligned} \lambda_z &= (p_i \phi'_P(z_i(\bar{x})))_{i \in \mathcal{I}}, \\ \lambda_y &= \mu_h (p_i \phi'_A(y_i(\bar{x})))_{i \in \mathcal{I}}. \end{aligned}$$

We conclude that only $(\mu_h > 0, \lambda_0 = 1)$ can occur under the standing assumptions. In this case $(\lambda_z, \lambda_y) \in \mathbb{R}^{2n}_{++}$ by assumptions on the priors and the welfare functional. ∎

Lemma 9. *The Hamiltonian is strictly concave in ω.*

Proof. From (21), we have for each fixed $x \in (0, \bar{x}]$

$$\frac{\partial H}{\partial \omega} = -v'(x-\omega) \langle \lambda_z, f(x|e) \rangle + u'(\omega) \langle \lambda_y, f(x|e) \rangle, \quad \omega \in (0, x), \tag{42}$$

and

$$\frac{\partial^2 H}{\partial \omega^2} = v''(x-\omega) \langle \lambda_z, f(x|e) \rangle + u''(\omega) \langle \lambda_y, f(x|e) \rangle. \tag{43}$$

By Lemma 8 and Assumption 2, we have $\langle \lambda_z, f(x|e)\rangle > 0$ and $\langle \lambda_y, f(x|e)\rangle > 0$. In addition, we also have $u'' < 0$ and $v'' \leq 0$ by Assumption 7. Hence $v''(x-\omega)\langle \lambda_z, f(x|e)\rangle \leq 0$ and $u''(\omega)\langle \lambda_y, f(x|e)\rangle < 0$, implying $\frac{\partial^2 H}{\partial \omega^2} < 0$ for all $\omega \in (0, x)$. ■

Theorem 10. *Define*

$$G(x|e) = \frac{\langle \lambda_y, f(x|e)\rangle}{\langle \lambda_z, f(x|e)\rangle}, \quad x \in I_x. \tag{44}$$

then the following holds for an optimal wage function.

1. *If the principal is risk-averse, then the optimal wage function takes the form*

$$w(x) = \Xi_x\left(G(x|e)\right), \quad x \in (0, \bar{x}], \tag{45}$$

where $\Xi_x : \mathbb{R}_{++} \to (0, x)$ is the inverse mapping of $\omega \mapsto \frac{v'(x-\omega)}{u'(\omega)}$. Moreover, the optimal wage is non decreasing in outcomes.

2. *If the principal is risk-neutral, then there exist $x_0 \in (0, \bar{x})$ such that an optimal wage function takes the form*

$$\begin{cases} w(x) = x & x \leq x_0 \text{ or } x \in (x_0, \bar{x}) \setminus J, \\ w(x) = \Xi\left(G(x|e)\right) & x \in J, \end{cases} \tag{46}$$

where $\Xi : \mathbb{R}_{++} \to (0, x)$ is the inverse mapping of $\omega \mapsto \frac{1}{u'(\omega)}$ and the set J is the countable union of open intervals defined by

$$J = \{x \in (x_0, \bar{x}) \mid u'(x)(G(x|e) < 1\}. \tag{47}$$

In particular, an optimal wage is differentiable on $(0, \bar{x})$ except at an at most countable set of points.

Moreover, x_0 is the smallest $x \in (0, \bar{x})$ such that $u'(x)(G(x|e) = 1$ and there exists a decreasing sequence $(x_l)_{l \geq 1}$ converging to x_0 with $u'(x_l)(G(x_l|e) < 1$ for $l \geq 1$.

An optimal effort satisfies $h(w, e) = 0$, i.e.,

$$\sum_{i=1}^{n} p_i\left(\phi_A(y_i(\bar{x}))\right) - c(e) = \bar{U},$$

where the optimal state vector $(y_i(\bar{x}))_{i \in \mathcal{I}}$ is evaluated at the optimal wage in each case. Moreover, either an optimal effort is not interior (and then belongs to $\{\underline{e}, \bar{e}\}$) or it is interior and satisfies (33).

Proof. Notice that $w(x) = 0$ cannot occur for any $x \in I_x$ since the Inada condition on u implies $\frac{\partial H}{\partial \omega}|_{\omega=0} = +\infty$, regardless of the principal's attitude towards risk.

1. When the principal is risk-averse, the Inada condition on $v(\cdot)$ implies that $w(x) = x$ cannot occur for any x since $\frac{\partial H}{\partial \omega}|_{\omega=x} = -\infty$. Thus the optimal wage satisfies $\frac{\partial H}{\partial \omega}|_{\omega=w(x)} = 0$, or

$$\frac{v'(x - w(x))}{u'(w(x))} = G(x|e). \tag{48}$$

Note that since $v'' \leq 0$ and $u' > 0$,

$$\frac{-v''(x-\omega)u'(\omega) - v'(x-\omega)u''(\omega)}{[u'(\omega)]^2} > 0 \tag{49}$$

for any fixed x, implying that the mapping $\omega \mapsto \frac{v'(x-\omega)}{u'(\omega)}$ is strictly increasing. Thus it has an inverse mapping Ξ_x, also strictly increasing, such that for each $x \in (0, \bar{x}]$, the optimal wage is uniquely defined by

$$w(x) = \Xi_x\left(G(x|e)\right), \quad x \in (0, \bar{x}]. \tag{50}$$

One deduces from the assumptions that $w(\cdot)$ is differentiable on its domain. Let $G'(x|e) = \frac{\partial G(x|e)}{\partial x}$, we have

$$w'(x) = \Xi_x'(G(x|e))G'(x|e), \tag{51}$$

where $\Xi_x' > 0$ since Ξ_x is a strictly increasing map, implying that $w'(x)$ has the same sign as $G'(x|e)$. Furthermore, differentiating both sides of (48) with respect to x and simplifying yield

$$w'(x) = \frac{r_v(x - w(x)) + \frac{G'(x|e)}{G(x|e)}}{r_v(x - w(x)) + r_u(w(x))}, \tag{52}$$

where $r_v(\cdot) = -\frac{v''(\cdot)}{v'(\cdot)} > 0$ and $r_u(\cdot) = -\frac{u''(\cdot)}{u'(\cdot)} > 0$ denote the degree of the absolute risk aversion of the principal and the agent, respectively. Observe that $G(x|e) = \frac{\langle \lambda_y, f(x|e)\rangle}{\langle \lambda_z, f(x|e)\rangle}$ is bounded above and below by positive constants since all the elements of the adjoint vectors are positive and finite, and the densities are positive and bounded. Since $w(x) \in (0, x)$ for all $x \in (0, \bar{x}]$, there exists $x_1 > 0$ such that $w'(x_1) > 0$, otherwise $w(x) \leq 0$ for all $x \in I_x$ since $w(0) = 0$, which is either inadmissible, or ruled out by a previous remark. From (51) we have $G'(x_1|e) > 0$. We would like to show that $G'(x|e) \geq 0$ for all $x \in I_x$. Suppose by contradiction, there exists $x_2 \in (0, \bar{x}]$ such that $G'(x_2|e) < 0$. By the continuity of $G'(\cdot|e)$, there exists $x_3 \in (0, \bar{x}]$ such that $G'(x_3|e) = 0$, which, via (51) implies $w'(x_3) = 0$. But $G'(x_3|e) = 0$ and $w'(x_3) = 0$ imply $r_v(\cdot) = 0$ in light of (52), a contradiction. Hence we must have $G'(x|e) \geq 0$ for all $x \in I_x$, implying via (51) that $w'(x) \geq 0$ for all $x \in I_x$.

2. When the principal is risk-neutral, $v' \equiv 1$. Equation (42) reads

$$\frac{\partial H}{\partial \omega} = \langle \lambda_z, f(x|e) \rangle \left(u'(\omega)G(x|e) - 1 \right), \tag{53}$$

and by strict convexity of the Hamiltonian, one gets that, for $x \in (0, \bar{x})$, the latter reaches its maximum at $\omega = x$ if and only if $u'(x)G(x|e) \geq 1$.

By the Inada condition on u, one deduces at once $w(x) = x$ for x small enough. Let x_0 be the largest $x \in I_x$ such that $w(x) = x$ on $[0, x_0]$. First notice that $x_0 < \bar{x}$, otherwise $w(x) = x$ on I_x, which is ruled out by a previous remark. We claim that $u'(x_0)G(x_0|e) = 1$. Indeed, one clearly has that $u'(x)G(x|e) \geq 1$ for $x < x_0$ and therefore $u'(x_0)G(x_0|e) \geq 1$. Furthermore by maximality of x_0, there exists a decreasing sequence $(x_l)_{l \geq 1}$ tending to x_0 such that $w(x_l) < x_l$ for $l \geq 1$, i.e., $u'(x_l)G(x_l|e) < 1$, yielding, as l tends to infinity that $u'(x_0)G(x_0|e) \leq 1$. Hence the claim and, as a byproduct, the characterization of x_0. Let the subset J of $(x_0, \bar{x})$ defined in (47). Since $u'(\cdot)G(\cdot|e)$ is continuous, J is the countable union of open intervals on which the optimal wage $w(x)$ is defined by $\frac{\partial H}{\partial \omega}|_{\omega = w(x)} = 0$, i.e., by risk-neutrality of the principal, yields

$$\frac{1}{u'(w(x))} = G(x|e), \quad x \in J. \tag{54}$$

Since $\omega \mapsto \frac{1}{u'(\omega)}$ is strictly increasing by the strict concavity of u, the inverse mapping $\Xi \equiv \left(\frac{1}{u'}\right)^{-1}$ is well-defined. Inverting both sides of (54), we arrive at the desired expression for $w(x)$. This completes the proof.

■

Remark 2. *One may wonder about the uniqueness of the solution. Clearly, a solution is determined by the triple (μ_h, x_0, e). In the case where e is not interior, we have two unknowns and two equations: one from the equality constraint $h(w, e) = 0$, and the other from the characterization of x_0 provided by the theorem, i.e., the study of the solution to equation $u'(x)G(x|e) = 0$. In the case where e is interior, we have an extra equation, namely (33). However, it seems difficult to determine all the solutions analytically.*

Remark 3. *With reasonable assumptions on the data of the problem, such as real-analyticity, one can conclude that there is a finite number of solutions to $u'(x)G(x|e) = 1$ on $(0, \bar{x}]$ and then J is made of a finite number of open intervals. Furthermore, notice that on $(0, \bar{x}]$, one has that $(u'G)' = u''G + u'G' = u'(\frac{G'}{G} - r_u)$. One deduces that if $\frac{G'}{G} \leq r_u$ then $J = (x_0, \bar{x})$ and x_0 is the unique solution to $u'(x)G(x|e) = 1$.*

When both DMs are ambiguity-neutral, we recover the result that is most analogous to the straight deductible result of Raviv (1979) in the context of insurance.

Corollary 1. ***[Raviv (1979)]*** *When both DMs are ambiguity-neutral and the principal is risk-neutral, the shape of the optimal contract is the following*

$$\begin{cases} w(x) = x & x \in (0, x_0] \\ w(x) = x_0 \in (0, x) & x \in (x_0, \bar{x}] \end{cases},$$

where x_0 is uniquely defined by $u'(x_0) = \frac{1}{\mu_h}$.

Proof. When both DMs are ambiguity-neutral,

$$G(x|e) = \frac{\mu_h \langle p, f(x|e) \rangle}{\langle p, f(x|e) \rangle} = \mu_h, \tag{55}$$

where $p \equiv (p_1, \ldots, p_n)$ is the vector of priors.

According to Remark 3, x_0 is the unique solution to

$$u'(x_0)\mu_h - 1 = 0. \tag{56}$$

By the monotonicity of u', for all $x \leq x_0$ we have $u'(x)\mu_h - 1 \geq 0$, implying that $w(x) = x$ for all $x \leq x_0$ and then for all $x > x_0$ we have $u'(x)\mu_h - 1 < 0$, implying that $w(x) = x_0$ for all $x > x_0$. ■

Remark 4. *Observe that the sharing rule (48) characterizes both efficient risk and ambiguity-sharing. It has a nice interpretation. The LHS of this equation is the relative marginal utilities, while the RHS is relative expected marginal welfare. The expectation is computed with respect to the posterior distribution. To see this, define the expected marginal welfare of the agent and the principal, respectively, be:*

$$A(x|e) = \frac{\sum_{i=1}^{n} p_i \phi'_A(y_i(\bar{x})) f_i(x|e)}{\sum_{i=1}^{n} p_i f_i(x|e)} \equiv \sum_{i=1}^{n} p_i(x|e) \phi'_A(y_i(\bar{x})), \tag{57}$$

$$P(x|e) = \frac{\sum_{i=1}^{n} p_i \phi'_P(z_i(\bar{x})) f_i(x|e)}{\sum_{i=1}^{n} p_i f_i(x|e)} \equiv \sum_{i=1}^{n} p_i(x|e) \phi'_P(z_i(\bar{x})). \tag{58}$$

where $p_i(x|e) = \frac{p_i f_i(x|e)}{\sum_{i=1}^{n} p_i f_i(x|e)}$ is the probability that state i occurs given that the outcome is x, which is by definition the Bayesian posterior probability. This "inference" that each contracting party has to make here is a direct consequence of the uncertainty on the distribution and that the state itself is not a contractible variable. Then (48) is equivalent to

$$\frac{v'(x - w(x)) P(x|e)}{u'(w(x)) A(x|e)} = \mu_h, \tag{59}$$

which tells us that at the optimum, the ratio between the product of marginal utility and expected marginal welfare of the agent and that of the principal is equalized across all levels of outcomes where optimal wage has an interior value. Notice that *under ambiguity neutrality of both DMs, $\frac{A(x|e)}{P(x|e)} = cons \equiv 1$ and (59) reduces to the famous Borch rule*

$$\frac{v'(x - w(x))}{u'(w(x))} = \mu_h, \tag{60}$$

according to Borch (1960). Hence (59) can be viewed as a modified Borch rule that an optimal contract has to satisfy under ambiguity.

Remark 5. *To clarify the notion of Bayesian inference mentioned above, consider a situation where the states are contractible variables, i.e., when the contract can be written as $\{w_i(\cdot), e\}_{i \in \mathcal{I}}$ instead of $\{w(\cdot), e\}$. In this case, we can slightly modify the state variables as*

$$\begin{aligned} z(x) &= (v(x - w_i(x)) f_i(x|e))_{i \in \mathcal{I}}, \\ y(x) &= (u(w_i(x)) f_i(x|e))_{i \in \mathcal{I}}, \end{aligned}$$

and and easily show that the sharing rule (59) holds for every state. In particular,

$$\frac{v'(x - w_i(x))\phi_P'(z_i(\bar{x}))}{u'(w_i(x))\phi_A'(y_i(\bar{x}))} = \mu_h, \quad \forall x \in I_x,\ i \in \mathcal{I}, \tag{61}$$

implying that the state-conditional ratio of marginal utilities $\frac{v'(x - w_i(x))}{u'(w_i(x))}$ is held constant across all $x \in I_x$, in each state $i \in \mathcal{I}$. Hence (61) can be viewed as a state-conditional Borch rule. Observe that the DMs no longer have to make an inference on the state based on the outcome as suggested by (48). Furthermore, when the principal is neutral to risk and ambiguity,

$$u'(w_i(x))\phi_A'(y_i(\bar{x})) = \mu_h, \quad \forall i \in \mathcal{I}, \tag{62}$$

implying that $w_i(x) = cons \equiv \bar{w}_i \in (0, \bar{x})$ for all x satisfying $w_i(x) \in (0, x)$. In view of Corollary 1, the optimal wage function under risk and ambiguity-neutral principal when the states are contractible has the form:

$$\begin{cases} w_i(x) = x & x \in (0, \bar{w}_i], \\ w_i(x) = \bar{w}_i & x \in (\bar{w}_i, \bar{x}], \end{cases} \quad i \in \mathcal{I}, \tag{63}$$

where $\bar{w}_i$ is the unique solution to $K_i(x) = 0$ where $K_i(x) = u'(x)\phi_A'(y_i(\bar{x}))\mu_h - 1$. Observe that if $y_i(\bar{x}) = \int_0^{\bar{w}_i} u(x) f_i(x|e)dx + (1 - F_i(\bar{w}_i))u(\bar{w}_i)$, then

$$\frac{\partial y_i(\bar{x})}{\partial \bar{w}_i} = u'(\bar{w}_i)(1 - F(\bar{w}_i)) > 0 \tag{64}$$

by Leibniz's integral rule. Thus the LHS of (62) is strictly decreasing in $\bar{w}_i$ by the strict concavity of $u(\cdot)$ and $\phi_A(\cdot)$. Hence in optimality $\bar{w}_i = \bar{w}_j \equiv \bar{w}$ for all $i, j \in \mathcal{I}$, further simplifying (63) to

$$\begin{cases} w_i(x) = x & x \in (0, \bar{w}], \\ w_i(x) = \bar{w} & x \in (\bar{w}, \bar{x}], \end{cases} \quad \forall i \in \mathcal{I}. \tag{65}$$

Observe that the contract (63) is robust to the principal's ambiguity attitude. When the principal is ambiguity-averse, we simply modify $K_i(x) = u'(x)\frac{\phi'_A(y_i(\bar{x}))}{\phi'_P(z_i(\bar{x}))}\mu_h - 1$, which is also strictly decreasing in x. Hence, if the states are contractible (markets are complete), we conclude from (65) and Corollary 1 that the shape of the optimal contract is robust to ambiguity when the principal is risk-neutral.

5 Binary ambiguous state case with risk-neutral principal

Under a risk-averse principal, we have shown in Theorem 10 that G' is always non negative, regardless of the number of ambiguous states and, since the sign of G' determines the sign of w', an optimal wage was always non decreasing.

For $n = 2$, we prove in the next lemma that G' has a constant sign a hence deduces some information on the sign of an optimal wage under a risk-neutral principal.

Lemma 11. *Under Assumption 3, in the binary state case $n = 2$, the sign of $G'(x|e)$ does not depend on $x \in I_x$. As a consequence, the following holds: either $G' \leq 0$, in which case x_0 is the unique solution of $u'(x)G(x|e) = 1$ and J defined in (47) is equal to $(x_0, \bar{x})$; or $G' > 0$, in which case an optimal wage is non decreasing.*

Proof. Denote $\lambda_z = (\lambda_z^1, \ldots, \lambda_z^n)$ and $\lambda_y = (\lambda_y^1, \ldots, \lambda_y^n)$. Then $G'(x|e)$ can be expressed as:

$$G'(x|e) = \frac{\sum_{1\leq i<j\leq n}(\lambda_y^i\lambda_z^j - \lambda_y^j\lambda_z^i)f_j^2(x|e)\ell'_{ij}(x|e)}{\left(\sum_{i=1}^n \lambda_z^i f_i(x)\right)^2}. \tag{66}$$

Hence for $n = 2$

$$G'(x|e) = \frac{\left(\lambda_y^1\lambda_z^2 - \lambda_y^2\lambda_z^1\right)\left[f_2(x|e)\right]^2 \ell'_{12}(x|e)}{\left(\sum_{i=1}^2 \lambda_z^i f_i(x|e)\right)^2},$$

where $\ell'_{12}(x) \geq 0$ by Assumption 3. Clearly, the sign of $G'(x|e)$ depends on the sign of $\lambda_y^1\lambda_z^2 - \lambda_y^2\lambda_z^1$, which is independent of x. Combining that result with Theorem 10 yields the rest of the statement of the lemma. ■

Remark 6. *Denote $A'(x|e) = \frac{\partial A(x|e)}{\partial x}$ and $P'(x|e) = \frac{\partial P(x|e)}{\partial x}$. We have*

$$A'(x|e) = \frac{\sum_{1\leq i<j\leq n} p_i p_j\left(\phi'_A(y_i(\bar{x})) - \phi'_A(y_j(\bar{x}))\right) f_j^2(x|e)\ell'_{ij}(x|e)}{\langle p, f(x|e)\rangle^2}, \tag{67}$$

$$P'(x|e) = \frac{\sum_{1\leq i<j\leq n} p_i p_j\left(\phi'_P(z_i(\bar{x})) - \phi'_P(z_j(\bar{x}))\right) f_j^2(x|e)\ell'_{ij}(x|e)}{\langle p, f(x|e)\rangle^2}, \tag{68}$$

where $\ell'_{ij}(x|e) \geq 0$ by Assumption 3 and $p_i p_j > 0$ for all $i, j \in \mathcal{I}$ by Assumption 1. Hence again in the binary state case the signs of $A'(x|e)$ and $P'(x|e)$ are independent of x.

Lemma 12. *In the binary state case, $A'(x|e) \leq 0$ and $P'(x|e) \leq 0$ for all $x \in I_x$.*

Proof. Note that by integration by parts (IBP), we always have

$$y_1(\bar{x}) = u(w(\bar{x})) - \int_0^{\bar{x}} u'(w(x))w'(x)F_1(x|e)dx, \tag{69}$$

$$z_1(\bar{x}) = (\bar{x} - w(\bar{x})) - \int_0^{\bar{x}} (1 - w'(x))F_1(x|e)dx. \tag{70}$$

Hence

$$y_1(\bar{x}) - y_2(\bar{x}) = \int_0^{\bar{x}} u'(w(x))w'(x)\left[F_2(x) - F_1(x)\right]dx, \tag{71}$$

and

$$z_1(\bar{x}) - z_2(\bar{x}) = \int_0^{\bar{x}} (1 - w'(x))\left[F_2(x|e) - F_1(x|e)\right]dx. \tag{72}$$

Recall that when $n = 2$, by Lemma 11 the sign of $G'(x|e)$ is constant with respect to x. Suppose $G'(x|e) \geq 0$, then (51) implies $w'(x) \geq 0$ for $x \in (x_0, \bar{x})$ (and thus for all $x \in I_x$). Thus $y_1(\bar{x}) \geq y_2(\bar{x})$ by (71) since $u' > 0$, $w' \geq 0$ and $F_2(x) \geq F_1(x) \geq 0$ for all $x \in I_x$ by Assumption 3. Hence $\phi'_A(y_1(\bar{x})) - \phi'_A(y_2(\bar{x})) \leq 0$ by the concavity of ϕ_A, implying $A'(x|e) \leq 0$ in light of (67). Moreover since $G(x|e) = \mu_h \frac{A(x|e)}{P(x|e)}$, differentiating with respect to x and simplifying yield:

$$\frac{G'(x|e)}{G(x|e)} = \frac{A'(x|e)}{A(x|e)} - \frac{P'(x|e)}{P(x|e)} \tag{73}$$

$$\iff \frac{P'(x|e)}{P(x|e)} = \frac{A'(x|e)}{A(x|e)} - \frac{G'(x|e)}{G(x|e)}, \tag{74}$$

implying $P'(x|e) \leq 0$ since $A'(x|e) \leq 0$ and $G'(x|e) \geq 0$.

Consider next the case $G'(x|e) < 0$, again from (51) we have $w'(x) < 0$ for $x \in (x_0, \bar{x})$. Thus $w'(x) \leq 1$ for all $x \in I_x$, and we have from (72) that $z_1(\bar{x}) \geq z_2(\bar{x})$. Hence $P'(x|e) \leq 0$ in light of (68). On the other hand (73) implies $\frac{A'(x|e)}{A(x|e)} = \frac{G'(x|e)}{G(x|e)} + \frac{P'(x|e)}{P(x|e)} < 0$ since $G'(x|e) < 0$ and $P'(x|e) \leq 0$. Thus $A'(x|e) < 0$.

We conclude that regardless of the sign of $G'(x|e)$, when $n = 2$ we always have $P'(x|e) \leq 0$ and $A'(x|e) \leq 0$ for all $x \in I_x$. ∎

Remark 7. *As a direct consequence of Lemma 12, in the binary state case, we always have $y_1(\bar{x}) \geq y_2(\bar{x})$ and $z_1(\bar{x}) \geq z_2(\bar{x})$.*

Example 3. *Consider the binary state case with power welfare function* $\phi_J(U) = U^{1-\sigma_J}$*, where* $\sigma_J \in [0,1)$ *is the degree of relative ambiguity aversion of decision maker* J*, where* $J \in \{A, P\}$*. Define* $\hat{y} = \frac{y_1(\bar{x})}{y_2(\bar{x})}$ *and* $\hat{z} = \frac{z_1(\bar{x})}{z_2(\bar{x})}$*. Then* $\hat{y} \geq 1$ *and* $\hat{z} \geq 1$ *by the remark above. Note that* $\lambda_y^1 \lambda_z^2 - \lambda_y^2 \lambda_z^1 = p_1 p_2 \mu_h S(\sigma_P, \sigma_A)$*, where*

$$S(\sigma_P, \sigma_A) = \phi_A'(y_1(\bar{x}))\phi_P'(z_2(\bar{x})) - \phi_A'(y_2(\bar{x}))\phi_P'(z_1(\bar{x})).$$

Thus $S(\sigma_P, \sigma_A)$ *bears the same sign as* $G'(x|e)$*. Upon simplification we have*

$$S(\sigma_P, \sigma_A) = (1-\sigma_A)(1-\sigma_P)\frac{1}{y_2 z_1}\left(\frac{\hat{z}^{\sigma_P}}{\hat{y}^{\sigma_A}} - 1\right).$$

Clearly, the sign of $S(\sigma_P, \sigma_A)$ *depends on the sign of* $\frac{\hat{z}^{\sigma_P}}{\hat{y}^{\sigma_A}} - 1$*. Taking* $\sigma_P = 0$*, the sign of* $S(\sigma_P, \sigma_A)$ *is simply that of* $1 - \hat{y}^{\sigma_A}$ *which can be made negative. By contrast, if* $\sigma_A = 0$*, the sign of* $S(\sigma_P, \sigma_A)$ *is that of* $\hat{z}^{\sigma_P} - 1$ *which can be made positive. In other words,* G' *can take both signs in the binary case and its sign depends crucially on the degree of ambiguity aversion.*

Definition 1. *For lack of better terminologies, we say that the ambiguity-aversion effect of the agent dominates that of the principal if*

$$G'(x|e) \leq 0, \qquad \forall x, \tag{75}$$

and vice versa. We say that the two ambiguity - aversion effects offset each other if the above holds with equality.

Remark 8. *In the binary state case, (75) is independent of* x *and in light of Example 3 is more likely to hold the more ambiguity-averse the agent and the less ambiguity-averse the principal.*

Remark 9. *When (75) holds and under the principal's risk-neutrality and the agent's ambiguity dominance assumptions,* x_0 *is uniquely defined by equation* $u'(x)(G(x|e) = 1$ *and an optimal contract (46) can be fully characterized by*

$$\begin{cases} w(x) = x & x \in (0, x_0], \\ w(x) = \Xi\left(G(x|e)\right) & x \in (x_0, \bar{x}], \end{cases} \tag{76}$$

where $\Xi : \mathbb{R}_{++} \to (0, x)$ *is the inverse mapping of* $\omega \mapsto \frac{1}{u'(\omega)}$*.*

Proposition 3. *In the binary state case, when the principal is ambiguity neutral, the variation in optimal wage whenever* $w(x) \in (0, x)$ *is the following.*

1. *If both DMs are ambiguity -averse, optimal wage is non decreasing in outcomes on* $[x_0, \bar{x}]$ *if and only if the ambiguity-aversion effect of the principal dominates that of the agent;*

2. *If the principal is ambiguity-neutral and the agent is ambiguity-averse, optimal wage is non increasing in outcomes. Moreover, optimal wage is constant if and only if ambiguity has a one-sided structure, i.e., either ambiguity is concentrated only outcomes beyond x_0, or concentrated only on outcomes above x_0.*

Proof. We consider each case separately.

1. To prove the first statement of the proposition, note that risk-neutrality of the principal implies $r_v = 0$. Thus (52) becomes

$$w'(x) = \frac{\frac{G'(x|e)}{G(x|e)}}{r_u(w(x))}. \tag{77}$$

   Since $r_u(\cdot) > 0$, optimal wage is non decreasing on $[x_0, \bar{x}]$ if and only if $G'(x|e) \geq 0$, i.e., if and only if the ambiguity aversion effect of the principal dominates that of the agent by Definition 1. By the same token, optimal wage is non - increasing on $[x_0, \bar{x}]$ if the ambiguity-aversion effect of the agent dominates.

2. For the second statement of the proposition, observe that when the principal is neutral to both risk and ambiguity, (77) simplifies to

$$w'(x) = \frac{\frac{A'(x|e)}{A(x|e)}}{r_u(w(x))}, \tag{78}$$

   implying that $w'(x) \leq 0$ on $[x_0, \bar{x}]$ since $A'(x|e) \leq 0$ for all $x \in I_x$ by Lemma 12. Thus the only non decreasing candidate solution satisfies $w'(x) = 0$ for all $x \in [x_0, \bar{x}]$, which satisfies (78) if and only if $A'(x|e) = 0$ for all $x \in [x_0, \bar{x}]$. Note that in the binary case (71) can be re-written as

$$\begin{aligned} y_1(\bar{x}) - y_2(\bar{x}) \quad &= \int_0^{x_0} u'(x)\,[F_2(x) - F_1(x)]\,dx \\ &+ \int_{x_0}^{\bar{x}} u'(w(x))w'(x)\,[F_2(x) - F_1(x)]\,dx. \end{aligned}$$

   Since $w'(x) = 0$ on $[x_0, \bar{x}]$,

$$y_1(\bar{x}) - y_2(\bar{x}) = \int_0^{x_0} u'(x)\,[F_2(x|e) - F_1(x|e)]\,dx,$$

   which is zero if and only if $F_2(x|e) = F_1(x|e)$ a.e. $x \in (0, x_0]$. Thus from (67) we have $A'(x|e) = 0$ on $[x_0, \bar{x}]$ if and only if either

$$\ell'_{12}(x|e) = 0, \quad \forall x \in (x_0, \bar{x}] \tag{79}$$

or

$$F_2(x|e) = F_1(x|e), \quad \text{a.e. } x \in (0, x_0]. \tag{80}$$

Condition (79) is satisfied if ambiguity is concentrated only on $(0, x_0]$, while (80) is satisfied if ambiguity is concentrated only on $(x_0, \bar{x}]$. Thus if the distribution of outcome has either of these one-sided ambiguous structures, then then the optimal wage contract is identical to the unambiguous case as expressed in Corollary 1.

Remark 10. *If ambiguity contaminates both sides of the support, then there exists a sub-interval of $(x_0, \bar{x}]$ where $w'(x) < 0$. For example, suppose there exists an interval $J_x = [x_1, x_2]$ where $x_1 \in (0, x_0)$ and $x_2 \in (x_0, \bar{x})$ satisfying $\ell'_{12}(x|e) > 0$ for all $x \in J_x$, then $w'(x) \geq 0$ for all $x \in I_x$ implies $y_1(\bar{x}) - y_2(\bar{x}) > 0$, which in turn implies $\phi'_A(y_1(\bar{x})) - \phi'_A(y_2(\bar{x})) > 0$ by the strict concavity of ϕ_A. Necessarily $A'(x|e) < 0$ on J_x, which in light of (78) implies that $w'(x) < 0$ on $[x_2, \bar{x}]$, contradicting the hypothesis that $w'(x) \geq 0$ for all $x \in I_x$. Hence, if the one-sided ambiguity structure is violated, there exists a subset of outcomes where optimal wage is strictly decreasing.*

■

6 Optimal wage under a modified admissible set

In this section, we modify one assumption, namely Assumption 5 and add another Inada condition on the utility function u in Assumption 7. We modify also the outcome set so that $I_x = [0, \infty)$. In particular, the followings hold.

Assumption 9. *The wage is a measurable function $w : I_x \to \mathbb{R}_{++}$ satisfying $w(x) \geq 0$ for all $x \in I_x = [0, \infty)$.*

Essentially, this modification allows the wage to be greater than the outcome for some outcomes; it is no longer constrained point-wise.

Assumption 10. *The agent has utility function $u : \mathbb{R}_+ \to \mathbb{R}$, which is at least C^2, satisfying $u(0) = 0$, $u' > 0 > u''$ and the Inada conditions $\lim_{w \to 0} u'(w) = +\infty$, and $\lim_{w \to \infty} u'(w) = 0$.*

Under the revised assumptions, the admissible control $\mathcal{U}$ coincides with the set $M(I_x)$, hence $\mathcal{U}$ is compact. The first major change occurs in the maximum condition of the PMP. In particular, equation (22) now reads

$$H(X(x), w(x), \lambda(x), \lambda_0, x) = \max_{\omega \in I_x} H(X(x), \omega, \lambda(x), \lambda_0, x) \tag{81}$$

for almost every x in I_x. Under the modified maximum condition, the second part of Theorem 10 changes to the following.

Proposition 4. *Consider the principal-agent model with a risk-neutral principal and a risk-averse agent. For each x in I_x, the unique optimal wage satisfies*

$$w(x) = \Xi(G(x|e)), \quad \forall x \in I_x. \tag{82}$$

An optimal wage satisfies the equality constraint $h(w, e) = 0$. Moreover, either an optimal effort is not interior and takes value in $\{\underline{e}, \bar{e}\}$, or it is interior and satisfies (33).

Proof. Observe that with the Inada conditions imposed at the boundary of the control set, corner solution cannot occur for any outcome. In particular, we have $H'(0) = +\infty$ for all x.[6] Since $H'(\omega)$ is strictly decreasing, for each fixed x in I_x, there exists a unique $w(x)$ in $(0, \bar{x})$ satisfying $H'(w(x)) = 0$, which is equivalent to $u'(w(x))G(x|e) - 1 = 0$. Recall that Ξ is the inverse mapping to $\omega \mapsto \frac{1}{u'(\omega)}$, which is strictly increasing. Hence $H'(w(x)) = 0$ is equivalent to (82), as desired. The proof for the result pertaining to the optimal level of effort is as before. ■

Corollary 2. *Under ambiguity neutrality, fixed wage is optimal. In particular,*

$$w(x) = \bar{w} = \Xi(\mu_h), \quad \forall x \in I_x.$$

Proof. The corollary follows immediately from the fact that under ambiguity neutrality (of both DMs), $G = \mu_h$. ■

Proposition 5. *Consider the binary-state principal-agent model with a risk-neutral principal and a risk-averse agent. If the principal is ambiguity-neutral and the agent ambiguity-averse, then a fixed wage contract is optimal. In particular,*

$$w(x) = \tilde{w} = \tilde{\Xi}(\mu_h), \quad \forall x \in I_x, \tag{83}$$

where $\tilde{\Xi}$ is the inverse mapping of $\omega \mapsto \frac{1}{u'(\omega)\phi'_A(u(\omega))}$. An optimal effort satisfies $h(\tilde{w}, e) = 0$, i.e.,

$$\phi_A(u(\tilde{w})) - c(e) = \bar{U}.$$

Moreover, either an optimal effort is not interior and takes value in $\{\underline{e}, \bar{e}\}$, or it is interior and satisfies (33).

Proof. First, consider the case of an ambiguity-neutral principal. Since $P' = 0$, we have that $\mu_h G' = A'$, which implies that $G' \leq 0$ by virtue of Lemma 12. Hence $w'(x) = \Xi'(G(x|e))G'(x|e) \leq 0$. Since w is positive-valued, the only admissible wage that satisfies $w'(x) \leq 0$ is $w'(x) = 0$ for all x. Denote the fixed wage by $\tilde{w}$. Recall that $\lambda_y = (p_i \phi'_A(y_i))_{i \in \mathcal{I}}$. Under constant wage,

[6] The optimal wage is also bounded above since the Inada condition implies $H'(\infty) < 0$.

$y_i = u(\tilde{w})$ for all i, hence $A'(x)$ reduces to just $\frac{\phi'_A(u(\tilde{w}))\langle p, f(x|e)\rangle}{\langle p, f(x|e)\rangle} = \phi'_A(u(\tilde{w}))$. Thus the condition $H'(w(x)) = 0$ simplifies to

$$\frac{1}{u'(\tilde{w})\phi'_A(u(\tilde{w}))} = \mu_h. \tag{84}$$

It is easy to check that the mapping $\omega \mapsto \frac{1}{u'(\omega\phi'_A(u(\omega))}$ is strictly increasing due to the strict concavity and monotonicity of u and ϕ_A. Hence its inverse mapping $\tilde{\Xi}$ exists and we could invert (84) to obtain (83). As before, if e takes an interior value, then the triple $(\mu_h, \tilde{w}, e)$ is pinned down by three equations, namely (84), (33) and $h(\tilde{w}, e) = 0$. ■

7 Optimal contract under moral hazard

In this section, we examine the impact of moral hazard on the optimal wage function derived in the previous section. We will consider only the case of two ambiguous states and two effort levels, i.e., $e \in \{e^H, e^L\}$ where $e^H > e^L \geq 0$. Observe that in order to implement high effort level e^H in this case, the principal cannot pay a fixed wage as in the previous setting since the agent can get away unpenalized with a a low effort level e^L. In other words, to implement the low effort, the principal just needs to offer a fixed wage (that is just enough to satisfy the participation constraint of the agent). The difficulty arises only when the principal wants to demand high effort from the agent. The problem faced by the principal that is neutral to both risk and ambiguity in this case reads:

$$\max_{w(\cdot)} \int (x - w(x)) f(x|e^H) dx \tag{P'}$$

$$\text{s.t.} \quad \sum_i p_i \phi_A(y_{i,e^H}) - c(e^H) \geq \bar{U}, \tag{IRC}$$

$$\sum_i p_i \phi_A(y_{i,e^H}) - c(e^H) \geq \sum_i p_i \phi_A(y_{i,e^L}) - c(e^L), \tag{ICC}$$

where

$$y_{i,e} \equiv \int u(w(x)) f_i(x|e) dx, \quad e \in \{e^H, e^L\}, \tag{85}$$

and

$$f(x|\cdot) \equiv \sum_i p_i f_i(x|\cdot). \tag{86}$$

Let λ and μ be the Lagrange multipliers associated with the IRC and ICC above. The necessary condition satisfied by an optimal wage function $w(\cdot)$ is:

$$\frac{1}{u'(w(x))} = \mu + \lambda \left(1 - \frac{\sum_i p_i \phi'_A(y_{i,e^L}) f_i(x|e^L)}{\sum_i p_i \phi'_A(y_{i,e^H}) f_i(x|e^H}\right). \tag{87}$$

Notice that under ambiguity neutrality, the fraction in the bracket on the RHS of (87) reduces to just the ambiguity-neutral likelihood ratio:

$$L(x) \equiv \frac{f(x|e^L)}{f(x|e^H)}, \tag{88}$$

which is often assumed to be decreasing in outcome. It is immediate to see that the ambiguity-neutral likelihood ratio being decreasing in outcome implies that the wage is increasing in outcome. In other words, just as in the well-known result in the moral hazard literature in absence of ambiguity, an optimal contract in this case also involves rewarding the agent for the good outcomes and penalizing him for the bad ones. In other words, the shape of the optimal contract under ambiguity and ambiguity neutrality is identical to that in absence of ambiguity.

We now consider the case of strict ambiguity aversion on the part of the agent, i.e., the case where ϕ_A is strictly concave. The fraction in the bracket on the RHS of (87) is the ambiguity-averse likelihood ratio (AALR), which we shall denote by $\hat{L}$:

$$\hat{L}(x) \equiv \frac{\sum_i p_i \phi'_A(y_{i,e^L}) f_i(x|e^L)}{\sum_i p_i \phi'_A(y_{i,e^H}) f_i(x|e^H)}. \tag{89}$$

Furthermore, we define the state-conditional likelihood ratio:

$$L_i(x) \equiv \frac{f_i(x|e^L)}{f_i(x|e^H)}, \quad i \in \{1,2\}, \tag{90}$$

which is decreasing by the LRD assumption. In particular, higher effort shifts the outcome distribution in the sense of LRD, which means $L'_i(x) \leq 0$ mathematically.

It can then be shown that:

$$\hat{L}(x) = \sum_i \hat{p}_i(x|e^H) R_i L_i(x), \tag{91}$$

where

$$\hat{p}_i(x|e^H) \equiv \frac{p_i \phi'_A(y_{i,e^H}) f_i(x|e^H)}{\sum_i p_i \phi'_A(y_{i,e^H}) f_i(x|e^H)} \tag{92}$$

is the ambiguity-biased posterior probability on the occurrence of state i, and

$$R_i \equiv \frac{\phi'_A(y_{i,e^L})}{\phi'_A(y_{i,e^H})}.$$

Differentiating the AALR yields:

$$\begin{aligned} \hat{L}'(x) &= [\hat{p}'_1(x|e^H) L_1(x) + \hat{p}_1(x|e^H) L'_1(x)] R_1 \\ &+ [\hat{p}'_2(x|e^H) L_2(x) + \hat{p}_2(x|e^H) L'_2(x)] R_2. \end{aligned}$$

Notice that $\hat{p}'_1(x|e) \geq 0$ due since $F_1(\cdot|e)$ dominates $F_2(\cdot|e)$ in the LR sense, for each e in $\{e^H, e^L\}$. Hence $\hat{p}'_2(x|e) = -\hat{p}'_1(x|e) \leq 0$. At this point we do not have enough information to conclude on the sign of $\hat{L}'$. To do this, we have to make further assumptions. Consider the following cases.

High effort is effective only in the bad state. In this case, we assume that in the good state $i = 1$, the likelihood ratio is not informative of the effort level take, i.e., $L'_1(x) = 0$ for all $x \in I_x$. Nevertheless, it is very telling of the effort level under the bad state $i = 2$, so that $L'_2(x) \leq 0$ for all $x \in I_x$. We then have:

$$\begin{aligned}\hat{L}'(x) &= [\hat{p}'_1(x|e^H)L_1(x)R_1 \\ &+ [\hat{p}'_2(x|e^H)L_2(x) + \hat{p}_2(x|e^H)L'_2(x)]R_2,\end{aligned}$$

which still has not allowed us to reach a conclusion since the first term of this expression is positive while the second one is negative. If we push a bit further and assume that effort can *perfectly mitigate* the effect of the bad state, i.e., assume that $\frac{f_2(\cdot|e^H)}{f_1(\cdot|e^H)}$ is constant, then $\hat{p}'_1(x) = 0$, implying $\hat{p}'_2(x) = 0$ and so

$$\hat{L}'(x) = \hat{p}_2(x|e^H)L'_2(x)R_2 \leq 0. \tag{93}$$

Thus the wage is increasing in outcome in this case. This implies that $R_2 \geq 1$: the principal puts even higher rewards to the ambiguity-averse agent to motivate hard word.

8 Conclusion

Borrowing the optimal control framework, we reformulate the principal-agent problem as a Mayer's problem to prove the existence of an optimal wage function in the symmetric information case. On the basis of the existence result, we employ the Pontryagin's Maximum Principle to characterize the solution. Our approach, which is most similar to Raviv (1979) represents a contribution to the existing literature in a number of ways. First, we have shown that an optimal wage is non decreasing in outcomes when the principal is risk-averse, regardless of the DMs' attitudes towards ambiguity and the number of ambiguous states. In other words, non decreasing wage is robust to ambiguity aversion when the principal is risk-averse. Second, we do not ex-ante assume an interior solution, which is not an innocuous assumption in presence of ambiguity aversion when the principal is risk-neutral. This is because the expected marginal welfare depends on the shape of the optimal wage function. Had we assumed interior solution, we would have concluded from (78) that constant wage were robust to ambiguity when the principal is neutral to risk and ambiguity while the agent is averse to both, regardless of the structure of ambiguity. Clearly this is not the case even in the case of two ambiguous states considered in Proposition 3.

The main limitation of our research is the generalizability of the result to more than two ambiguous states in the case of a risk-neutral principal. We await future research to shed light on this issue.

Bibliography

L. Berger and V. Bosetti. Ellsberg re-revisited: An experiment disentangling model uncertainty and risk aversion. *FEEM Working Paper*, 2016.

K. Borch. The safety loading of reinsurance premiums. *Scandinavian Actuarial Journal*, 1960(3-4):163–184, 1960.

S. Chakravarty and J. Roy. Recursive expected utility and the separation of attitudes towards risk and ambiguity: an experimental study. *Theory and Decision*, 66(3):199–228, 2009.

F. H. Clarke. *Optimization and nonsmooth analysis*, volume 5. Siam, 1990.

D. Ellsberg. Risk, ambiguity, and the savage axioms. *The quarterly journal of economics*, pages 643–669, 1961.

P. Klibanoff, M. Marinacci, and S. Mukerji. A smooth model of decision making under ambiguity. *Econometrica*, 73(6):1849–1892, 2005.

E. B. Lee and L. Markus. Foundations of optimal control theory. Technical report, Minnesota University Minneapolis Center for Control Sciences, 1967.

A. Raviv. The design of an optimal insurance policy. *The American Economic Review*, 69(1):84–96, 1979.

E. Trélat. *Contrôle optimal: théorie & applications*. Vuibert, 2008.

E. Wolfstetter. *Topics in microeconomics: Industrial organization, auctions, and incentives*. Cambridge University Press, 1999.