Essays on Contract Theory and Organizational Economics

by

Rongzhu Ke

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2009

ARCHIVES

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
OCT 20 2009
LIBRARIES

© Rongzhu Ke
All rights reserved.

The author hereby grants to Massachusetts Institute of Technology permission to reproduce and
to distribute copies of this thesis document in whole or in part.

Signature of Author ....
Department of Economics
15 July 2009

Certified by ....
Bengt Holmstrom
Paul. A. Samuelson Professor of Economics
Thesis Supervisor

Certified by ....
Victor Chernozhukov
Professor of Economics
Thesis Supervisor

Certified by ....
Panle Jia
Assistant Professor of Economics
Thesis Supervisor

Accepted by ....
Esther Duflo
Abdal Latif Jameel Professor of Poverty Alleviation and Development Economics
Chairman, Department Committee on Graduate Students

Essays on Contract Theory and Organizational Economics

by

Rongzhu Ke

Submitted to the Department of Economics
on 15 July 2009, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Chapter 1 develops a non-parametric methodology for identifying contract optimality in the presence of moral hazard. Following the first order approach, a standard method of computing optimal contracts, the paper first proves two theoretical properties of the solutions to the moral hazard problem. First, we show that the profit loss (relative to the optimal contract) for given effort level has a unique lower bound. The second property is an equivalence between the first order condition (Mirrlees-Holmstrom Condition) and the Cramer-Rao Lower Bound (CRLB). These two properties provide the foundations for (1) identifying optimality, and (2) performing statistical inference on the agent's primitives based on an observed sequence of pairs of outputs and payments. The paper shows that under some weak conditions, contract optimality is identified, as long as the output generating process is additive in effort and noise. Identification does not require the agent's effort to be observed by the principal or the econometrician, and requires no knowledge of (1) the details of the contract, (2) the agent's cost of effort, (3) the agent's monetary utility, or (4) the distribution of output. Based on the approach proposed in this paper, we test contract optimality for a piece-rate contract, and estimate bounds on the profit loss for cotton weavers in Zhejiang Province, China.

Chapter 2 develops a new method to justify the validity of the first order approach (FOA). We first prove that checking the validity amounts to checking the existence of a fixed point of the agent's best reponse against a *Mirrlees-Holmstrom ($\mathcal{MH}$) class contract* offered by the principal, given some specifications of complemetary conditions. The main advantage of the current approach is the relaxation of the global concavity of agent utility. We show that under a set of mild conditions, the fixed point approach is applicable and the solution to the principal-agent problem exists. In particular, if the log likelihood ratio is monotonically increasing in output but decreasing in effort, the best response correspondence agrainst a $\mathcal{MH}$ contract has and only has one unique fixed point. Our approach unifies Jewitt's (1988) and Rogerson's (1985) proofs of validity of FOA, and provides a general method to judge validity of FOA. Based on the fixed-point approach, with some additional specifications, we restore Jewitt's (1988) results to situations where the distribution is not convex and the log likelihood ratio is not bounded from below (e.g., normal distribution), or there exists a limited liability constraint.

Furthermore, we generalize our results to a situation where the agent's utility is non-separable. In this fairly general environment, we prove a necessary and sufficient condition for the FOA to be valid, which provides an important method to identify the validity of FOA and compute the solution of the original problem. Finally, we provide a necessary and sufficient condition for a general non-linear bi-level optimization problem to be solvable based on FOA, without a convex constrained set.

Chapter 3 constructs a concrete mechanism/auction to explore the consequence of imposing the ex post participation constraint. The main findings are:

(1) In private good cases (symmetric or asymmetric), we can obtain ex post first best, ex post budget balance, at least interim incentive compatibility and ex post individual rationality (we call it ex post social effiicency), whenever the VCG mechanism runs expected surplus. And the mechanism generating an ex post monotonic payoff is generically unique (up to an ex ante side-pay). In addition, compared with standard auctions, our mechanism generates a risk-free revenue to the seller and ex post invidually rational payoff to the bidders.

(2) In a general preference case with externality, we show there exists an ex post socially efficient mechanism when the number of participants is sufficiently small ($n = 2$). And the choice of mechanism depends on whether the quantity is endogeneous or not.

(3) As an implication, we provide a general discussion on how divisibility, endowment distribution and preference affect the possibility of trade. For the negative result, we show a set of conditions for non-existence of an ex post socially efficient trade, such as either utility is linear or the lowest type agent's utility is independent of his endowment, which can be regarded as stronger version of no-trade theorem (Myerson-Satterthwaite, 1983). This proposition implies non-existence of an ex post socially efficient partner dissolving mechanism. For the positive side, we provide a sufficient condition for existence of ex post socially efficient trade mechanism and show an explicit example.

JEL Code: D86 D81 C12 D82

Thesis Supervisor: Bengt Holmstrom
Title: Paul. A. Samuelson Professor of Economics

Thesis Supervisor: Victor Chernozhukov
Title: Professor of Economics

Thesis Supervisor: Panle Jia
Title: Assistant Professor of Economics

Acknowledgement 1 *I am grateful to my advisors Bengt Holmstrom, Victor Chernozhukov, and Panle Jia for their guidance and supervision. I would also like to thank Jerry Hausman, Hanming Fang, Whitney Newey, Glenn Ellison, Robert Gibbons, Abhijit Banerjee, Anna Mikusheva, Stephen Ryan, Randall Lewis, Michael Powell, Alp Simsek, Moshe Cohen, Konrad Menzel, Florian Ederer, Xianwen Shi, Enying Zheng, and Jialan Wang for their helpful comments. I especially appreciate Jin Li for his insightful suggestions.*

I owe very much to my former advisors and friends, before or after my coming to MIT. Professor Weiying Zhang, Xiangrong Jin, Jinchuan Shi and Li-an Zhou are always encouraging me. I also thank my friends, Li Ding, Xiangsheng Li, Hongjun Zhong, Miaojun Wang, Wenfeng Yang, Hui Shan, Ding Wu, Nancy Qian, and Jin Lu for their helps. There are many names I remember and I wish there was more space to list all of them.

Finally, I thank my parents for their endless love. My wife and parents-in-law are taking care of us at MIT incredibly well. Without their support, nothing is possible.

Chapter 1

Identifying Contract Optimality Non-parametrically with Moral Hazard: First Order Approach and Statistical Inference

1.1 Introduction

This paper explores statistical inference in moral hazard problems. In a principal-agent context with moral hazard, the principal needs to design an incentive scheme to implement a certain level of effort despite the fact that the agent's effort is unobservable (Mirrlees, 1971; Holmstrom, 1979; Grossman and Hart, 1983, *among others*). Recently, there has been increased interest in searching for testable empirical implications of contract theory (see Chiappori and Salanie, 2003, for a survey). However, there is still a gap between direct implications of theory and existing empirical strategies. In particular, empirical data[1] often lacks seemingly relevant information such as (1) the details of the payment schedule, (2) the agent's monetary utility, (3) the agent's cost of effort, or (4) the distribution of output, which hinders the ability to identify the primitives of the model or the contract optimality. As a result, a host of structural assumptions are employed to analyze a particular class (linear) of contracts and optimality is typically assumed[2].

In this paper, we provide a methodology for doing statistical inference without restricting ourselves to linear contracts, without relying on instrument variables, and with relatively weaker assumptions. More precisely, we address the following questions:

- (Q1) Suppose a researcher observes a sequence of output and payment data. Does there exist any possibility to rationalize the data such that the underlying contract is optimal?[3]

[1] For example, Paarsch and Shearer (1999, 2000) and Shearer (1998) evaluate the efficiency of piece rate contracts, Ackerberg and Botticini (1999, 2003) address matching of sharecropping contracts, and there is a large body of literature trying to estimate the effectiveness of CEO compensation (Murphy, 1999 for a survey).

[2] One exception is Paarsch and Shearer (1999, 2000), who use as an instrument information about outside reservation utility (minimum wage law) to estimate a lower bound on the profit loss from utilizing a strict piece rate compensation scheme instead of a more flexible linear contract (with a potentially non-zero intercept). Haley (2003) follows Paarsch and Shearer's approach to estimate the elasticity of the tree logger's response of to piece rates.

[3] Q1 can be rephrased as follows: with the presence of moral hazard, what are the restrictions (identification conditions) on the sequence of data $\{(x_i, w_i)\}_{i=1}^n$ we can test if (*and only if*) the data are generated by an optimal contract, without knowing agent's utility, cost of effort, production function, or contractual form? An analogue of this question in neoclassical microeconomics seems to be the test of the generalized axiom of revealed preference (GARP), where economists ask what the identification condition is if a quantity-price bundle $\{(q_i, p_i)\}_{i=1}^n$ is generated by GARP of a certain individual. However, the question regarding optimal contract is much more difficult. Besides unobservability of utility and cost of effort, there are at least several other important differences: (1) the action of the agent also is unobservable due to moral hazard (in GARP, action q_i is observable); (2) the outside economist may only observe wage w_i, rather than the functional form of contract $s_i(x_i)$ that the agent faces (in GARP, p_i is observed); (3) the researchers's purpose is not to test the agent's individual decision, but to test how the contract optimally incorporates the agent's best response. These difficulties make test of contract optimality a challenging issue.

- (Q2) In the case where we can reject optimality, can we bound the profit loss of the suboptimal contract compared with the unobserved optimal counterfactual?

- (Q3) Are the primitives identified (and under what conditions)?

- (Q4) If there is profit loss, can we identify the mechanism leading to that loss?

These questions are of both theoretical and practical interest in many fields where designing incentive schemes plays a crucial role, such as labor economics, corporate finance[4], income taxation, social insurance and health care (see the survey by Chiappori and Salanie, 2003). In this paper we develop a unified framework to address these issues[5].

The general framework is based on the first order approach (FOA) to solving for the optimal contract (Mirrlees, 1971; Holmstrom, 1979; Rogerson, 1985, Jewitt, 1988, *among others*). In this paper, we first establish a linkage between the first order approach and the Cramer-Rao Lower Bound, allowing us to design a test for the optimality of a contract under some regularity conditions. If optimality is not rejected, we then establish uniqueness of the solution, which allows us to back out parameters of the underlying model, allowing us to carry out statistical inference.

This paper demonstrates that given only data on output and compensation, the econometrician can say whether or not the contract offered to the agent is optimal, without making any parametric assumptions about the utility of money and the disutility of effort, or the distribution of output. Additionally, we can non-parametrically identify the score function of effort (up to a multiplicative scalar) and the agent's inverse marginal utility (up to an affine transformation) under the null hypothesis that the contract is optimal. This allows us to put a lower bound on the profit loss of using a suboptimal contract. The above results hold if (1) the output is generated by a single agent facing the same (potential stochastic) contract and choosing the same level of effort across observations, (2) output is additively separable in effort and an i.i.d. error term, and (3) the agent's utility is separable in money and disutility of effort. As

[4]In corparate finance, CEO compensation is a key question; however, a method to identify the optimality itself is still absent (see Prendergast, 1999; Murphy, 1999 for the survey).

[5]Building off this framework, in a separate paper, we address co-existence of moral hazard and adverse selection (Ke, 2008).

a demonstratration of our approach, we apply our techniques to examine a piece rate contract from a textile firm in Zhejiang Province, China.

This paper is organized as follows: in Part 2, we characterize the moral hazard problem and its solution using FOA, list the regularity conditions and derive the uniqueness of the optimal contract and the equivalence theorem. In Part 3, we develop a parametric testing procedure when the monetary utility and score function are parameterized by unknown parameters, and we also provide a Monte-Carlo simulation for a baseline model. In Part 4, we develop a non-parametric testing procedure when both the monetary utility and score functions are unknown. In Part 5, we extend the results to a heterogeneous data generating process. In Part 6, we provide an empirical example, using piece-rate data collected from Zhejiang Province, China. In Part 7, we conclude and discuss future research. Technical proofs are provided in the Appendices.

1.2 Characterization, Equivalence Theorem and Uniqueness of Lower Bound

1.2.1 Characterization of moral hazard (first order approach)

In a standard moral hazard problem setting (Holmstrom, 1979), there are a principal and an agent, whose monetary utility functions are $v(.) \in \mathcal{V}$ and $u(.) \in \mathcal{U}$, respectively. The agent's output X is randomly distributed over region $\mathcal{X} \subset \mathbb{R}$, with probability density function (p.d.f.) $f(x, a)$, given the agent's effort $a \in \mathbb{A} \subset \mathbb{R}_+$, where $\mathbb{A}$ is an open interval. We assume the support $\mathcal{X}$ does not depend on the effort level a and p.d.f. $f(x, a)$ is continuous and differentiable in a up to some appropriate order. The agent's effort is unobserved by the principal, and the principal makes a take-it-or-leave-it contract $s(x) \in \mathcal{S}$ with $\mathcal{S}$ being a measurable functional space. Assume that the disutility of effort $c(a)$ is separable from the monetary utility. The principal solves the following optimization problem:

$$(P1) \quad \max_{\{a,s(x)\}} \int v(x - s(x)) f(x, a) dx$$

subject to the following individual rationality (IR) and incentive compatibility (IC) constraints for the agent,

$$\int [u(s(x)) - c(a)]f(x,a)dx \geq \underline{U} \tag{IR}$$

$$\int [u(s(x)) - c(a)]f(x,a)dx \geq \int [u(s(x)) - c(\tilde{a})]f(x,\tilde{a})dx, \ \forall a, \tilde{a} \in \mathbb{A} \cup \{0\}, \tag{IC}$$

where $\underline{U}$ is the outside reservation utility, and the choice set $\mathbb{A} \cup \{0\}$ implies that $a = 0$ could be a default choice by the agent.

In theory, because the IC constraint listed above is an infinitely dimensional constraint, under some conditions, we can use the first order condition of IC instead, which is called the first order approach (FOA). The validity of FOA is an important theoretical issue (Mirrlees, 1974; Rogerson, 1985; and Jewitt, 1988 *among others*), but the discussion of it is beyond the scope of this paper (See Ke (2008) for more discussion). Instead, we try to enable the use of empirical data to answer whether FOA is valid based on the procedures to be developed later. Here we make the following regularity assumptions:

A1: Agent is risk averse and the principal is risk neutral;

A2: disutility of effort is increasing and weakly convex in effort a, namely, $c'(a) > 0$, *and* $c''(a) \geq 0$;

A3: expected output is increasing and weakly concave in effort a, namely, (i) $\frac{\partial}{\partial a}\mathbb{E}X > 0$ *and* $\mathbb{E}X < \infty$ any $a < \infty$, *and (ii)* $\frac{\partial^2}{\partial a^2}\mathbb{E}X \leq 0$;

A4: the score is well-defined everywhere, namely $\frac{\partial}{\partial a}\log f(x,a) > -\infty$ *for any* $x \in \mathcal{X}$;

A4′: or the payment is uniformly bounded from below, namely, $s(x) \geq \underline{s} > -\infty$ *for any* $x \in \mathcal{X}$.

A1 is conventional but not a necessary condition. What is necessary is $-\frac{v''(.)}{v'(.)} + (-\frac{u''(.)}{u'(.)}) > 0$, which means that the principal can be risk averse or risk taking. A2 and A3 are standard and particularly, A3-i implies the identifiability of the distribution function, that is, $\mathbb{E}X$ increasing in a implies $f(x,a) \neq f(x,a')$ *for all* $a \neq a'$. A4 may be the most restrictive one, precluding the most familiar normal distribution. But A4 can be dropped if limited liability constraint A4′ holds. The economic intuition behind A4 or A4′ is that the principal cannot punish the agent by a negative infinite fine, even if the agent's performance is bad. Interestingly, A4′ can bring some more information about the identification of the contract optimality, which will be discussed later in a real example (piece rate contract).

For convenience, we define the following regularity conditions.

Definition 1.1: *A moral hazard problem is regular, if A1, A2, A3 and A4 or A4′ are met.*

If moral hazard is regular[6] and the first order approach is valid, then the following conditions are listed (Holmstrom, 1979):

$$\frac{v'(x-s(x))}{u'(s(x))} = \lambda + \mu \frac{f_a(x,a)}{f(x,a)} \tag{1.1}$$

$$\int u(s(x)) f_a(x,a)dx - c'(a) = 0 \tag{1.2}$$

$$\int v(x-s(x)) f_a(x,a)dx + \mu \left(\int u(s(x)) f_{aa}(x,a)dx - c''(a) \right) = 0 \tag{1.3}$$

$$\int [u(s(x)) - c(a)] f(x,a)dx \geq \underline{U}. \tag{IR}$$

And we have the second order condition as:

$$\int u(s(x)) f_{aa}(x,a)dx - c''(a) < 0. \tag{1.4}$$

Equation (1) is the first order condition with respect to contractual form $s(x)$, which is pointwise. Equation (2) is the local IC constraint stating that the unobserved action a should be of the agent's best interest and this constraint will be global if the concavity constraint (4) is met. Equation (3) is the adjoint condition for the action a to be implemented. At the optimal point, the local IC constraint (2) and IR constraint will be binding, and thus the system is solvable if all parameters $(v, u, c, f, \underline{U})$ are known. We denote the solution to the problem (P1) as $V(v, u, c, f, \underline{U})$.

The advantage of the above characterization is its theoretical generality, despite the restriction of the validity of FOA. However, usually, it is not easy to have an explicit solution. Therefore, at the first glance, a comparative statics seem almost unavailable, which may be the primary reason that the above approach is seldom used to analyze the contract in practice. So far, there has been little empirical research directly based on the above theoretical setting.

[6] We are aware that conditions A1-A4' are not sufficient to guarantee the validity of FOA, so we assume it is valid for a moment and then ask whether the real data can justify the validity of FOA.

We come up with an idea to take advantage of the generality of the above approach, in the sense that we can test optimality and bound profit loss without complete knowledge of the agent's utility, cost of effort or even the production function. So we first prove the uniqueness of the lower bound of profit loss, and second build an equivalence theorem between contract optimality and the Cramer-Rao Lower Bound (CRLB) theorem.

1.2.2 Unique Lower Bound of Profit Loss

For convenience, hereafter, let $\frac{v'(.)}{u'(.)} = h(.)$ (according to A1, $\frac{1}{u'(.)} = h(.)$) and $l_a(x,a) \equiv \frac{\partial}{\partial a}\log f(x,a) = \frac{f_a(x,a)}{f(x,a)}$. Throughout this paper, we will put hat to the observed variables versus their counterfacutals. This subsection deals with some important properties of moral hazard problem (P1) with infinite data. Then we will move to the finite data case in later sections. Before proceeding, we define a concept of optimality. Similarly to Grossman-Hart (1983), the moral hazard problem in our characterization can be decomposed into a two-step optimization procedure (Kim, 1995). The first step is to find a minimum expected payment scheme $s^*(x)$ implementing a given effort level a given IC and IR constraint binding. This type of optimality is defined formally as follows.

Definition 1.2: *A contract $w = s^*(x)$ is called the conditional constrained optimum (CCO), if and only if there is no other incentive compatible contract $\tilde{s}(x) \neq s^*(x)$ with positive probability such that (i) $\int(x - s^*(x))f(x,a)dx < \int(x - \tilde{s}(x))f(x,a)dx$ given the same effort level a being implemented, and (ii) the agent has the same utility $\int u(s^*(x))f(x,a)dx - c(a) = \int u(\tilde{s}(x))f(x,a)dx - c(a)$.*

With infinite data, conceptually, we can back out $s(x)$ by observing output and payment (or at least the distribution of w conditional on x), and we also are able to back out $f(x,a)$. We ask whether with infinite data, we can find utility $(u(.), c(.))$ such that observed contract is optimal. We call the data "rationalizable", which is defined as follows.

Definition 1.3: *An observed output payment pair (x,w) is rationalizable if there exists any utility function (u,c) satisfying regularity condition in definition 1 such that $w = s^*(x)$ is the optimal contract consistent with (u,c) and implements p.d.f. of x.*

Difficulty arises because the optimality itself can not be observed directly, so to check whether data is rationalizable, we have to check the first order conditions. The important property is that, if the data satisfies condition (1), say, $\hat{s}(x)$ satisfying $h(\hat{s}(x)) = \hat{\lambda} + \hat{\mu} l_a(x,a)$ for given $h(.)$ and $l_a(x,a)$ then the following theorem shows that $\hat{s}(x)$ is the unique conditional constrained optimal contract, regardless of whether the validity of the first order approach.

Theorem 1 ***Theorem 1.1:*** *If moral hazard problem is regular, then when (infinite) data (x, w) satisfy the first order condition (1), then w is the unique conditional constraint optimal contract implementing the effort a, i.e. p.d.f. $f(x,a)$.*

Proof. Step 1: Identifiability of effort.

By assumption A3, effort $f(x,a)$ is identifiable, i.e., with infinite data, $\int[\log f(x,a) - \log f(x,\hat{a})]f(x,\hat{a})dx = 0$ if and only if $\hat{a} = a$ with probability 1. Therefore, we know observed output is generated by effort a.

Step 2: Effort is a fixed point of the agent's best response correspondence

We know the contract w satisfies first order condition (1), so a is a fixed point, i.e., $a \in \arg\max \int u(w)f(x,a)dx - c(a)$. Therefore, w will be the cheapest contract to implement a. We prove it by contradiction as follows (see chapter 2 for more discussions).

Note that the local IC constraint (2) is the necessary condition for agent's best response. Therefore, suppose there is an optimal contract w^* implementing the same effort level a, and keeping the agent's utility the same as $\underline{U}$, too, we have

$$\begin{aligned} \int u(w^*)l_a(x,a)f(x,a)dx - c'(a) &= 0 = \int u(r(q))l_a(x,a)f(x,a)dx - c'(a) \\ \text{and} \int u(w^*)f(x,a)dx - c(a) &= \underline{U} = \int u(r(q))f(x,a)dx - c(a). \end{aligned}$$

By these two equalities, we have,

$$\int q[u(r(q)) - u(w^*))]f(x,a)dx = 0, \tag{1.5}$$

where $q(x) = \lambda + \mu l_a(x,a)$, $r(.) = h^{-1}(.)$. The profit difference under contract w^* and $r(q)$ is follows:

$$\begin{aligned}
\Delta\Pi &= \int (x - w^*) f(x,a)dx - \int (x - r(q)) f(x,a)dx \\
&= \int (x - w^*) f(x,a)dx + \mu[\int u(w^*) l_a(x,a) f(x,a)dx - c'(a)] + \lambda[\int u(w^*) f(x,a)dx - c(a) - \underline{U}] \\
&\quad -\{\int (x - r(q)) f(x,a)dx + \mu[\int u(r(q)) l_a(x,a) f(x,a)dx - c'(a)] + \lambda[\int u(r(q)) f(x,a)dx - c(a) - \underline{U} \\
&= \int \{r(q) - qu(r(q)) - [w^* - qu(w^*)]\} f(x,a)dx - \int q[u(r(q)) - u(w^*))] f(x,a)dx \\
&= \int \{r(q) - qu(r(q)) - [w^* - qu(w^*)]\} f(x,a)dx.
\end{aligned}$$

Note that $r(q)$ is the unique pointwise minimizer of function $s - qu(s)$ when $q > 0$, and $r(q) = \underline{s}$ when $q < 0$. Therefore, $\Delta\Pi < 0$ when there is positive probability $w^* \neq r(q)$, a contradiction with $\Delta\Pi \geq 0$. Q.E.D. ■

The key point of the above theorem is that we need not check the validity of the first order approach, which has received many concerns. The reason for us to overcome this issue is due to the identifiabity of distribution $f(x,a)$ based on assumption A3, which implies that observed output data is generated by the agent's true best response, rather than just the solution to the local IC constraint (2). Based on this fact, and local IC constraint (2) holding as a necessary condition, we can prove that there does not exist any contract cheaper than the current one, say, w, to implement effort a as we identified through output x.

The above theorem is useful for inference. To rationalize the data, we only need to check whether data fit the first order condition (1). If condition (1) holds, then we can conclude that the contract is optimal given utility function $u(.)$ regardless of the validity of FOA. And if we find $\lambda > 0$ and $\mu > 0$, then the IR constraint and IC contraint should be binding at optimum point.

Given theorem 1, we can deal with the second question, that is whether there exists contract $\hat{s}(x) \neq s(x)$ implementing a and the agent's utility is the same as $\underline{U}$, given utility $(u(.), c(.))$ and effort a being implemented by a contract $s(x)$. If a is implemented by an optimal contract $s(x)$, there does not exist another optimal contract $\hat{s}(x)$ to implement a as well (otherwise, it will be a contridiction with optimality of $s(x)$ by theorem 1). But it is less obvious whether there exists a sub-optimal contract $\hat{s}(x)$ to implement the same effort a. Particularly, we are interested in a certain class of contract, which is called Mirrlees-Holmstrom contract class.

Definition 1.4: *The MH class is all contract collections such that*

$$\mathcal{FC} = \{s(x) \in \mathcal{S} : s(x) = \hat{u}'^{-1}(\frac{1}{\hat{\lambda} + \hat{\mu} l_a(x,a)}), \hat{u} \in \mathcal{U},\ f \in \mathcal{F},\ a \in \mathbb{A},\ \hat{\lambda}, \hat{\mu} \in \mathbb{R}_+\}$$

A contract belonging to the MH contract class is indexed by inverse utility function $\hat{u}'$, distribution function $f(x,a)$, and three positive scalar parameters, a, $\hat{\lambda}$ and $\hat{\mu}$. The key component of this class of contract is the score function $l_a(x,a)$, which can be identified using output data under a very weak condition. For any contract within the feasible class, there exists certain utility function $\hat{u}(.)$ such that under this contract, a will be implemented. The following proposition shows that there only exists a unique utility function $(u(.), c(.))$ rationalizing the contract $\hat{s}(x)$ provided score function $l_a(x,a)$.

Proposition 1.1: *If moral hazard problem (P1) is regular and FOA is valid, given* $s(x)$ *and* $l_a(x,a)$*, there only exists a unique utility function* $(u(.), c(.))$ *rationalizing* $s(x)$ *and keeping the agent the same utility* $\underline{U}$. **(See Appendix A1)**

The idea of proof of proposition 1 is similar to that of theorem 1. Suppose there is another contract $\hat{s}(x) = \hat{r}(\hat{q})$ implementing the same effort level a, and keeping the agent's utility the same as $\underline{U}$, too, we claim if the above equality holds, then $h(.) = \hat{h}(.)$, $\lambda = \hat{\lambda}$ and $\mu = \hat{\mu}$, i.e., $\hat{s}(x) = s(x)$. The intuition is that for any two contracts from the feasible class, if they implement the same effort, and assign the same utility $\underline{U}$ to the agent, then when one of them is optimal, another must be optimal too.

If a contract is optimal, the lower bound of profit loss should be zero. So theorem 1 and proposition 1 can be used to make an inference about the lower bound of the profit loss and the identification of utility as we will see later. Note that the monetary utility $u(s)$ and the disutility of effort $c(a)$ are unobserved by the outside econometrician. The above theorem means that if we find a parameter $(\hat{u}, \hat{c})$ such that (6), (9) and (10) are satisfied based on data $(x_i, w_i)_{i=1}^n$, then we can not find any other utility (u, c) to rationalize the data. So keeping the agent's utility unchanged, the underlying contract which generates $\{w_i\}_{i=1}^n$ is the best contract we can find.

1.2.3 Equivalence of FOC and Cramer-Rao Lower Bound theorem

We suppose a sequence of data $\{x_i, w_i\}_{i=1}^n$ has been observed, where x_i comes from p.d.f. $f(x, a)$ given a, and w_i is the amount of payment generated by an unknown contract, either deterministic or stochastic. For notational convenience, we follow the convention of using subscript i to denote the i-th observation of a random sample, and use $\hat{\mathbb{E}}$ to denote the empirical analogue of expectation $\mathbb{E}$, i.e. $\hat{\mathbb{E}}x_i = \frac{1}{n}\sum_{i=1}^n x_i$. Under the environment of moral hazard, the variation of x_i comes from the exogenous randomization due to either the principal's imperfect measurement of effort or any uncontrollable random effect during production process, even though the effort as a parameter may be invariant. And x_i will drive the variation of w_i through some unknown functional relationship $s(x_i)$ and perhaps some other random factors like entry error or measurement error. For each pair (x_i, w_i), the first order condition (1) can be written as follows[7]:

$$h(w_i) = \mu l_a(x_i,a) + \lambda \tag{1.6}$$

The task is to identify whether or not the underlying contract $w_i = s(x_i)$ is optimal. The following theorem shows an equivalent representation between the first order condition (1) and the Cramer-Rao Lower Bound.

Theorem 1.2: *Assuming data set is i.i.d., then $\frac{1}{n}\sum h(w_i)$ is the best unbiased estimate of λ ($Var[\frac{1}{n}\sum h(w_i)]$ attains CRLB), i.e.*

$$\rho \equiv \frac{Cov(h(w_i), l_a(x_i,a))}{\sqrt{Var(h(w_i))\mathbb{E}[l_a(x_i,a)]^2}} = 1, \tag{1.7}$$

if and only if the first order condition (1) holds with probability 1 associated with $\mu > 0$.

Proof. Since data set is i.i.d., $Var(\frac{1}{n}\sum h(w_i)) = \frac{1}{n}Var(h(w_i))$, we deal with $h(w_i)$.

If part:

In general, suppose $\tau(a)$ is a continuous function of the parameter a, then the Cramer-Rao

[7] In this formula, x_i is not the only random variable. w_i can be generated by any unknown stochastic process too. The only restriction on w_i is that we assume what the researcher observed is the one that the agent has received. If we assume there is error in documentation, like $w_i = w_i^* + \epsilon_i$ and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, our approach is still applicable in the sense that testing $h(w_i^*) = \mu l_a(x_i,a) + \lambda$ by testing $h(w_i) - \frac{1}{2}h''(w_i)\sigma^2 = \mu l_a(x_i,a) + \lambda$. Another possibility is to introduce zero-mean and zero-correlated-to-effort noise $\mathbb{E}_x[\mathbb{E}_\epsilon \epsilon_i / x_i] = 0$ and $\frac{\partial \mathbb{E}_x[\mathbb{E}_\epsilon \epsilon_i / x_i]}{\partial a} \equiv 0$, whose variance asymptotically converges to zero.

Lower Bound of any unbiased estimator of $\tau(a)$ is $\frac{\tau'(a)^2}{\mathbb{E}[l_a(x_i,a)]^2}$. According to this definition, we want to show $Var(h(w_i)) = \frac{\lambda'(a)^2}{\mathbb{E}[l_a(x_i,a)]^2}$ first. Note that $\mathbb{E}l_a(x_i,a) = 0$, so $\frac{1}{n}\sum h(w_i)$ is an unbiased estimator of λ, i.e., $\mathbb{E}h(w_i) = \lambda$. Meanwhile, since the support does not depend on effort, we have

$$\lambda'(a) = \frac{\partial \mathbb{E}h(w_i)}{\partial a} = Cov(h(w_i), l_a(x_i,a)) = \mathbb{E}[h(w_i)l_a(x_i,a)]$$

and $Var\,(l_a(x_i,a)) = \mathbb{E}[l_a(x_i,a)]^2$. From the Cauchy-Schwarz Inequality,

$$Var(X) \geq \frac{[Cov(X,Y)]^2}{Var\,(Y)}$$

we have,

$$Var(h(w_i)) \geq \frac{(\mathbb{E}[h(w_i)l_a(x_i,a)])^2}{\mathbb{E}[l_a(x_i,a)]^2}. \tag{1.8}$$

When the first order condition holds with probability 1, $\mathbb{E}[h\,(w_i)\,l_a(x_i,a)] = \mu\mathbb{E}[l_a(x_i,a)]^2$, we have $\frac{(\mathbb{E}[h(w_i)l_a(x_i,a)])^2}{\mathbb{E}[l_a(x_i,a)]^2} = \mu^2\mathbb{E}[l_a(x,a)]^2$ and

$$Var(h(w_i)) = Var(\lambda + \mu l_a(x_i,a)) = \mu^2\mathbb{E}[l_a(x_i,a)]^2.$$

This means that in inequality (8), the right hand side is equal to the left hand side. Second, it is straightforward that when $\mu > 0$, we have $\rho \equiv \frac{Cov(h(w_i),l_a(x_i,a))}{\sqrt{Var(h(w_i))\mathbb{E}[l_a(x_i,a)]^2}} = 1$. This completes the if part.

Only if part:

$\rho = 1$ implies $Var(h(w_i))$ attaining CRLB, i.e. $Var(h(w_i)) = \frac{Cov(h(w_i),l_a(x_i,a))^2}{Var(l_a(x_i,a))}$, which is true with probability 1 if and only if $h(w_i) = A + Bl_a(x_i, a)$. It turns out $A = \mathbb{E}h(w_i) = \lambda$, since $\mathbb{E}l_a(x_i,a) = 0$, $Var(h(w_i)) = B^2\mathbb{E}[l_a(x_i,a)]^2$ and $\mathbb{E}[h\,(w_i)\,l_a(x_i,a)] = B\mathbb{E}[l_a(x_i,a)]^2$. Finally, by the statement of the theorem, the square root of $Var(h(w_i))$ should be non-negative, which means $B = \mu > 0$. Q.E.D. ■

The above theorem means if data are generated by an optimal contract, then the variation of score should be able to explain variation of compensation completely. We need not know how w_i is generated, we only need to check the relationship between x_i and w_i.

The equivalence indicated by theorem 2 also allows us to link economic theory to statistic inference to recast a new interpretation of contract optimality. It is well known (Holmstrom, 1979; Milgrom, 1981; Holmstrom and Hart, 1987) that the shape of the incentive scheme will depend on the likelihood ratio (or score function) through which the principal can learn about the effort being exerted. A typical well-known insight is that monotone likelihood ratio property (MLRP) implies monotonicity of wage scheme. However, these explanations have not been tested by any real data yet. Based on the above theorem, correspondence between contract optimality and statistical inference can be thought as follows: If an incentive scheme is optimal, then the average marginal incentive cost, say $\frac{1}{n}\sum h(w_i)$, should be the best unbiased estimator of the shadow price of the agent's participation constraint, say, λ. In other words, the optimality means that $\frac{1}{n}\sum h(w_i)$ should contain all information that is useful to estimate the shadow price of the agent's participation of contract.

Meanwhile, the above theorem is also useful for hypothesis testing. First of all, theorem 2 leads to an information matrix type of testing principle (White, 1982) that one can apply to test contract optimality. Note that CRLB is the lower bound of the variance of all unbiased estimators, therefore, theorem 2 builds a bridge between contract theory and statistical inference. To test and identify the agency theory model is then to test and identify whether the marginal incentive cost $h(w_i)$ is perfect linearly correlated to the score function $l_a(x_i, a)$. Meanwhile, for equality (6), note that neither nominator nor denominator depends on cost function $c(a)$, shadow prices λ and μ; this means we can test the equality (6) without knowing $c(a)$, λ and μ. The inference therefore is made by testing whether the correlation coefficient between $h(w_i)$ and $l_a(x_i,a)$ is significantly close to 1[8].

Besides equation (1), there are some issues worth noting. Firstly, the existence of an optimal contract requires $\mu > 0$, or equivalently, the $Cov(h(w_i), l_a(x_i,a))$ to be positive. As Jewitt (1988) shows, $\mu > 0$ comes from condition (2), which can be regarded as the agent's best response to the contract. The justification of condition (2) is fairly weak, namely, the agent's action space $\mathbb{A}$ is compact. And when IC constraint holds, (2) is legitimate, without assuming first order stochastic dominance. Since we do not know the cost function, we can not justify $\mu > 0$ by

[8] In empirical research, $h(w_i)$ and $l_a(x_i,a)$ might not be completely known. But as we will see in the next section, there still exist some statistical procedures to estimate ρ, and test $\rho = 1$.

solving (P1). But to justify $\mu > 0$, it suffices to test that

$$M \equiv \int u(w) f_a(x,a) dx > 0 \tag{1.9}$$

Secondly, the second order condition (4) is sufficient IC constraint to be necessary (FOA valid at least locally). For $c''(a) \geq 0$, we want the following inequality to hold:

$$K \equiv - \int u(w) f_{aa}(x,a) dx > 0. \tag{1.10}$$

Our goal is to use the data to test (6), (9) and (10). The next section will start to deal with these issues.

1.3 Identification When Monetary Utility and Score Function Is Parameterized by Unknowns Parameters

This section uses parametric technique. We assume that all observations are homogeneous, generated by the same contract, cost function, utility, and production function, based on the same effort. In terms of statistics, this means that output $\{x_i\}_{i=1}^n$ is i.i.d. drawn from p.d.f. $f(x,a)$ for a fixed unknown effort a. The variation in output x_i here only comes from the random factor during the production process. Wage w_i varies in response to output x_i and possibly to other unobserved factors. We will relax homogeneity assumption in section 6, allowing heterogeneity in the data generating process. And through out this paper, we assume that the cost function $c(a)$ and the outside reservation $\underline{U}$ are unknown[9], but the restriction on output distribution and monetary utility will be relaxed step by step. In this section, we assume p.d.f. $f(x,a)$ is unknown but the score function can be parameterized. We provide a base line model first, and extensions followed.

[9]Unknown means we do not know the functional form except for some qualitive properties like concavity/convexity, increasing/decreasing, etc.

1.3.1 Benchmark: Known Monetary Utility and A Sufficient Statistic of Effort

In the benchmark case, we assume the agent's monetary utility is completely known, and the production function unknown but there exists a sufficient statistic for the effort. To understand the implication of the sufficient statistic, we can decompose the p.d.f. of output as follows:

$$f(x,a) = \frac{e^{\omega(x,a)}v(x)}{\int e^{\omega(x,a)}v(x)dx}, \tag{1.11}$$

where the real valued function $v(x) > 0$, and $v(x)$ and $\omega(x,a)$ such that $\int e^{\omega(x,a)}v(x)dx < \infty$.

When a sufficient statistic for the effort level a exists, $\omega(x,a) = \omega_1(x)\omega_2(a)$ with $\omega_1(x)$ being known though $\omega_2(a)$ and $v(x)$ remain unknown. In this case, by the factorization theorem, $\frac{1}{n}\sum_{i=1}^{n}\omega_1(x)$ contains all information about effort a. An useful example of sufficient statistic is that $\omega_1(x) = x$, which represents the exponential family according to Brown (1986).

We also assume the information quantity exists:

A5: Variance of the score is finite, namely, $Z \equiv \mathbb{E}l_a(x_i,a)^2 < \infty$ *for* $a < \infty$.

A5 is fairly general, and employed throughout this paper.

Test for the Optimality

We can run a three-step procedure.

(1) Estimate the score function $l_a(x,a)$.

When $\omega(x,a) = \omega_1(x)\omega_2(a)$, from (11), we have,

$$l_a(x,a) = \omega_2'(a)[\omega_1(x) - \mathbb{E}\omega_1(x)]$$

Since a is a parameter of $f(x,a)$, likelihood equality implies that $\mathbb{E}l_a(x,a) = 0$. Therefore, the effort level a can be determined by moment condition $T(a) \equiv \mathbb{E}\omega_1(x)$ since $\mathbb{E}\omega_1(x)$ is a monotonically increasing function of a. The moment estimator of T is,

$$\widehat{T} = \frac{1}{n}\sum_{i=1}^{n}\omega_1(x).$$

Note $\widehat{T} = \frac{1}{n}\sum_{i=1}^{n}\omega_1(x)$ is a consistent estimator of $\mathbb{E}\omega_1(x)$[10], so that $l_a(x,a)$ can be estimated consistently up to a scalar parameter $\omega_2'(a)$. Since $\omega_2'(a)$ does not affect the asymptotic z-value of estimator $l_a(x,\hat{a})$, we normalize $\omega_2'(a)$ as $\frac{1}{Var(\omega_1(x))}$ in the following context. Accordingly, the asymptotic distribution of T is,

$$\sqrt{n}(\hat{T} - T) \to^d \mathcal{N}(0, Var(\omega_1(x_i))).$$

And for each fixed x, we have,

$$\sqrt{n}(l_a(x,\hat{a}) - l_a(x,a)) \to^d \mathcal{N}(0, Z^{-1})$$

where $Z = \mathbb{E}l_a^2$ is information quantity, and $\hat{Z} = \frac{1}{n-1}\sum_{i=1}^{n} l_a(x_i,\hat{a})^2$ is its consistent estimator. The asymptotic distribution of $\hat{Z}$ follows,

$$\sqrt{n}(\hat{Z} - Z) \to^d \mathcal{N}(0, Var(l_a^2(x_i,a)) + 4Z^{-1}[\mathbb{E}l_a(x_i,a)l_{aT}(x_i,a)]^2) = \mathcal{N}(0, Var(l_a^2(x_i,a))).$$

where $l_{aT}(x_i,a) = \frac{\partial}{\partial T}l_a(x_i,a)$ (hereafter, we use a subscript to denote partial derivative, throughout this analysis).

(2) Test significance of moral hazard and the second order condition.

Let $\hat{M} = \frac{1}{n}\sum_{i=1}^{n} u(w_i)l_a(x_i,\hat{a})$, by the continuous mapping theorem,

$$\sqrt{n}(\hat{M} - M) \to^d \mathcal{N}(0, nAVar(\hat{M}))$$

with $nAVar(\hat{M}) = Var(u(w_i)l_a(x_i)) + Z^{-1}[\mathbb{E}(u(w_i)l_{aT}(x_i))]^2$.

To test the second order condition, we need to know the object

$$\hat{K} = -\frac{1}{n}\sum_{i=1}^{n} u(w_i)[l_{aa}(x_i,\hat{a}) + l_a^2(x_i,\hat{a})],$$

where $l_a^2(x_i,\hat{a})$ is estimated, but $l_{aa}(x_i,a) = -\mathbb{E}l_a^2 - \frac{\omega_2''(a)}{\omega_2'(a)}l_a$ contains one unknown parameter

[10] If the production functional form $f(x,a)$ is known, we are able to estimate the unobservable effort a by a standard econometric process, like maximum likelihood estimate (MLE) or generalized moment method (GMM). There are many methods to estimate a. The conditions for MLE consistency and asymptotic normality have been well-studied in the statistical literature (See Casella and Berger, 2002 Chapter 10 for the discussion).

$\frac{\omega_2''(a)}{\omega_2'(a)}$ that can not be estimated parametrically. To circumvent this barrier, we derive a sufficient condition for $K > 0$. Note that if $\mathbb{E}x_i$ is concave in a, then, the sufficient condition for $K > 0$ is

$$\frac{\mathbb{E}x_i(l_a^2 - Z)}{\mathbb{E}x_i l_a} M > \mathbb{E}u(w_i)(l_a^2 - Z).$$

Let $\mathcal{K} = \mathbb{E}x_i(l_a^2 - Z)M - \mathbb{E}u(w_i)(l_a^2 - Z)\mathbb{E}x_i l_a$, thus $\mathcal{K}$ can be estimated by $\hat{\mathcal{K}} = \hat{\mathbb{E}}x_i(\hat{l}_a^2 - \hat{Z})\hat{M} - \hat{\mathbb{E}}u(w_i)(\hat{l}_a^2 - \hat{Z})\hat{\mathbb{E}}x_i\hat{l}_a$ whose asymptotic distribution is

$$\sqrt{n}(\hat{\mathcal{K}} - \mathcal{K}) \to^d \mathcal{N}(0, nAVar(\hat{\mathcal{K}}))$$

where

$$\begin{aligned} nAVar(\mathcal{K}) &= Var\{x_i(l_a^2 - Z)M + ul_a\mathbb{E}x_i(l_a^2 - Z) - u(l_a^2 - Z)\mathbb{E}x_i l_a - x_i l_a \mathbb{E}u(w_i)(l_a^2 - Z)\} \\ &\quad + Z^{-1}\{2\mathbb{E}x_i l_a l_{aT} M + \mathbb{E}x_i(l_a^2 - Z)M_T - 2\mathbb{E}u(w_i)l_a l_{aT}\mathbb{E}x_i l_a - \mathbb{E}u(w_i)(l_a^2 - Z)\mathbb{E}x_i l_{aT}\}^2 \end{aligned}$$

Using the sample analogue of $AVar(\hat{M})$ and $AVar(\hat{\mathcal{K}})$, we can test the hypotheses (one-sided):

$$H_0^M : M(a) > 0$$

and

$$H_0^K : K(a) > 0.$$

(3) Test equation (6), contract optimality.

Let

$$\hat{Q} = \frac{1}{n-1}\sum_{i=1}^{n}[h(w_i) - \hat{\mathbb{E}}h(w_i)]^2, \tag{1.12}$$

and

$$\hat{J} = \hat{\mathbb{E}}(h(w_i) - \hat{\mathbb{E}}h(w_i))l_a(x_i, \hat{a}) \tag{1.13}$$

be the sample analogue of $Var(h(w_i)$ and $Cov(h(w_i), l_a(x_i, a))$ respectively, and let

$$\hat{\rho} = \frac{\hat{J}}{\sqrt{\hat{Z}\hat{Q}}} \tag{1.14}$$

be the estimator of correlation coefficient between $h(w_i)$ and $l_a(x_i, a)$. If $\hat{\rho}$ significantly approaches 1, if and only if the first order condition holds. We assume moments of $h(w_i)$ and $l_a(x_i, a)$ exist up to some appropriate order (typically the fourth moments exist $\mathbb{E}h^4 < \infty$ and $\mathbb{E}l_a^4 < \infty$). We have the following proposition.

Proposition 1.2: *To test the first order condition (1), is to test $\hat{\rho} \rightarrow^p 1$, where $\hat{\rho}(\hat{T}) = \frac{\hat{J}(\hat{T})}{\sqrt{\hat{Z}(\hat{T})\hat{Q}}}$ has the following asymptotic distribution:*

(i) if $\rho < 1$,

$$\sqrt{n}(\hat{\rho}(\hat{T}) - \rho(T)) \rightarrow^d \mathcal{N}\left[0, \frac{1}{4QZ}Var(2(h-\lambda)l_a - \frac{J}{Q}(h-\lambda)^2 - \frac{J}{Z}l_a^2)\right]$$

(For convenience, in the final expression of asymptotic objects, we suppress the arguments w or x if there is no confusion).

(ii) If $\rho = 1$, under the null hypothesis:

$$2n(1 - \hat{\rho}(\hat{T})) \rightarrow^d \chi_1^2.$$

Proof. (See Appendix A2) ■

We can either use the sample analogues or bootstrap to estimate $nAVar(\hat{\rho})$.

Estimate the profits loss and agency costs

We can also estimate the loss of efficiency due to misspecification of the contract according to definition 2. In the benchmark case, since utility is known, we can have a consistent estimator of the profit loss, in comparison with the CCO.

Proposition 1.3: *Compared with the CCO, (i) the profit loss of the observed contract is estimated by*

$$\widehat{\Delta\Pi} = \hat{\mathbb{E}}[x_i - (\widehat{s^*}(x_i)] - \hat{\mathbb{E}}[x_i - w_i] = \hat{\mathbb{E}}w_i - \hat{\mathbb{E}}h^{-1}[\widehat{\lambda^*} + \widehat{\mu^*}l_a(x_i, \hat{a})]$$

where $\hat{a}$ is the consistent estimator of the effort a, and $(\widehat{\lambda^*}, \widehat{\mu^*}, \widehat{s^*}(x_i))$ solves the following *equations*

$$\begin{aligned} h^{-1}[\widehat{\lambda^*} + \widehat{\mu^*} l_a(x_i, \hat{a})] &= \widehat{s^*}(x_i) \text{ for almost every } x_i, \\ \hat{\mathbb{E}} u(\widehat{s^*}(x_i)) l_a(x_i, \hat{a}) &= \hat{\mathbb{E}} u(w_i) l_a(x_i, \hat{a}), \\ \text{and } \hat{\mathbb{E}} u(\widehat{s^*}(x_i)) &= \hat{\mathbb{E}} u(w_i); \end{aligned}$$

(ii) the asymptotic distribution of $\widehat{\Delta\Pi}$ is,

$$\begin{aligned} \sqrt{n}(\widehat{\Delta\Pi} - \Delta\Pi) &\rightarrow {}^d\mathcal{N}(0, nAVar(\widehat{\Delta\Pi})) \text{ if } \Delta\Pi = \mathbb{E}w_i - \mathbb{E}s^*(x_i) > 0, \\ \text{or } \frac{\widehat{\Delta\Pi}^2}{AVar(\widehat{\Delta\Pi})} &\rightarrow {}^d\chi_1^2 \text{ if } \Delta\Pi = 0; \end{aligned}$$

(iii) the validity of FOC at $s^*(x)$ is justified by testing (9) and (10) based on the following asymptotic distributions:

$$\begin{aligned} \sqrt{n}(\widehat{M^*(\hat{a})} - M^*(a)) &\rightarrow {}^d\mathcal{N}(0, nAVar(\widehat{M^*(\hat{a})})) \\ \text{and } \sqrt{n}(\widehat{K^*(\hat{a})} - K^*(a)) &\rightarrow {}^d\mathcal{N}(0, nAVar(\widehat{K^*(\hat{a})})) \end{aligned}$$

where,

$$\widehat{M^*(\hat{a})} = \hat{\mathbb{E}} u(\widehat{s^*}(x_i)) l_a(x_i, \hat{a})$$

and

$$\widehat{K^*(\hat{a})} = -\hat{\mathbb{E}} u(\widehat{s^*}(x_i))[l_a^2(x_i, \hat{a}) + l_{aa}(x_i, \hat{a})].$$

(Proof and the expression of asymptotic variances are in Appendix A3)

Using the sample analogues, we can have form $n\widehat{AVar}(\widehat{\Delta\Pi})$ as the consistent estimator of $nAVar(\widehat{\Delta\Pi})$. Thus the profit loss can be estimated[11].

Additionally, based on the above derivation, we are also able to estimate the agency cost AC, which is the distance between the first best contract and the second best contract, given

[11]We can not test $\widehat{K^*(\hat{a})} > 0$ directly, but we can test it based on $\widehat{\mathcal{K}^*(\hat{a})} > 0$.

the same utility and the same effort a being implemented. The theoretical expression of AC is,

$$AC = \int s^*(x) f(x,a) dx - w^{fb},$$

where w^{fb} solves $u(w^{fb}) = \int u(s^*(x)) f(x,a) dx$ and $s^*(x)$ is the second best contract. Meanwhile, it is not difficult to have an estimator of the profit distance between observed data and the potential first best contract, which represents the total profit loss due to misspecification of the optimal contract and the agency cost.

Corollary 1.1: *Conditional on the effort level a, (i) the agency cost is estimated by*

$$\widehat{AC} = \hat{\mathbb{E}}\widehat{s^*}(x_i) - \hat{w}^{fb},$$

where $\widehat{s^*}(x)$ is estimated CCO and $\hat{w}^{fb}$ solves $u(\hat{w}^{fb}) - c(a) = \hat{E}u(w_i) - c(a)$, and the asymptotic distribution of $\widehat{AC}$ is

$$\sqrt{n}(\widehat{AC} - AC) \to^d \mathcal{N}(0, nAVar(\widehat{AC})).$$

(ii) The total profit loss compared with the potential first best contract is

$$\widehat{TL} = \hat{\mathbb{E}}w_i - u^{-1}\left(\hat{\mathbb{E}}u(w_i)\right),$$

and the asymptotic distribution of $\widehat{TL}$ is

$$\sqrt{n}(\widehat{TL} - TL) \to^d \mathcal{N}(0, nAVar(\widehat{TL})),$$

where $TL = \int wf(x,a)dx - u^{-1}\left(\int u(w) f(x,a) dx\right)$. **(Proof see Appendix A4)**

It should be noted that $\frac{1}{n}\sum w_i - w^{fb}$ is the distance to the conditional first best contract, given the effort level a. It is difficult to obtain the unconditional profit loss since alternatives of effort level remain unknown based on a single sequence of data. Despite this difficulty, we can partially identify whether the output is higher or lower than the unconditional optimal effort level by testing the adjoint equation (3). We will discuss this issue in section 4.

A simulated example 1: $u = 2\sqrt{w}$, $l_a(x,a) = \frac{x-a}{Var(x)}$

We provide a Monte Carlo experiment as follows. The theoretical counterpart of this example is considered by Holmstrom (1979). In this example, suppose the utility function $u = 2\sqrt{w}$ is known and average output is a sufficient statistic for effort, but p.d.f. itself and the contract details are also unknown. We do Monte Carlo simulations to simulate 1000 sequences of data, with each containing 100 observations. The results are reported in table 1. It can be seen, in both contract, $\hat{M}$ and $\hat{\mathcal{K}}$ are significantly positive in most trials. $\hat{M}$ is positive at 0.1 significant level in all simulations. $\hat{\mathcal{K}}$ is positive at 0.1 significant level in 99.7% of simulations in contract A and 98.0% in contract B (the rest are at least 0.2 significance level). For contract A, $\hat{\rho}$ are significantly close to 1, the finite sample z value of $1 - \hat{\rho}$ is almost zero under alternative hypothesis, and p value of chi-square test is almost one under the null hypothesis. For contract B, though $\hat{\rho}$ value looks high, it significantly deviates from zero, indicated by both z test and chi-square test[12]. Monte Carlo simulations indicates that $\hat{\rho}$ test behaves very well in detecting contract optimality, although in contract B, the profit losses are actually not significant (up to 2.09% per capita loss).

[12] The true contract in A is optimal contract, $s(x) = (\frac{x}{10} + 1)^2$, while the true contract in B is a piece-rate contract $s(x) = 0.5093x$. Given the same utility (u, c), the piece rate contract will implement effort level $a^{BR}(B) = 10$, the same as contract A does (See Technical Supplement 2 for details).

Table 1. Testing Results for Monte Carlo Simulations

	Contract A				Contract B			
variables	Med.	s.d.	% of 0.05	% of 0.10	Med.	s.d.	% of 0.05	% of 0.
$\hat{\mu}$	9.6063	2.9498			9.7701	2.1270		
z val. $\hat{\mathcal{M}}$	2.1724	0.1938	98.7	100	2.4271	0.2035	99.9	100
z val. $\hat{\mathcal{K}}$	2.8836	1.2244	96.8	99.7	2.2089	0.6976	86.1	98.0
$\hat{\rho}=1$	1.0000	9.5165e-017			0.9610	0.0083		
z val 1-$\hat{\rho}$	0.0000	1.8918e-013	100		32.5455	4.0758	0	
p of χ^2-test	1.0000	6.6173e-008	100		0.0052	0.0053	0	
$\widehat{\Delta\Pi}(\%)$	0.38	0.0011			2.09	0.0050		
z val. $\widehat{\Delta\Pi}$	0.0246	0.0019	100		0.1317	0.0150	100	
p of χ^2-test	0.9804	0.0015	100		0.8952	0.0119	0	33.8

Note: The fourth (eighth) column displays the percentage of simulations where null hypothesis can not be rejected at 0.05 significance level, and the fifth (ninth) column displays that at 0.1 significance level.

1.3.2 Unknown Parameters in $u(w,\gamma)$ and $l_a(x,\mathbf{T})$: Incentive, Selection and Matching

This subsection extends the benchmark model to a situation in which parameters appear in monetary utility and score function that can not be estimated through the first stage estimation. Suppose the monetary utility $u(w,\gamma)$ can be parameterized by unknown parameters $\gamma \in \mathbf{\Gamma} \subset \mathbb{R}^{|\mathbf{\Gamma}|}$, and $l_a(x,a)$ can be parameterized as $l_a(x,a) = \omega_a(x,\mathbf{T}) - \mathbb{E}\omega_a(x,\mathbf{T})$[13] where $\mathbf{T} \equiv (a,\boldsymbol{\theta})$, and $\boldsymbol{\theta} \in \mathbf{\Theta} \subset \mathbb{R}^{|\mathbf{\Theta}|}$. We assume that $\mathbf{\Theta}$ and $\mathbf{\Gamma}$ are compact. In the empirics, it is common that the agent's risk aversion coefficient is unknown, such as in constant relative

[13] In this formula, the likelihood equality still applies. For example, in Logistic distribution, $\omega_a(x,\boldsymbol{\theta},a) = \frac{2e^{-\frac{x-a}{\theta}}}{1+e^{-\frac{x-a}{\theta}}}$, the moment estimator $\frac{1}{n}\sum\omega_a(x,\boldsymbol{\theta},a)$ provided θ will implies MLE of effort. But in general, given the value of $(\boldsymbol{\theta},a)$, the MLE of a can not be solved by the equation $\frac{1}{n}\sum\omega_a(x,\boldsymbol{\theta},a) = \mathbb{E}\omega_a(x,\boldsymbol{\theta},a)$ provided $\boldsymbol{\theta}$. One particular example is curved exponential family. For example, if $f(x,a)$ is the curved normal, then $l_a(x) = \frac{1}{a^3}(x^2-ax-a^2)$ here,$\omega_a(x,\boldsymbol{\theta},a) = x^2-ax$, which implies that $\frac{1}{n}\sum\omega_a(x,\boldsymbol{\theta},a)$ is unbiased estimator of $\mathbb{E}[x^2-ax]$ for given a, but the fact that $\hat{a}$ solves $\frac{1}{n}\sum l_a(x,a)=0$ does not result in consistent estimator of a. So the specification here is weaker than the requirement that MLE of a is implied by $\frac{1}{n}\sum l_a(x,a)=0$. Of course, if $\omega_a(x,a,\theta)$ does not contains a (or a only enters ω_a as a multiplicable term), this means $\omega_a(x,\theta)$ is a sufficient statistic of a, given $\boldsymbol{\theta}$, then in this case MLE of a is implied by $\frac{1}{n}\sum l_a(x,a)=0$.

risk aversion (CRRA) or constant absolute risk aversion (CARA). The unknowns in the utility and production functions bring in a set of identification issues on whether these unknowns are endogenous or exogenous to the principal's choice of contract. In empirics, this issue is called matching. We deal with these issues in next subsections step by step.

Test for the optimality and Bound of Losses for Given Type of Agent

Before proceeding to derivations, we make the following identification condition:

A6: Distribution function of output is identifiable, namely, $f(x, a, \boldsymbol{\theta}) \neq f(x, \tilde{a}, \tilde{\boldsymbol{\theta}})$ for $(a, \boldsymbol{\theta}) \neq (\tilde{a}, \tilde{\boldsymbol{\theta}}) \in \mathbb{A} \times \boldsymbol{\Theta}$.

In order to do estimation, we form a criterion function as follows:

$$\Lambda(\gamma, \mathbf{T}) = \mathbb{E}\left(\frac{h(w_i, \gamma) - \mathbb{E}h(w_i, \gamma)}{\mathbb{E}h(w_i, \gamma) l_a(x, \mathbf{T})} - \frac{l_a(x_i, \mathbf{T})}{\mathbb{E}l_a(x_i, \mathbf{T})^2}\right)^2, \tag{1.15}$$

which will be non-negative and achieve minimum value zero if and only if the first order condition (1) holds with probability one[14]. For convenience, denote $\delta \equiv (\gamma, \mathbf{T})$. Note that $\Theta \times \Gamma$ is compact, so there exist some δ minimizing $\Lambda(\gamma, \mathbf{T})$. We can compute $\rho(\delta)$ value corresponding to the minimizer δ and test the null hypothesis $H_0^\rho : \rho(\delta) = 1$.

We denote the sample analogue of $\Lambda(\gamma, \mathbf{T})$ as

$$\hat{\Lambda}(\gamma, \mathbf{T}) = \hat{\mathbb{E}}\left(\frac{h(w_i, \gamma) - \hat{\mathbb{E}}h(w_i, \gamma)}{\hat{\mathbb{E}}h(w_i, \gamma) l_a(x_i, \mathbf{T})} - \frac{l_a(x_i, \mathbf{T})}{\frac{1}{n-1}\sum l_a(x_i, \mathbf{T})^2}\right)^2. \tag{1.16}$$

For the uniform convergence theorem to be applicable, we also make the following assumption.

Assumption 7: $\sup_{\boldsymbol{\delta}} \left\| \frac{\partial^2 \hat{\Lambda}(\boldsymbol{\delta})}{\partial \delta \partial \delta'} - \frac{\partial^2 \Lambda(\boldsymbol{\delta})}{\partial \delta \partial \delta'} \right\| \to^p 0$ and $\sqrt{n}\hat{\Lambda}_{\boldsymbol{\delta}}(\delta) \to^d \mathcal{N}(0, \Omega)$, *and* $\frac{\partial^2 \Lambda(\boldsymbol{\delta})}{\partial \delta \partial \delta'}$ *are non-singular (assume Lindberg-Feller or Liapounov Central Limit Theorem is applicable), where $\Lambda(\delta)$ is defined as formula (15).*

We have the following theorem:

Theorem 1.3: *If the moral hazard problme is regular, FOA is valid, and A5-A6 hold, when*

[14] When the first order condition (1) holds, the distance between the two sides of equation (1) should be zero, namely, $\mathbb{E}[\lambda + \mu l_a(x) - h(s(x), \gamma))]^2 = 0$. In order to avoid the trivial solution that $\mu = 0$ and $h(s(x), \gamma))$ is a constant, we re-scale the objective function as (15). Additionally, there are many criterion functions that can be employed, like the L_∞ norm or others. Here, for simplicity we use the L_2 norm as an example. The comparision between different criterion functions is not the primary purpose of this paper.

score function $\omega_a(x,T)$ is known (up to some unknown parameter(s) $T \in \Theta \times A$), and the utility function $u(w,\gamma)$ is known up to unknown parameter(s) γ, then

(i) contract $s(x)$ is optimal only if hypothesis $H_0^\rho : \rho(\gamma, T) = 1$ is accepted, where $\rho(\gamma,T)$ is approximated by

$$\hat{\rho}(\hat{\gamma}, \hat{\mathbf{T}}) = \frac{\widehat{Cov}(\omega_a(x, \hat{\mathbf{T}}), h(w, \hat{\gamma}))}{\sqrt{\widehat{Var}(\omega_a(x, \hat{\mathbf{T}}))\widehat{Var}(h(w, \hat{\gamma}))}}$$

with $(\hat{\gamma}, \hat{T}) \in \arg\min_{(\gamma,\mathbf{T})} \hat{\Lambda}(\gamma, T)$.

(ii) With the additional assumption A7, $\hat{\rho}(\hat{\gamma}, \hat{T})$ has asymptotic distribution,

$$\sqrt{n}(\hat{\rho} - \rho) \rightarrow^d \mathcal{N}(0, nAVar(\hat{\rho})) \text{ if } \rho < 1,$$

or

$$n(1-\hat{\rho}) \rightarrow^d \mathcal{Q}\left[\frac{1}{2}\begin{pmatrix} \sqrt{n}\hat{\mathbb{E}}l_a \\ \sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \end{pmatrix}' \begin{pmatrix} Z^{-1} & 0 \\ 0 & -\rho_{\boldsymbol{\delta}\boldsymbol{\delta}'} \end{pmatrix} \begin{pmatrix} \sqrt{n}\hat{\mathbb{E}}l_a \\ \sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \end{pmatrix}\right] \quad \text{if } \rho = 1;$$

where Q is a quadratic form of multivariate normal distribution with degree of freedom $r+t+1$, and

$$\sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \rightarrow^d \mathcal{N}\left(0, [\Lambda_{\boldsymbol{\delta}\boldsymbol{\delta}'}(\boldsymbol{\delta})]^{-1}\Omega[\Lambda_{\boldsymbol{\delta}\boldsymbol{\delta}'}(\boldsymbol{\delta})']^{-1}\right)$$

with $Z = El_a^2$, $Z^{-1} = nAVar(\hat{T})$ and $\Gamma = nAVar(\hat{\gamma})$. **(See Appendix A5 for the proof)**

The intuition behind the above test is that we pick the parameters which make the data most likely to be consistent with an optimal contract. If these parameters can not support our hypothesis, then there exists no possibility for the contract to be optimal within the whole parameter space, i.e., data are not rationalizable. This test is conservative in the sense that we might fail to reject optimality when the contract is not optimal given the unobserved true tuple (u, f, c). In other words, δ_0 solving $\rho(\delta_0) = 1$ needs not to be the parameter(s) consistent with the true utility. However, under some conditions, δ_0 can be the true primitive parameters under the null hypothesis. These conditions will be discussed below and proved more generally in section 4.

Remark 1.1: *If the p.d.f. of output, $f(x, a, \theta)$ has some unknown parameters θ in terms of $f(x, \tau(a, \theta))$, and $f(x, \tau) \neq f(x, \tilde{\tau})$ if $\tau \neq \tilde{\tau}$, then assumption A6 may not hold. In this case,* *we can test the first order condition (1) by* $E[h(w_i)l_\tau(x_i,a)] = \sqrt{\mathbb{E}[l_\tau(x_i,a)]^2}\sqrt{Var(h(w_i))}$ *and treat* $l_\tau(x_i,a)$ *as* $l_a(x_i,a)$ *without knowing* $a(\theta)$.

The further task is to estimate the loss of profit. Because the utility is not completely known, we are only able to bound the losses. Note that the profit loss will be a function of unknown parameters δ, written as,

$$\Delta\Pi(\boldsymbol{\delta}) = \mathbb{E}w_i - \mathbb{E}s^*(x_i,\boldsymbol{\delta}),$$

where $s^*(x_i,\boldsymbol{\delta})$ is the solution to problem (P1) given effort a.

We can find the lower bound and upper bound of $\Delta\Pi(\boldsymbol{\delta})$ within parameter space $\boldsymbol{\Gamma}\times\boldsymbol{\Theta}$, and subject to the constraints

$$\int u(s^*(x,\boldsymbol{\delta}),\gamma)l_a(x,T)f(x,a)dx = \int u(w,\gamma)l_a(x,T)f(x,a)dx, \tag{1.17}$$

$$\text{and} \int u(s^*(x,\boldsymbol{\delta}),\gamma)f(x,a)dx = \int u(w,\gamma)f(x,a)dx, \tag{1.18}$$

where the first constraint means that $s^*(x,\boldsymbol{\delta})$ and w implement the same effort level a, and the second one means that the agent is indifferent between $s^*(x,\boldsymbol{\delta})$ and w[15]. So we construct the Lagrangian as follows,

$$\begin{aligned} & L^*(\boldsymbol{\delta},\lambda^*,\mu^*) \\ = & \int [w - h^{-1}(\lambda^* + \mu^*[\omega_a(x,\mathbf{T}) - \mathbb{E}\omega_a(x_i,\mathbf{T})])f(x,a)dx \\ & +\lambda^*[\int u(h^{-1}(\lambda^* + \mu^*[\omega_a(x,\mathbf{T}) - \mathbb{E}\omega_a(x_i,\mathbf{T})]),\gamma)f(x,a)dx - \int u(w,\gamma)f(x,a)dx] \\ & +\mu^*\left[\begin{array}{c} \int u(h^{-1}(\lambda^* + \mu^*[\omega_a(x,\mathbf{T}) - \mathbb{E}\omega_a(x_i,\mathbf{T})]),\gamma)[\omega_a(x,\mathbf{T}) - \mathbb{E}\omega_a(x_i,\mathbf{T})]f(x,a)dx \\ -\int u(w,\gamma)[\omega_a(x,\mathbf{T}) - \mathbb{E}\omega_a(x_i,\mathbf{T})]f(x,a)dx \end{array}\right]. \end{aligned}$$

Because Θ and Γ are compact, we can find a lower bound and an upper bound of the profit loss by choosing δ, that is, $(\boldsymbol{\delta}_{\min},\lambda^*(\boldsymbol{\delta}_{\min}),\mu^*(\boldsymbol{\delta}_{\min})) \in \arg\min_{\boldsymbol{\delta}\in\Gamma\times\mathbf{T},\lambda>0,\mu>0} L^*(\boldsymbol{\delta},\lambda^*,\mu^*)$ s.t.

[15]The reason why we only search the exetreme values within space $\Theta\times\Gamma$ instead of whole contract space $\mathcal{S}$ is because we parameterize the primitives within the parameter space. Let $\mathcal{FC}_{\Gamma\times\Theta}$ be the subset of feasible class indexed by parameterized utility family and score family, then given this restriction, for any suboptimal $s(x)\in\mathcal{S}$ which can not be rationalized by parameter space $\Gamma\times\Theta$, we have $\sup_{\boldsymbol{\delta}\in\Gamma\times\Theta}\Delta\Pi(\boldsymbol{\delta}) \le \sup_{s\in\mathcal{S},s\notin\mathcal{FC}_{\Gamma\times\Theta}}\mathbb{E}w_i - \mathbb{E}s(x_i)$, and $0 = \inf_{\boldsymbol{\delta}\in\Gamma\times\Theta}\Delta\Pi(\boldsymbol{\delta}) \le \inf_{s\in\mathcal{S},s\notin\mathcal{FC}_{\Gamma\times\Theta}}\mathbb{E}w_i - \mathbb{E}s(x_i)$.

(17) and (18) and $(\boldsymbol{\delta}_{\max}, \lambda^*(\boldsymbol{\delta}_{\max}), \mu^*(\boldsymbol{\delta}_{\max})) \in \arg\max_{\boldsymbol{\delta}\in\Gamma\times\mathbf{T},\lambda>0,\mu>0} L^*(\boldsymbol{\delta}, \lambda^*, \mu^*)$ s.t. (17) and (18). The lower bound of profit loss is

$$\Delta\Pi(\boldsymbol{\delta})^{lb} = \Delta\Pi(\boldsymbol{\delta}_{\min}) = \mathbb{E}w_i - \mathbb{E}s^*(x_i, \boldsymbol{\delta}_{\min}) \tag{1.19}$$

and the upper bound of profit loss is

$$\Delta\Pi(\boldsymbol{\delta})^{ub} = \Delta\Pi(\boldsymbol{\delta}_{\max}) = \mathbb{E}w_i - \mathbb{E}s^*(x_i, \boldsymbol{\delta}_{\max}). \tag{1.20}$$

Particularly, if there exists some δ such that the first order condition (1) holds, $\Delta\Pi(\delta)^{lb} = 0$, achieving the lower bound of profit loss, otherwise, $\Delta\Pi(\delta)^{lb} > 0$. The reason of $\Delta\Pi(\delta)^{lb} > 0$ is that, if $h(w)$ does not attain CRLB, $s^*(x_i, \boldsymbol{\delta}_{\min})$ can not be the pointwise minimizer of principal's problem (P1), although $\delta_{\min}$ is the minimizer of $L^*(\boldsymbol{\delta}, \lambda^*, \mu^*)$.

The first order conditions can derived be as follows:

$$\begin{aligned} \int_{x\geq w^{-1}(\underline{s})} \frac{[\frac{\partial}{\partial\gamma}u(s^*,\gamma) - \frac{\partial}{\partial\gamma}u(w,\gamma)]}{\frac{\partial}{\partial s}u(s^*,\gamma)} f(x,a)dx &= 0 \\ \text{and} \int_{x\geq w^{-1}(\underline{s})} [u(s^*,\gamma) - u(w,\gamma)][\omega_{a\mathbf{T}}(x,\mathbf{T}) - \mathbb{E}\omega_{a\mathbf{T}}(x_i,\mathbf{T})] f(x,a)dx &= 0. \end{aligned}$$

These two first order conditions are not the first order condition of principal's original problem (P1).

The sample analogue of the Lagrangian $L^*(\delta, \lambda^*, \mu^*)$, constraint (17) and (18) and the two first order conditions can be replaced by their sample analogues. Therefore, bounding profit loss is feasible. The following proposition provides the asymptotic distribution of $\widehat{\Delta\Pi}(\boldsymbol{\delta})^{lb}$ and $\widehat{\Delta\Pi}(\boldsymbol{\delta})^{ub}$.

Proposition 1.4: *If the conditions in theorem 3 hold, and assuming condition for uniform convergence applies, then* $\widehat{\Delta\Pi}(\delta_k) = \hat{E}w_i - \hat{E}s^*(x_i, \delta_k)$ *(k=min or max) has the following asymptotic distribution:*

$$\sqrt{n}(\widehat{\Delta\Pi}(\boldsymbol{\delta}_k) - \Delta\Pi(\boldsymbol{\delta}_k)) \to^d \mathcal{N}(0, nAVar(\widehat{\Delta\Pi}(\boldsymbol{\delta}_k));$$

and $M^*(a) > 0$ *and* $K^*(a) > 0$ *justifying the existence of moral hazard and the second order* *condition.* (Proof in Technical Supplement 4).

One fact worthy noting is that if there only exists a unique δ_0 to solve $\Delta\Pi(\delta_0^{lb}) = 0$, then δ_0 should be the true parameters under the null hypothesis. Unfortunately, it is complicated to analytically justify that $\Delta\Pi(\delta_0^{lb}) = 0$ has a unique solution. Another condition is that if T_0 is the true parameter, then γ_0 will be the true utility parameters under the null hypothesis. This will be the case if either $f(x, a, \theta)$ is functionally known (then we can do MLE to estimate T_0 at the first stage) or there exists a sufficient statistic for effort (then we estimate T_0 by its moment). The following proposition summarizes the result of these two cases.

Proposition 1.5: *If the conditions in theorem 3 hold, then under the null hypothesis, and $f(x, a, \theta)$ is functionally known or there exists a sufficient statistic for effort a, then (i) $\gamma_0 = \arg\min_{\gamma \in \Gamma} \Lambda(\gamma, T)$ should be the true preference, and (ii) under assumption A7, $\hat{\gamma} \in \arg\min_{\gamma} \hat{\Lambda}(\gamma, \hat{T})$ has the following asymptotic distribution:*

$$n(\hat{\gamma} - \gamma_0) \rightarrow^d \mathcal{N}\left[0, \left(\mathbb{E}\frac{\partial\varepsilon_i}{\partial\gamma}\frac{\partial\varepsilon_i}{\partial\gamma'}\right)^{-1}\left(\mathbb{E}\frac{\partial\varepsilon_i}{\partial\gamma}\frac{\partial\varepsilon_i}{\partial\mathbf{T}'}\mathbf{Z}^{-1}\mathbb{E}\frac{\partial\varepsilon_i}{\partial\mathbf{T}}\frac{\partial\varepsilon_i}{\partial\gamma'}\right)\left(\mathbb{E}\frac{\partial\varepsilon_i}{\partial\gamma}\frac{\partial\varepsilon_i}{\partial\gamma'}\right)^{-1}\right],$$

where

$$\begin{aligned} \mathbb{E}\frac{\partial\varepsilon_i}{\partial\gamma}\frac{\partial\varepsilon_i}{\partial\gamma'} &= \frac{Var(h_\gamma)\left(\mathbb{E}l_a^2\right) - Var(h_\gamma l_a)}{\mu^2\left(\mathbb{E}l_a^2\right)^3} \\ \text{and } \mathbb{E}\frac{\partial\varepsilon_i}{\partial\gamma}\frac{\partial\varepsilon_i}{\partial\mathbf{T}'} &= \frac{1}{\mu\left(\mathbb{E}l_a^2\right)^3}\left[Cov(h_\gamma, l_a)Cov(l_a, l_{a\mathbf{T}'}) - \left(\mathbb{E}l_a^2\right)Cov(h_\gamma, l_{a\mathbf{T}'})\right]. \end{aligned}$$

and $\hat{\rho}$ has asymptotic distribution,

$$n(1-\hat{\rho}) \rightarrow^d \mathcal{Q}\left[\frac{1}{2}\begin{pmatrix} \sqrt{n}\hat{\mathbb{E}}l_a \\ \sqrt{n}(\hat{\mathbf{T}} - \mathbf{T}) \end{pmatrix}'\begin{pmatrix} Z^{-1} & 0 \\ 0 & -\rho_{\mathbf{TT}'} \end{pmatrix}\begin{pmatrix} \sqrt{n}\hat{\mathbb{E}}l_a \\ \sqrt{n}(\hat{\mathbf{T}} - \mathbf{T}) \end{pmatrix}\right] \quad \text{if } \rho = 1.$$

(Proof in Appendix A6).

Remark 1.2: *The estimate of the agency cost $\widehat{AC}$ and total loss $\widehat{TL}$ can be done similarly..*

Identification for Selection of Agent

We go further to proceed with the issue that the type of agent may be endogenous. The endogeneity can be due to the principal's selection of certain a type of agent to contract with.

If there exists a certain type of agent, say $\gamma_{\min}$, so that the null hypothesis $H_0^{\rho} = 0$ is accepted, is $\gamma_{\min}$ the most profitable agent in the principal's prospect? This is the identification issue between incentive and selection, which has been mentioned by literature dealing with empirical data. For example, in the agrarian contractual context, the principal (landlord) may choose the most profitable tenant based on tenants' degree of risk version. Let $V(\gamma)$ be the solution to problem (P1), then if the principal recognizes the profitability of selecting different agents, he should solve the further optimization problem:

$$(P2): \max_{\gamma \in \mathbf{\Gamma}} V(\gamma) \text{ s.t. IC and IR.}$$

The solution to (P2) means that the principal "chooses the right agent to offer the right contract".

Our identification of selection depends on additional restriction over γ related to the solution to problem (P2). Given $\mathbf{\Gamma}$ is compact, to find $\gamma^* \in \arg\max_{\gamma \in \mathbf{\Gamma}} V(\gamma)$ s.t. IC and IR, we only need to search the saddle point(s) and the boundary of $\mathbf{\Gamma}$. Thus we can find γ^* by the first order condition w.r.t. γ. According to the envelope theorem, an estimable condition can be found as the following lemma states.

Lemma 1.1: *If $s^*(x)$ is the optimal contract regarding agent's utility $u(s^*(x), \gamma)$, therefore γ^* is the most profitable agent selected by the principal (matching with contract $s^*(x)$) if*

$$V_{\gamma}^*(\gamma, \mathbf{T}) = \frac{\partial u(\underline{s}, \gamma)}{\partial \gamma} \int_{w<\underline{s}} (\lambda + \mu\omega_a(x, \mathbf{T})) f(x, a) dx + \int_{w \geq \underline{s}} \frac{\frac{\partial}{\partial \gamma} u(s^*(x), \gamma)}{\frac{\partial}{\partial s} u(s^*(x), \gamma)} f(x, a^*) dx = 0. \quad (1.21)$$

(Proof see Appendix A7)

Formula (21) in lemma 1 is strikingly simple, whose sample analogue is $\widehat{V_{\gamma}^*}(\hat{\gamma}, \hat{\mathbf{T}}) = \Pr(w \geq \underline{s}) \hat{\mathbb{E}}_{w \geq \underline{s}} \frac{\frac{\partial}{\partial \gamma} u(w_i, \gamma)}{\frac{\partial}{\partial w} u(w_i, \gamma)} + \Pr(w < \underline{s}) \frac{\partial u(\underline{s}, \gamma)}{\partial \gamma} \hat{\mathbb{E}}_{w<\underline{s}} (\hat{\lambda} + \hat{\mu} l_a(x, \hat{a}))$. One simple way to test selection is to test $H_0^{v^*}: V_{\gamma}^*(\gamma_0, \mathbf{T}_0) = 0$ where $(\gamma_0, \mathbf{T}_0)$ is found by $\rho(\gamma_0, \mathbf{T}_0) = 1$ and $(\gamma_0, \mathbf{T}_0)$ is not on the boundary of parameter space $\mathbb{A} \times \Theta \times \Gamma$. But in some cases, if asymptotic distribution of $\widehat{V_{\gamma}^*}(\hat{\gamma}, \hat{\mathbf{T}})$ does not exist (for example, CRRA utility under null hypothesis $\gamma_0 = 0$), then we test selection in another way. This is to use the estimators of $\mathbf{T}$ from theorem 3 to replace the object in (21), and solve γ^* by comparing the value function at saddle points with its value at boundary. Particularly, if $\underline{s}$ is not effective, then we can solve γ^* without involving $\mathbf{T}$. Based on γ^* found in (P2), we can test the null hypothesis $H_{0s} : \rho(\gamma^*, T_0) = 1$. We summarize the above results as follows.

Proposition 1.6: *Under the conditions that theorem 3 holds, (i) when there exist* γ_0 *such that* $H_0 : \rho(\gamma_0) = 0$ *accepts, then the principal does not solve the problem (P2), if either hypothesis* $H_0^{v^*} : V_\gamma^*(\gamma_0, T_0) = 0$ *or* $H_{0s} : \rho(\gamma^*) = 0$ *is rejected. (ii)* V_γ^* *can be estimated* $\widehat{V_\gamma^*}(\hat{\gamma}, \hat{T})$ *and by* $\rho(\gamma^*)$ *can be approximated by* $\hat{\rho}(\hat{\gamma}^*)$ *and their asymptotic distributions can be established.* (Proof in Technical Supplement 5).

Meanwhile, we also can estimate the profit loss due to either suboptimality of incentive or selection, or both of them. When the null hypothesis H_0^ρ is rejected, the bound of the profit loss is

$$\widehat{\Delta\Pi} = \underbrace{\frac{1}{n}\sum[w_i - \widehat{s^*}(x_i)]}_{\text{incentive error}} + \underbrace{\frac{1}{n}\sum[\widehat{s^*}(x_i) - \widehat{s^{**}}(x_i, \widehat{\gamma^{**}})]}_{\text{selection error}}$$

where $\widehat{s^*}(x)$ is defined as proposition 4, and $\widehat{s^{**}}(x_i)$ can be found as follows:

$$\begin{aligned} h(\widehat{s^{**}}(x_i), \widehat{\gamma^{**}}) &= \widehat{\lambda^{**}} + \widehat{\mu^{**}} l_a(x, \hat{\mathbf{T}}) \\ \hat{\mathbb{E}} u(\widehat{s^{**}}(x), \widehat{\gamma^{**}}) l_a(x, \hat{\mathbf{T}}) &= \hat{\mathbb{E}} u(w_i, \widehat{\gamma^{**}}) l_a(x, \hat{\mathbf{T}}) \\ \hat{\mathbb{E}} u(\widehat{s^{**}}(x), \widehat{\gamma^{**}}) &= \hat{\mathbb{E}} u(w_i, \widehat{\gamma^{**}}) \end{aligned}$$

with $\widehat{\gamma^{**}} \in \arg\max_{\gamma \in \mathbf{\Gamma}} V(\gamma)$ and $(\widehat{\lambda^{**}}, \widehat{\mu^{**}})$ and $\hat{\mathbf{T}}$ defined as in proposition 4.

The idea of the above decomposition is to construct the optimal contract based on the most profitable type of agent (the solution is unique for $s^{**}(x)$), under the same IC constraint and IR constraint. By this decomposition, we can quantify the mechanism leading to efficiency loss (gains).

Identification for selection of technology

Another important issue is that the parameter(s) in the production function, namely $\boldsymbol{\theta}$, can be a decision variable too, though exogenous to the agent. The choice of production function or technology is to choose the "information system" (Holmstrom, 1979; Kim, 1995), which indicates the degree of informativeness of the output[16]. Paralleling the previous section, the matching of technology to the agent and/or contract should also be detected. Let $V(\boldsymbol{\theta})$ be the solution to problem (P1) when $\boldsymbol{\theta}$ is parameter(s) in distribution function or score function. If the principal indeed is doing optimalization over $\boldsymbol{\theta}$, he should solve the following optimization problem:

$$(P3): \max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} V(\boldsymbol{\theta}) \text{ s.t. IC and IR.}$$

The solution to (P3) means the principal "choosing the right technology to match the right contract".

Similarly, $\boldsymbol{\theta}^* \in \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} V(\boldsymbol{\theta})$ s.t. IC and IR can be found by the first order condition, compared with the boundary of parameter(s). By the envelop theorem, the first order condition becomes:

$$V_{\boldsymbol{\theta}}^* = \int (x - s^*(x)) f_{\boldsymbol{\theta}}(x, \mathbf{T}) dx + \lambda \int u(s^*(x), \gamma) f_{\boldsymbol{\theta}}(x, \mathbf{T}) dx + \mu \int u(s^*(x), \gamma) f_{a\boldsymbol{\theta}}(x, \mathbf{T}) dx = 0, \tag{1.22}$$

where,

$$\begin{aligned} \frac{f_{\boldsymbol{\theta}}(x,\mathbf{T})}{f(x,\mathbf{T})} &= l_{\boldsymbol{\theta}} = \frac{\partial \log f(x,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \omega_{\boldsymbol{\theta}}(x,a) - \mathbb{E}\omega_{\boldsymbol{\theta}}(x,a), and \\ \frac{f_{a\boldsymbol{\theta}}(x,\mathbf{T})}{f(x,\mathbf{T})} &= l_{a\boldsymbol{\theta}} + l_a l_{\boldsymbol{\theta}} = \omega_{\boldsymbol{\theta}a}(x,a) - \mathbb{E}\omega_{\boldsymbol{\theta}a}(x,a)\omega_a(x,a) + (\omega_a(x,a) - \mathbb{E}\omega_a(x,a))(\omega_{\boldsymbol{\theta}}(x,a) - \mathbb{E}\omega_{\boldsymbol{\theta}}(x,a)), \end{aligned}$$

so that $V_{\boldsymbol{\theta}}^*$ can be replaced by its empirical moment.

We can test whether $\boldsymbol{\theta}$ is the most profitable technology matching with current contract or the agent. The following proposition states the results.

Proposition 1.7: *The technology is perfectly matched with the contract if the following hypothesis:*

$$H_{0m} : V_{\boldsymbol{\theta}}^* = 0,$$

is accepted, where $\hat{V}_{\boldsymbol{\theta}}^*$ has asymptotic distribution

$$\sqrt{n}(\hat{V}_{\boldsymbol{\theta}}^* - V_{\boldsymbol{\theta}}^*) \rightarrow^d \mathcal{N}(0, nAVar(\hat{V}_{\boldsymbol{\theta}}^*)).$$

[16]For example, the landlord may choose different crops, because different crops may vary in their dependence on weather condition; therefore, the precision of measurement of the effort through the output will be different.

(Proof in Technical Supplement 6)

1.3.3 Testing $M > 0$, $K > 0$ and Adjoint Equation (3)

The remaining subsections are to test $M(a) > 0$ and $K(a) > 0$. It is easy to test $\hat{M} > 0$ by using a sample analogue of M and asymptotic variance of $\hat{M}$ can be computed by bootstrapping.

Proposition 1.8: *When $l_a(x,a)$ is parameterized by formula $l_a(x,a) = \frac{1}{\eta(\mathbf{T})}[(\varpi(x,T) - E\varpi(x,T))]$, $M(a) > 0$ if and only if the sign of $Cov(h(w,\gamma), \varpi(x,T))$ is the same as that of $Cov(x, \varpi(x,T))$.*

Proof. Because $l_a(x,a)$ is not completely identified (up to a multiplicative constant), we can determine the sign of $\frac{\mu}{\eta(\mathbf{T})}$, which is determined by $\frac{\mu}{\eta(\mathbf{T})} = \frac{Cov(h(w,\gamma),\varpi(x,\mathbf{T})}{\mathbb{E}(\varpi(x,\mathbf{T})-\mathbb{E}\varpi(x,\mathbf{T}))^2}$. The sign of $\frac{Cov(h(w,\gamma),\varpi(x,\mathbf{T})}{\mathbb{E}(\varpi(x,\mathbf{T})-\mathbb{E}\varpi(x,\mathbf{T}))^2}$ can be tested based on its sample analogue. Without loss of generality, suppose $\frac{\mu}{\eta(\mathbf{T})} > 0$ is the case, then we need to check whether $\eta(\mathbf{T}) > 0$. By assumption A3,

$$\frac{\partial \mathbb{E} x_i}{\partial a} = \frac{1}{\eta(\mathbf{T})} \int x(\varpi(x,\mathbf{T}) - \mathbb{E}\varpi(x,\mathbf{T}))f(x,a)dx > 0,$$

thus $Cov(x_i, \varpi(x_i,\mathbf{T})) > 0$ if and only if $\eta(\mathbf{T}) > 0$. Similarly, we can show the case $\frac{\mu}{\eta(\mathbf{T})} < 0$. Q.E.D. ∎

The difficulty to test $\hat{K} > 0$ is due to absence of an estimator of $l_{aa}(x,a)$. We discuss two cases, depending on whether $\varpi(x,\mathbf{T})$ contains a or not. We have the following proposition.

Proposition 1.9: *The second order condition can be tested based on the sample analogues:*

(i) When $\varpi(x,T)$ does not contain a, to test $K > 0$, it suffices to test

$$\frac{Cov(x_i, (\varpi(x_i,\mathbf{T}) - \mathbb{E}\varpi(x_i,\mathbf{T}))^2)}{Cov(x, \varpi(x_i,\mathbf{T}))} Cov(u(w_i), \varpi(x_i,\mathbf{T})) > Cov(u(w_i), (\varpi(x_i,\mathbf{T}) - \mathbb{E}\varpi(x_i,\mathbf{T}))^2). \tag{1.23}$$

(ii) When $\varpi(x,T)$ contains a, and the sign of $Cov(x_i, (\varpi(x_i,T) - E\varpi(x_i,T))^2)$ is the same with that of $Cov(u(w_i), (\varpi(x_i,T) - E\varpi(x_i,T))^2)$, the sufficient condition for $K > 0$ is that the sign of $Cov(x_i, \frac{\partial}{\partial a}\varpi(x_i,T)) \frac{Cov(u(w_i),(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))^2)}{Cov(x_i,(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))^2)} - Cov(u(w_i), \frac{\partial}{\partial a}\varpi(x_i,T))$ is the same as that of $Cov(x_i, \varpi(x_i,T))$. **(Proof in Appendix A7)**

We can use the sample analogues of the above objects, and the asymptotic variance of the above objects can be obtained by bootstrapping[17].

Once the null hypothesis H_0^ρ is accepted, i.e. the contract is the conditional constrained optimum, we can partially identify whether the effort level is optimal or not by testing the following two inequalities implied by adjoint equation (3):

$$\int (x - s(x)) f_a(x,a)dx + \mu \int u(s(x)) f_{aa}(x,a)dx > 0$$

and

$$\int (x - s(x)) f_a(x,a)dx > 0.$$

If $\int (x - s(x)) f_a(x,a)dx \leq 0$, we can conclude that the effort level a is higher than the optimal level. In this respect, one can increase the effort level until $\int (x - s(x)) f_a(x,a)dx + \mu \int u(s(x)) f_{aa}(x,a)dx$ is no longer significantly positive (See Technical Supplement 7 for the details).

1.4 Non-parameteric Identification With Unknown Utility $u(w)$ and Score $l_a(x,a)$

This section considers a more general case, in which both utility $u(w)$ and score $l_a(x,a)$ are unknown. In general, without putting any further assumptions on agent's utility, cost of effort and score function, it is impossible for the econometrician to fully identify contract optimality and the moral hazard problem. Although with infinite data, we can recover the distribution of output, say, $f(x,a)$, but without variations of a, the score function $l_a(x,a)$ is not identified based on the single sequence of data generated by the same a. This situation differs from identification of the adverse selection problem (d'Haultfoeuille and Fevrie, 2007), where the quantile of agents' type can be fully identify based on distribution of trade data, provided single-crossing property. Given this difficulty, we can not jointly identified two functions $h(w)$ and $l_a(x,a)$ only based on one first order condition (1). We formally state it as follows.

[17] When the sign of $Cov(x, (\varpi(x,\hat{\mathbf{T}}) - \hat{\mathbb{E}}\varpi(x,\mathbf{T}))^2)$ is not the same with that of $Cov(u(w), (\varpi(x,\hat{\mathbf{T}}) - \hat{\mathbb{E}}\varpi(x,\mathbf{T}))^2)$, or when either we are not sure whether $\varpi(x,\mathbf{T})$ contains a, or the functional form $\frac{\partial}{\partial a}\varpi(x,\mathbf{T})$ is unavailable, we need some additional conditions. For example $s(x)$ non-decreasing and $l_{aax}(x,a) < 0$. Under these two conditions, $Cov(u(s(x)), l_{aa}(x,a)) < 0$, implying $\mathbb{E}u(w)l_{aa}(x,a) \leq \mathbb{E}u(w)\mathbb{E}l_{aa}(x,a) = \mathbb{E}u(w)\mathbb{E}l_a^2(x,a)$, thus the sufficient condition for $K > 0$ is $Cov(u(w), (\varpi(x,\mathbf{T}) - \mathbb{E}\varpi(x,\mathbf{T}))^2) < 0$.

Proposition 1.10: *Given a single sequence of data being observed, $h(w)$ and $l_a(x, a)$ can not be jointly identified without knowning one of these two functions.*

The questions turn out to be: what is a weaker condition to identify $h(w)$ and $l_a(x, a)$? The condition provided in this paper is that output is generated by effort and noise additively. Formally, we write as follows,

$$x = m(a) + \varepsilon \tag{1.24}$$

where $m(a) = \mathbb{E}x_i$ is an unknown monotonically increasing function of effort level a (according to assumption A3); and ε is random noise coming from an unknown density. This specification is general enough to cover many distributions used in empirical studies. The left hand side of the above equation can be generalized to any parameterized monotone transformation $y(x, \theta)$ or even unknown monotone transformation $y(x)$. We will discuss these generalization in another paper (Ke, 2008). For simplicity, we derive all the theoretical results based on formula (24).

If (24) holds, the unknown score function can be estimated non-parametrically up to a positive scale constant[18] as follows,

$$l_a(x, a) = -\frac{f'(\varepsilon)}{f(\varepsilon)} m'(a).$$

Without loss of generality, we normalize $m'(a) = 1$ since we need not estimate $m'(a)$. For convenience, since a in this section is a nuisance variable, we suppress $l_a(x, a)$ as $l_a(x)$.

For the d-th order derivative of density $f(\varepsilon)$, if we choose band width $b \to 0$ such that $nb^{2d+1} \to \infty, nb^{2d+1}b^2 \to 0$ as $n \to \infty$, then we have the asymptotic distribution (See Pagan and Ullah, 1989, pp56, Li and Racine, 2007)

$$\sqrt{nb^{2d+1}}(\hat{f}^{(d)} - f^{(d)}) \to^d \mathcal{N}\left(0, f \int (K^{(d)}(\varphi))^2 d\varphi\right), \tag{1.25}$$

where $\hat{f}^{(d)}$ is the kernel estimator

$$\hat{f}^{(d)}(\varepsilon_i) = \frac{(-1)^d}{nb^{d+1}} \sum_{j=1, j\neq i}^{n} K'(\frac{\varepsilon_j - \varepsilon_i}{b})$$

[18]Denote the noise's c.d.f (p.d.f.) as $G(\varepsilon)$ ($g(\varepsilon)$) temporoally, then, $F(x, a) = \Pr(X \leq x) = \Pr(\varepsilon \leq x - m(a)) = G(x - m(a))$, thus $f(x, a) = g(x - m(a))$, and $\frac{df}{da} = -m'(a)g'(\varepsilon)$.

and $K(\varphi)$ is a kernel function with some appropriate properties such as:

$$\int K^{(d)}(\varphi)\varphi^j d\varphi = \begin{cases} 0 & \text{for j=0, 1, 2,...,d-1} \\ (-1)^d d! & \text{for j=d} \\ 0 & \text{for j=d+1} \\ \text{constant} & \text{for j=d+2} \end{cases}$$

Therefore, choosing $b \propto n^{-\frac{1}{7}-\alpha}$ ($\alpha > 0$), and using leave-one-out estimator $\hat{l}_a = -\frac{\hat{f}'(\varepsilon_i)}{\hat{f}(\varepsilon_i)}$, thus yields $\lim_{n\to\infty}(\mathbb{E}\hat{l}_a - l_a)^2 \to 0$, and $\lim_{n\to\infty} Var(\hat{l}_a) = 0$. The pointwise asympotic distribution is as follows,

$$\sqrt{nb^3}(\hat{l}_a - l_a) \to \mathcal{N}(0, \frac{1}{f(\varepsilon)}\int [K'(\varphi)]^2 d\varphi).$$

Based on well established nonparametric estimation techniques, we state the following theorem.

Theorem 1.4: *If the moral hazard problem is regular, FOA is valid, assumption A5 holds, and the output is generated by equation (24), then*

(i) contract optimality can be tested by testing the null hypothesis $H_0^\rho : \rho(h, l_a) = 1$ *and* $\rho(h, l_a)$ *can be estimated by*

$$\hat{\rho}(\hat{h}, \hat{l}_a) = \frac{\widehat{Cov}(\hat{h}(w_i), \hat{l}_a(x_i))}{\sqrt{\widehat{Var}(\hat{h}(w_i))\widehat{Var}(\hat{l}_a(x_i))}}$$

where

$$\hat{h}(w_i) \in \arg \min_{h\in\mathcal{H}, h_w>0} \hat{\mathbb{E}}(\frac{h(w_i) - \hat{\mathbb{E}}h(w_i)}{\widehat{Cov}(\hat{h}(w_i), \hat{l}_a(x_i))} - \frac{\hat{l}_a(x_i)}{\widehat{Var}(\hat{l}_a(x_i))})^2$$

is a monotone smooth non-parametric estimator of marginal incentive cost $h(w_i)$ *from a compact space* H.

(ii) Under the null hypothesis, $\hat{\rho}(\hat{h}, \hat{l}_a)$ *has the following asymptotic distribution,*

$$\hat{\rho} - 1 + \frac{1}{2Z}\left(\frac{1}{nb^3}\mathbb{E}\frac{1}{f(\varepsilon_i)}\int [K'(\varphi)]^2 d\varphi + \frac{1}{4}b^4\left(\int \varphi^2 K(\varphi)d\varphi\right)^2 Var(\frac{[f'(\varepsilon_i)f''(\varepsilon_i) - f(\varepsilon_i)f'''(\varepsilon_i)]}{f(\varepsilon_i)^2})\right)$$
$$\to {}^d\mathcal{N}\left(0, \frac{1}{4nZ^2}\left(\frac{24}{nb^3}\mathbb{E}\frac{l_a^2(x_i)}{f(\varepsilon_i)}\int [K'(\varphi)]^2 d\varphi] + O(\frac{1}{nb}) + o(b^4)\right)\right).$$

(Proof see Appendix A9)

The above theorem provides an optimality test without functional form specifications of marginal utility or score function. The power to reject the null hypothesis comes from the fact that wage should be a smooth monotonically increasing function of score. This is the well-known monotone likelihood ratio property advocated by Milgrom (1982). The violation of optimality happens when the payment usually is not an increasing function of the score (maybe an increasing function of performance).

Because implementing theorem 4 involves a two-step non-parametric procedure, it is desirable to find some ways to reduce the computational cost by testing the relationship between payment and score directly. One very loose testing is to test $Cov(w_i, l_a(x_i)) > 0$, but the stronger testing of correlation should be a rank order correlation test (Kendall, 1955). The discussion of those tests is not the main purpose of this paper. The above test is conservative, as we have noticed before. However, under the null hypothesis that the contract is optimal if we rationalize data by some utility function (u, c), theorem 1 leads us to conclude that (u, c) is unique though it can not be completely estimated. And the estimated $\hat{h}(w_i)$ in theorem 4 identifies the true $h(w_i)$ up to an affine transformation. So we can construct confidence interval to cover the true $h(w_i)$ based on the data. We state it as follows.

Theorem 1.5: *If the conditions in theorem 4 hold, then under the null hypothesis $H_0^\rho : \rho = 1$, based on*

$$\hat{h}(w_i) \in \arg \min_{h \in \mathcal{H}, h_w > 0} \hat{\mathbb{E}}(\frac{h(w_i) - \hat{\mathbb{E}}h(w_i)}{\widehat{Cov}(\hat{h}, \hat{l}_a(x_i))} - \frac{\hat{l}_a(x_i)}{\widehat{Var}(\hat{l}_a(x_i))})^2,$$

we can identify $h(w)$ (up to positive affine transformation) by the asymptotic distribution:

$$\begin{aligned} & n\sqrt{b^3 b_h}\left(\hat{h}(w) - h(w) - \frac{1}{2}b_h^2 \frac{h''(w)\phi(w) + 2h'(w)\phi'(w)}{\phi(w)} \int \varphi^2 K_r(\varphi) d\varphi\right) \\ \to & \ {}^d\mathcal{N}\left(0, \frac{1}{\phi(w) f(\varepsilon)} \int [K'(\varphi)]^2 d\varphi \int K_r(\varphi)^2 d\varphi\right), \end{aligned}$$

where b is defined as in theorem 4, $\lim_{n \to \infty} b_h \to 0$ and $n\sqrt{b^3 b_h} \to \infty$, and $\phi(w)$ is p.d.f. of payment schedule w. **(Proof in Appendix A10)**.

Another step is to bound the profit loss based on the consistent estimator $\hat{l}_a(x_i)$, the idea is the same as in section 4, except replacing $h(w_i)$ and $l_a(x_i)$ by their non-parametric estimators. We summarize as follows.

Theorem 1.6: *If the conditions in theorem 4 hold, then the profit loss of observed data* $\{x_i, w_i\}_{i=1}^n$ *can be bounded by the extrema of* $L^*(\hat{h}, \lambda^*, \mu^*)$*, and the asymptotic variance of* $\widehat{\Delta\Pi}$ *can be bootstrapped accordingly.* **(Proof in Appendix A11)**

One caveat of the above theorem is that we do not have theoretical judgement about how sharp the bounds are, because they depend on the data and the particular model. Finding some good theoretical properties of bound of profit loss is a future research question. One fact worth noting is that we can find the most profitable utility by solving problem (P2) over a compact function space, similar to what we did in the previous section. The profit loss can decomposed into selection error and incentive error as described in proposition 6.

Although above results are very general, there are two special cases of interest. The first case is that utility is unknown but score can be parameterized by some unknown parameters. In this case, we can find $(\hat{h}(w_i), \hat{\mathbf{T}}) \in \arg\min_{h \in \mathcal{H}, h_w > 0, \mathbf{T} \in A \times \Theta} \hat{\Lambda}(h, \mathbf{T})$ without non-parametric estimation of $\hat{h}(w_i)$, and test the null hypotheses by replacing $l_a(x, \hat{\mathbf{T}})$ with non-parametric estimator $\hat{l}_a$.

Corollary 1.2: *If the conditions in theorem 4 hold, when the score function can be parameterized by* $l_a(x, T) = \varpi(x, T) - E\varpi(x, T)$*, then contract optimality can be tested by testing the null hypothesis* $H_0^\rho : \rho(h, l_a) = 1$*. And* $\rho(h, l_a)$ *can be estimated by*

$$\hat{\rho}(\hat{h}(w_i, \hat{\gamma}), \hat{l}_a) = \frac{\widehat{Cov}(\hat{h}(w_i, \hat{\gamma}), \hat{l}_a(x_i))}{\sqrt{\widehat{Var}(\hat{h}(w_i, \hat{\gamma}))\widehat{Var}(\hat{l}_a(x_i))}},$$

where $(\hat{\gamma}, \hat{T}) \in \arg\min_{(\gamma, \mathbf{T})} \hat{\Lambda}(\gamma, T)$*. And the asymptotic variance of* $\hat{\rho}(\hat{h}, \hat{l}_a)$ *can be obtained.*

The second case is that the monetary utility function is parameterized by some unknown parameters. In this case, we can infer the most favorable score function $l_a(x, \mathbf{T})$ theoretically up to some parameters $\mathbf{T}$. The results are summarized as follows.

Corollary 1.3: *If the conditions in theorem 4 hold, when and the monetary utility function can be parameterize by* $u(w, \gamma)$*, then contract optimality can be tested by testing the null hypothesis* $H_0^\rho : \rho(h, l_a) = 1$*. And* $\rho(h, l_a)$ *can be estimated by*

$$\hat{\rho}(\hat{h}(w_i, \hat{\gamma}), \hat{l}_a(x_i)) = \frac{\widehat{Cov}(\hat{h}(w_i, \hat{\gamma}), \hat{l}_a(x_i))}{\sqrt{\widehat{Var}(\hat{h}(w_i, \hat{\gamma}))\widehat{Var}(\hat{l}_a(x_i))}},$$

where $(\hat{\gamma}, \hat{T}) \in \arg\min_{(\gamma,\mathbf{T})} \hat{\Lambda}(\gamma, T)$. *And the asymptotic variance of* $\hat{\rho}(\hat{h}, \hat{l}_a)$ *can be obtained.*

Remark 1.3: *When the agent's utility is not separable, like* $u(s(x), a)$, *the first order condition (1) becomes*

$$\frac{v'(x - s(x))}{u_w(s(x), a)} - \mu \frac{u_{wa}(s(x), a)}{u_w(s(x), a)} = \lambda + \mu l_a(x).$$

Let $h(s(x), a, \mu) = \frac{v'(x-s(x))}{u_w(s(x),a)} - \mu \frac{u_{wa}(s(x),a)}{u_w(s(x),a)}$, *the basic conclusions still go through, as long as* $h(s(x), a, \mu)$ *can be parameterized. Particularly, the derivation of test will be identical when the utility function is log-separable, say,* $u(w, a) = u(w)c(a)$.

Remark 1.4: *The conclusion can also easily extend to the Glossman-Hart approach of characterization of moral hazard, if FOA is valid.*

Remark 1.5: *The model can be used to deal with optimal insurance deduction considered by Holmstrom (1979).*

Remark 1.6: *It may be easier to implement all the procedures when the action space is binary, say, agent can choose* a_H *or* a_L. *In this situation, the likelihood ratio will be* $l_a(x) = \frac{f(x,a_H) - f(x,a_L)}{f(x,a_H)}$, *and theorems still holds and the algebra is similar.*

1.5 Heterogeneous Data-generating Process

This section extends the previous result to the situation that the data-generating process is heterogeneous. The heterogeneity of data-generating process can due to non-i.i.d. shock, preference and/or productivity heterogeneity, or contract variety. In addition, it is very important for the theory to distinguish the principal's observed heterogeneity from the agent's privately known heterogeneity. In this section, we deal with commonly observed heterogeneity, which sheds important light on the principle of informativeness (whether additional observables should enter the contract), heterogeneous shock, and non-deterministic contract process. We leave the agent's privately known preference heterogeneity issue to another paper, where moral hazard and adverse selection coexist (Ke, 2008).

1.5.1 Commonly Observed Heterogeneity

Suppose both the principal and agent observe some additional information $\mathcal{Z} \in \mathbb{Z} \subset \mathbb{R}^{|\mathbb{Z}|}$ besides output x. $\mathcal{Z}$ could be multidimensional signal, $|\mathbb{Z}|$ may be finite or infinite, and assume $\lim_{n\to\infty} \frac{1}{n} |\mathbb{Z}| = 0$. There are several ways to introduce $\mathcal{Z}$, depending on how $\mathcal{Z}$ affects contracting and the agent's behavior. We adopt a more general form that both principal and agent observe $\mathcal{Z}$ before the agent puts effort, and the principal should sign the contract based on $(x, \mathcal{Z})$ and implement effort level according to $\mathcal{Z}$. Suppose the joint distribution of output is $f(x,\mathcal{Z}, a)$, therefore the first order condition will be[19]:

$$\frac{v'(x - s(x, \mathcal{Z}))}{u'(s(x, \mathcal{Z}))} = \lambda(\mathcal{Z}) + \mu(\mathcal{Z}) \frac{f_a(x, \mathcal{Z}, a(\mathcal{Z})))}{f(x, \mathcal{Z}, a(\mathcal{Z}))} \tag{1.26}$$

We can also introduce preference heterogeneity into the model. Thus λ and μ become a function of $\mathcal{Z}$ (or subset of $\mathcal{Z}$). The utility can be parameterized as $u(w, \gamma\,|\mathcal{Z})$.

Since $\mu(\mathcal{Z})$ and $a(\mathcal{Z})$ only vary in $\mathcal{Z}$, when the number of the observations with the same $\mathcal{Z}$ is large enough, we can do the same analysis within each group $\mathcal{Z}$ as we did in section 4. This approach has some disadvantages. For example, when number of observations from the same $\mathcal{Z}$ varies, the quality of estimation also varies. And running testing procedures case by case will lose some information that can be utilized more efficiently. Instead, we can use all the information to estimate unknowns. We only discuss the parametric identification here, assuming utility is known up to some unknown parameter(s) γ, and score function $l_a(x,\mathcal{Z}, a(\mathcal{Z})) = \frac{f_a(x,\mathcal{Z},a(\mathcal{Z}))}{f(x,\mathcal{Z},a(\mathcal{Z}))}$ can be parameterized by $\omega_a(x,\mathcal{Z}, \mathbf{T}) - \mathbb{E}_x[\omega_a(x_i,\mathcal{Z}, \mathbf{T})\,|\mathcal{Z}]$. Suppose $\mathcal{Z}$ is a discrete (or discretized) variable and let $\rho_{\mathcal{Z}} = \frac{J(\mathcal{Z})}{\sqrt{Q(\mathcal{Z})Z(\mathcal{Z})}}$. Note that $\bar{\rho} \equiv \sum_{\mathcal{Z}} \frac{n_{\mathcal{Z}}}{n} \rho_{\mathcal{Z}} = 1$ if and only if for all $\rho_{\mathcal{Z}} = 1$ since $\rho_{\mathcal{Z}} \leq 1$ for all $\mathcal{Z}$. To test the optimality is to test whether $\bar{\rho} = 1$ by using its sample analogue. Theorem 3 can be extended as follows.

Theorem 1.7: *If the conditions that theorem 3 hold for every Z, given the information $(w_i, x_i, Z)_{i=1}^n$,*

(i) the contract $s(x, Z)$ is optimal only if hypothesis $H_0 : \bar{\rho}(\delta) = 1$ is accepted, where $\bar{\rho}(\delta)$

[19] The classic model (Holmstrom, 1979) is that $\mathcal{Z}$ happens after effort has been made. For example, the weather changes after tenant finishes his field work. In this case, cost function $c(a)$, and monetary utility function $u(w)$ do not depend on $\mathcal{Z}$, nor the effort level. Another type of modeling is that $\mathcal{Z}$ affects effort level but will be verified by both parties ex post. In this case, the contract will be more complicated, we leave this to another paper. In practice, which type of $\mathcal{Z}$ makes more sense depends on the context.

can be approximated by

$$\widehat{\rho}(\hat{\gamma},\hat{\mathbf{T}})=\sum_{\mathcal{Z}}\frac{n_{\mathcal{Z}}}{n}\left(\frac{\widehat{Cov}(\omega_a(x_i,\mathcal{Z},\hat{\mathbf{T}}),h(w_i,\hat{\gamma}\,|\mathcal{Z})\,|\mathcal{Z})}{\sqrt{\widehat{Var}(\omega_a(x_i,\mathcal{Z},\hat{\mathbf{T}})\,|\mathcal{Z})\widehat{Var}(h(w_i,\hat{\gamma}))\,|\mathcal{Z})}}\right)$$

with $(\hat{\gamma},\hat{T})\in\arg\min_{(\gamma,\mathbf{T})}\hat{\Lambda}(\gamma,T)$, *where,*

$$\hat{\Lambda}(\gamma,\mathbf{T})=\sum_{\mathcal{Z}}\frac{n_{\mathcal{Z}}}{n}\sum_{i=1}^{n_{\mathcal{Z}}}\left(\frac{h(w_i,\gamma\,|\mathcal{Z})-\bar{h}(\mathcal{Z})}{\sum_{i=1}^{n_{\mathcal{Z}}}h(w_i,\gamma\,|\mathcal{Z})(\omega_a(x_i,\mathcal{Z},\mathbf{T})-\bar{\omega}_a(\mathcal{Z}))}-\frac{(\omega_a(x_i,\mathcal{Z},\mathbf{T})-\bar{\omega}_a(\mathcal{Z}))}{\frac{n_{\mathcal{Z}}}{n_{\mathcal{Z}}-1}\sum_{i=1}^{n_{\mathcal{Z}}}\omega_a(x_i,\mathcal{Z},\mathbf{T})-\bar{\omega}_a(\mathcal{Z}))^2}\right)^2$$

and $n_{\mathcal{Z}}$ *is the number of observations with the same* Z.

(ii) With the additional assumptions,

$$\sup_{\boldsymbol{\delta}}\left\|\frac{\partial^2\hat{\Lambda}(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}-\frac{\partial^2\Lambda(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}\right\|\to^p 0 \text{ and } \sqrt{n}\hat{\Lambda}_{\boldsymbol{\delta}}(\boldsymbol{\delta})\to^d\mathcal{N}(0,\Omega)$$

under the null hypothesis, $\hat{\rho}(\hat{\gamma},\hat{T})$ *has asymptotic distribution*

$$\sqrt{n}(\widehat{\rho}-\bar{\rho})\to^d\mathcal{N}(0,nAVar(\hat{\rho}))$$

and $\hat{\delta}$ *has asymptotic distribution*

$$\sqrt{n}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta})\to^d\mathcal{N}\left(0,[\Lambda_{\boldsymbol{\delta}\boldsymbol{\delta}'}(\boldsymbol{\delta})]^{-1}\Omega[\Lambda_{\boldsymbol{\delta}\boldsymbol{\delta}'}(\boldsymbol{\delta})']^{-1}\right).$$

(Proof in Appendix A12).

Meanwhile, we can estimate the bounds of the profit loss similar to what we have done in the previous sections. The Lagrangian is,

$$\begin{aligned}
&L^*(\boldsymbol{\delta},\lambda^*(\mathcal{Z}),\mu^*(\mathcal{Z}))\\
=\;&\sum\frac{n_{\mathcal{Z}}}{n}\{\int[w-h^{-1}(\lambda^*+\mu^*[\omega_a(x,\mathcal{Z},\mathbf{T})-\mathbb{E}\omega_a(x_i,\mathbf{T})])f(x,\mathcal{Z},a)dx\\
&+\lambda^*(\mathcal{Z})[\int u(h^{-1}(\lambda^*+\mu^*[\omega_a(x,\mathcal{Z},\mathbf{T})-\mathbb{E}\omega_a(x_i,\mathcal{Z},\mathbf{T})]),\gamma)f(x,\mathcal{Z},a)dx-\int u(w,\gamma)f(x,\mathcal{Z},a)dx]\\
&+\mu^*(\mathcal{Z})\left[\begin{array}{c}\int u(h^{-1}(\lambda^*+\mu^*[\omega_a(x,\mathcal{Z},\mathbf{T})-\mathbb{E}\omega_a(x_i,\mathcal{Z},\mathbf{T})]),\gamma)[\omega_a(x,\mathcal{Z},\mathbf{T})-\mathbb{E}\omega_a(x_i,\mathcal{Z},\mathbf{T})]f(x,\mathcal{Z},a)dx\\ -\int u(w,\gamma)[\omega_a(x,\mathcal{Z},\mathbf{T})-\mathbb{E}\omega_a(x_i,\mathcal{Z},\mathbf{T})]f(x,\mathcal{Z},a)dx\end{array}\right.
\end{aligned}$$

where for each $\mathcal{Z}$, IC constraint and IR constraint are binding. Therefore, we can find the extreme value of $L^*(\boldsymbol{\delta}, \lambda^*(\mathcal{Z}), \mu^*(\mathcal{Z}))$ by choosing $(\boldsymbol{\delta}, \lambda^*(\mathcal{Z}), \mu^*(\mathcal{Z}))$, and the lower bound of profit loss is,

$$\Delta\Pi(\boldsymbol{\delta})^{lb} = \Delta\Pi(\boldsymbol{\delta}_{\min}) = \sum \frac{n_{\mathcal{Z}}}{n}\{\mathbb{E}w_i - \mathbb{E}s^*(x_i, \mathcal{Z}, \boldsymbol{\delta}_{\min})\}; \tag{1.27}$$

and the upper bound of profit loss is

$$\Delta\Pi(\boldsymbol{\delta})^{ub} = \Delta\Pi(\boldsymbol{\delta}_{\max}) = \sum \frac{n_{\mathcal{Z}}}{n}\{\mathbb{E}w_i - \mathbb{E}s^*(x_i, \mathcal{Z}, \boldsymbol{\delta}_{\max})\}. \tag{1.28}$$

Based on the above estimations, we can also figure out which group can be the most profitable. The testing for selection and matching can be done similarly.

Theorem 7 can be extended to deal with non-parametric identification where both the monetary utility and score function are unknown. If we assume that the output is generated by effort $\mathcal{Z}$ and random noise as follows

$$y(x) = m(a) + \mathcal{Z}'\beta + \varepsilon, \tag{1.29}$$

where $y(.)$ is an unknown monotone function, and $m(a)$ the p.d.f. of ε are unknown. The model (29) falls into transformation models in econometrics literature (Horowitz, 1996, Chen, 2002, among others). It is known that β can be efficiently estimated without specifying the error function, and $y(.)$ and $l_a(x)$ can be estimated consistently. Particularly, when $y(x) = x$, the model becomes a classical adaptive estimation in econometrics (Newey, 1988, Ai, 1997, *among others*). Discussion of those estimations is beyond the scope of this paper, but in our real example, we will perform an addaptive estimation, where individual heterogeneity is allowed.

1.5.2 Heterogeneous Shock to Each Observation

The analysis can also be extended to the situation where additional individual specific shock exists. We suppose the shock ε is drawn according to p.d.f. $p(\varepsilon)$, which is uncontrollable by the agent such as unconscious mistake, trembling hand, and so on. Then the p.d.f. of output becomes $f(x,\mathcal{Z}, a\,|\varepsilon)$ given the shock ε, where ε are unobserved by all contract parties before effort is put. In this case, the first order condition for contract optimality is

$$\frac{1}{u'(s(x,\mathcal{Z}))} = \lambda + \mu\frac{\int f_a(x,\mathcal{Z},a\,|\varepsilon)p(\varepsilon)d\varepsilon}{\int f(x,\mathcal{Z},a\,|\varepsilon)p(\varepsilon)d\varepsilon}.$$

Then we can repeat all the analyses in the previous sections by using marginal p.d.f. $\int f(x,\mathcal{Z},a\,|\varepsilon)p(\varepsilon)d\varepsilon$, instead of $f(x,\mathcal{Z},a)$. If both $f(x,\mathcal{Z},a\,|\varepsilon)$ and $p(\varepsilon)$ is functionally specified, then we can estimate the parameters by maximizing conditional likelihood or maximum marginal likelihood. This excise could be used in cross section data analysis.

1.5.3 Non-deterministic Contract Process

If contract is not a deterministic process, where the payment w_i is generated by some stochastic process. In this case, the contract optimality refers to the optimality of non-stochastic moment $s(x_i,\mathcal{Z}) = \mathbb{E}[w_i\,|x_i\,,\mathcal{Z}]$ because w_i itself is not a measurable function. We can characterize the problem slightly different as problem (P1). Let the conditional distribution of w given $s(x,\mathcal{Z})$ be $\phi(w\,|s(x,\mathcal{Z}))$, hence,

$$s(x,\mathcal{Z}) = \int w\phi(w\,|s(x,\mathcal{Z}))dw.$$

Therefore the first order condition for the principal's profit maximization problem is

$$\frac{1}{\frac{\partial}{\partial s}\int u(w)\phi(w\,|s(x,\mathcal{Z}))dw} = \lambda + \mu\frac{f_a(x,\mathcal{Z},a)}{f(x,\mathcal{Z},a)}.$$

If contract is a deterministic process, $\frac{\partial}{\partial s}\int u(w)\phi(w\,|s(x,\mathcal{Z}))dw = u'(s)$, back to the condition (1). If $\frac{\phi_s(w|s(x,\mathcal{Z}))}{\phi(w|s(x,\mathcal{Z}))} = \frac{\partial \log\phi(w|s(x,\mathcal{Z}))}{\partial s}$, $l_a(x,\mathcal{Z},a)$ and $u(w)$ is functionally specified, then our analyses in section 3 are still applicable. If $w_i = s(x_i) + \nu_i$, $x_i = g(\mathcal{Z}_i) + \varepsilon_i$ (ν_i and ε_i are noise term form unknown distribution) and $u(w)$ are unknown, in principle, the analyses in section 4 will apply. This framework could be used to analyze cross-sectional contract, such as sharecropping, insurance and CEO payment. We will leave this to future research.

1.5.4 Asymmetric Information

A more general case discussed by Holmstrom (1979) is to consider the existence of post contract asymmetric information, corresponding to Harris and Raviv's (1979) model 2. Some recent case studies regarding piece rate contract (Paarsch and Shearer, 1999, 2000), can be regarded as this type of inquiry. In Paarsch and Shearer's model, the principal can not observe the realization of the shock that the agent observes, given the contract has been written and there is no renegotiation after the agent realizes the shock. In this situation, the test for the optimality is still similar but more complicated. We leave this interesting topic to future research.

1.6 A Real Evidence

1.6.1 Data Description

The data were collected from a textile factory in Zhejiang Province, China. The payroll data document 50 cotton weavers' monthly output, monthly wage and working days in each month from June 2006 to May 2008 (See Table 2). All these workers are female, and paid by piece rate contract at the same percentage 18.75% per unit of output. At the beginning of the observation window, their working experience, age, and education level were also documented.

Table 2. Summary Statistics of Cotton Weaver's Contract Data

Variables	obs	Mean	std. dev	Min	Max
Age	1200	38.36	5.83	30	49
Education (Year)	1200	9	1.83	5	12
Ini. Experience (Year)	1200	3	0.84	1	3
Monthly wage (Yuan)	1200	1346.93	368.20	108	2415.74
Monthly output (Meter)	1200	481.05	131.50	38.57	862.76
Work days	1200	25.83	2.35	8	29

1.6.2 Test for Contract Optimality with Wage-Output Data Only

The first test is to identify the contract optimality using wage-output data only, without utilizing any additional information partition. We assume the output generating process is additive as follows

$$x_i = a + \varepsilon_i$$

with ε from an unknown p.d.f.

First of all, note that the prestage Kolmogrov-Smirnov normality test is not rejected[20]. Inspired by this, according to theory, if the score is close to linear function of x, then the inverse marginal utility function should be linear in wage. So log utility is the most favorable functional form to rationalize the data. Our first estimation is to test whether log utility can rationalize the data.

We estimate the score function $-\frac{f'(\varepsilon)}{f(\varepsilon)}$ based on kernel estimation (leave-one-out). The band width is chosen to minimize the expected MSE[21] $\int \mathbb{E}(\hat{l}_a - l_a)^2 f(\varepsilon)d\varepsilon$, where

$$\begin{aligned}&\int \mathbb{E}(\hat{l}_a - l_a)^2 f(\varepsilon)d\varepsilon \\ = \; & \frac{1}{nb^3}\int [K'(\varphi)]^2 d\varphi \mathbb{E}\frac{1}{f(\varepsilon_i)} + \mathbb{E}\left(\frac{f'''(\varepsilon_i)}{f(\varepsilon_i)} - \frac{f'(\varepsilon_i)f''(\varepsilon_i)}{f(\varepsilon_i)^2}\right)^2 \left(\frac{b^2}{2}\int \varphi^2 K(\varphi)d\varphi\right)^2 . (1.30)\end{aligned}$$

Since our pre-assumption of the output is normal, we choose b according to the normal distribution of $f(\varepsilon)$ which results in $\int \left(\frac{f'''(\varepsilon)}{f(\varepsilon)} - \frac{f'(\varepsilon)f''(\varepsilon)}{f(\varepsilon)^2}\right)^2 f(\varepsilon)d\varepsilon = \frac{28}{\sigma^6}$. Using Gaussian kernel, and using sample range to replace $\int d\varepsilon$, we have

$$b_0 = \left(\frac{\frac{3}{4\sqrt{\pi}}(\varepsilon_{\max} - \varepsilon_{\min})}{\frac{1}{4}\frac{28}{\hat{\sigma}^6}}\right)^{\frac{1}{7}} n^{-\frac{1}{7}} = \left(\frac{\frac{3}{\sqrt{\pi}}(\varepsilon_{\max} - \varepsilon_{\min})}{28}\right)^{\frac{1}{7}} \hat{\sigma}^{\frac{6}{7}} n^{-\frac{1}{7}} \cong 41.5742.$$

However, under this band width, the second term (biased term) in (30) will be 3 times larger than the first term, which is not reasonable since we want to avoid estimating high order derives in the biased term. So we adjust b_0 by $b = b_0(\log n)^{-\frac{1}{4}}$ which will make the second term only a tenth of the first term on average. So the final band width is

$$b = 25.4777.$$

[20]Our pre-stage test tells us that the Kolmogrov-Smirnov test for normality can not be rejected at p-value 0.65. And Pearson Chi-square test for normality is also not rejected.

[21]We can also choose b minimizing approximated mean intergrated squre error (AMISE, see Jones, 1991) of derivative estimator of $f'(\varepsilon)$ since estimation of $f'(\varepsilon)$ requires larger band width. Minizing AMISE based on the normal distribution associated with Gassian Kernel yields,

$$b = \left(\frac{\frac{3}{4\sqrt{\pi}}}{4\frac{15}{16\sqrt{\pi}}\frac{1}{\hat{\sigma}^7}}\right)^{\frac{1}{7}} n^{-\frac{1}{7}} = 5^{-\frac{1}{7}}\hat{\sigma}n^{-\frac{1}{7}} \cong 37.948,$$

which is close to our choice.

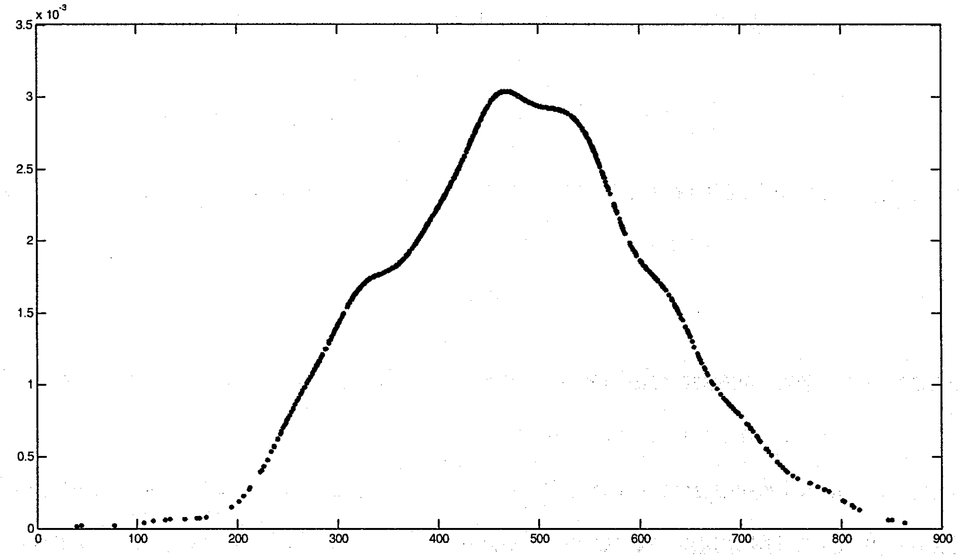


Figure 1-1: The kernel plot of density (b=25.4777)

Under this band width, we estimate $f(\varepsilon)$ and score (See Fig. 1 for p.d.f. plot and Fig 2. for score plot). The basic results are reported in Table 3.

In Table 3, we find that both $\hat{M}$ and $\hat{K}$ are significantly greater than zero, which indicates that moral hazard exists and the second order condition is met (at least locally). And our test for optimality $H_0 : \rho = 0$ is rejected at 0.01 confidence level, if the smaller order biasness of estimator $\hat{\rho}$ is not taken into consideration. When the bias is substracted, the null hypothesis is rejected at 0.05 confidence level[22]. And we also find that the profit loss is about 57 RMB, 1% of total profit, compared with a potential optimal contract which is consistent with log utility. The finding indiciates that log utility seems to explain the variation of wage very well in terms of small profit loss. However the test significant rejects that log utility is the most favorable type of utility, as we will see in next subsection.

[22]The boostrapped standard error is slightly higher than the theoretical asymptotic variance, but the null is still rejected even using bootstrapped standard error.

Table. 3. Optimality Test Using Wage-Output Information Only

b=25.48	Log utility	Unknown utility
$\hat{Z} \rightarrow \mathbb{E}l_a^2$	$6.1736{\times}10^{-5}(3.7297{\times}10^{-6})$	$6.1736{\times}10^{-5}(3.7297{\times}10^{-6})$
$\hat{\lambda} \rightarrow \mathbb{E}h$	$1.3469{\times}10^3$ (10.6289)	$5.8151{*}10^{-4}(1.6868 * 10^{-4})$
$\hat{\mu} \rightarrow \frac{Cov(h,l_a)}{\mathbb{E}l_a^2}$	4.2820×10^4 (5.4341×10^3)***	0.3509 (0.0139)***
$\hat{M}(\hat{a})$	0.0022 (0.0016)*	6.4956 (0.0301)***
$\hat{K}(\hat{a})$	2.0343×10^{-6} (5.8544×10^{-5})***	4.4946×10^{-4} (7.1030×10^{-4})
$\hat{\rho} = \frac{\hat{J}}{\sqrt{\hat{Z}\hat{Q}}}$	0.9138 (0.0239)***[0.0277]	0.8858 (0.042)***
Adjusted $\hat{\rho}$	0.9612 (0.0239)**[0.0277]	0.9723 (0.042)
$\widehat{s^*(x)}$	$10^3(1.2896 + 4.3024 l_a)$	n.a.
$\Pi^{sb} - \Pi^{ob}\{\%\}$	57.2442{0.98%}	55.3682{0.96%}
Adjoint Equation	11.2146 (?)	-1.0134 (46.3283)

Since the optimality hypothesis is rejected for log utility, we run a non-parametric optimality test. We try to find a smooth monotone function $h(w)$ to rationalize the first order condition, this is

$$h(w_i) = \lambda + \mu \hat{l}_a(x_i).$$

To estimate $h(w)$, we adopt the estimate proposed by Dette et al (2006). We use E-kernel, and choose the bandwith of b_h according to the critierion

$$b_h = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (l_a(x_{i+1}) - l_a(x_i))^2 = 3.7492 * 10^{-5},$$

which also numerically coincides with the criterion minimizing expected MSE of $\hat{h}(w)$,

$$b_h \in \arg\min_{b_h} \frac{b_h^4}{4} \int m_2^2 (2h'\phi')^2 \phi(w)^2 dw + \left(\int [K'(\varphi)]^2 d\varphi \int K_r(\varphi)^2 d\varphi \right) \frac{1}{nb_h nb^3} \int \frac{1}{\phi(w) f(\varepsilon)} \phi(w)^2 dw.$$

The results show that the recovered inverse marginal utility is almost linear function, and we are no longer able to reject the null hypothesis. The profit loss is even smaller (about 0.5%).

This shows that log utility explains the data very well, and pasts the non-parametric test (We also did test based on daily output, and the results are the same)[23].

1.6.3 Tests for Optimality by Using Individual Fixed Effect

The second exercise is to utilize additional information partitions. Assuming the output is generated by

$$x = \mathcal{Z}\beta + \varepsilon$$

where $\mathcal{Z}$ is individual dummy variable, and we let $a = \beta_1$ as effort parameter. We can do parametric or non-parametric test.

For parametric case, we assume score is linear in x, thus the most likely funtional form to rationalize the data is CRRA or log utility familiy. We run OLS to estimate β first, then the score can be estimated by,

$$l_a(x_i, \hat{\beta}) = \hat{\varepsilon}_i = x_i - \mathcal{Z}\hat{\beta}.$$

The OLS regression shows that the majorities of $\hat{\beta}_i$'s are significant, indicating the presence of heterogeity among agents. Our testing results are reported in Table 4. We find that in this parametric case, the null hypothesis is not rejected. This means that the heterogeneity among agent is not enough to beat the optimality of contract, even though these heterogeity is significant based on the output data. That is said, if we believe that the contract is optimal, the agent's utitity should be close to log or CRRA (these assumptions are employed in empirical macro literature and labor literature). We also run non-parametric testing, by doing adaptive estimation, but result seems similar. However, if we take the selection issue into considerations, then the data shows that principal may fail to match the right agent or technology to contract.

[23]We can also choose the bandwidth by Cross Validation $\min_h \int (\hat{f}_n'(x))^2 dx - \frac{2}{n} \sum \hat{f}_{-i}'(X_i)$.

Table 4. Test Results for Piece-Rate Contract (Linear Score).

	CRRA	log utility
$\hat{\gamma}_0$	0.4117(0.0047)***	0(*n.a.*)
$\hat{Z} \to \mathbb{E}l_a^2$	1.2422*10^4(507.6182)	1.2422*10^4(507.6182)
$\hat{Q} \to Var(h)$	131.9964(6.3309)	1.3557*10^5(5.4144*10^3)
$\hat{\lambda} \to \mathbb{E}h$	66.6460(0.4690)	1.3469*10^3(10.6289)
$\hat{\mu} \to \frac{Cov(h,l_a)}{\mathbb{E}l_a^2}$	0.0878(0.0181)***	2.7977(0.3686)***
$\hat{M}(\hat{a})$	530.7616(148.4872)***	29.2213(23.0592)*
$\hat{\mathcal{K}}(\hat{a})$	1.1690*10^8(1.4567*10^8)	1.9104*10^3(3.4128*10^3)
$\hat{\rho} = \frac{\hat{j}}{\sqrt{\hat{Z}\hat{Q}}}$	0.8516(0.8454)	0.8466(0.1073)*
$\hat{\mathcal{Q}}(0.99)$	0.3260**	0.3305**
$\widehat{s^*(x)}$	$(68.5600+0.0883l_a)^{\frac{1}{0.4117}}$	$10^3(1.3317 + 0.0028l_a)$
$\Pi^{sb} - \Pi^{ob}$	8.8016(1.6821)***	15.2283(8.6577)*
%	0.15%	0.26%
$\frac{(\hat{\mathbb{E}}w_i - \hat{\mathbb{E}}\hat{s}^*)^2}{AVar(\widehat{\Delta\Pi})} \to^d \chi_1^2$	0.0484[0.1741]	0.0340[0.1463]
$\hat{\Pi}^{fb} - \hat{\Pi}^{sb}$	23.5366(1.6821)***	42.0576(8.6577)***
$\hat{\Pi}^{fb} - \hat{\Pi}^{ob}$	32.3382(20.9582)*	57.2859(15.8125)***
$H_{0\gamma} : V_\gamma^* = 0$	1.3029*10^4(79.7825)***	n.a.
or $H_{0\gamma} : \hat{\gamma} = \gamma^*$	-	0.1380(1.4983*10^{-4})***
$H_{0\theta} : V_\theta^* = 0$	n.a.	n.a.
Adjoint Equation	4.6440*10^{10}(9.5883*10^9)***	1.4607*10^5(2.1829*10^4)***

1.6.4 Allowing Preference Heterogeneity

If we allow each individual to be different in their utility parameter within CRRA family. Then we choose a set of $\gamma = (\gamma_1, ..., \gamma_{50})$ to maximize the criterion function. And we also assume that score is linear function of output for each individual j,

$$l_a(x_{ji}, \mathcal{Z}_i, a_j) = x_{ji} - \mathbb{E}[x_{ji}/\mathcal{Z}_i]$$

By doing the same estimation, we have the following results:

Table 7. Testing Results for Heterogeneous Preference

variables	Min	Max	Median	Mean	std. dev
$\hat{\gamma}$	-9.1171*10^{-6}	6.7257*10^{-6}	-2.9671*10^{-8}	8.5095*10^{-8}	2.6726*10^{-6}
$\hat{Z} \to \mathbb{E}l_a^2$	5.1745	33.9309	12.2156	11.5514	5.0451
$\hat{Q} \to Var(h)$	40.5673	266.0103	90.5633	95.7705	39.5293
$\hat{\mu} \to \frac{Cov(h,l_a)}{\mathbb{E}l_a^2}$	2.6832	2.6835	2.6833	2.6833	3.5521*10^{-5}
$\hat{\rho}$	0.9583	0.9583	0.9583	0.9583	4.8361*10^{-14}
χ_1^2 [p-value]	0.0299[0.8628]				

The results shows that the p-value is significant larger than before and we can not reject the null hypothesis. The reason is that there are only 24 observations for each individual, which dramatically increases the degree of freedom.

1.7 Conclusions and implications

Moral hazard theory has been an important element of modern economics since the 1970s, while there has been increasing interest in testing contract theory in recent decades. But as Chiappori and Salanie (2003) highlight, testing of contract theory has been hampered due to self-selection, unobservable action and preference, and other factors. On the one hand, empirical regressions using exogenous variations face the difficulties of self-endogeneity and separation between the exogenous factors and variation of effort which also depends on those factors[24]. On the other hand, without variation of effort, it is hard to back out the alternatives of observed choice by the agent. These difficulties become somewhat of a barrier to examining contract theory and also weaken the generality of the conclusion drawn by a specific case study, such as piece rate (Paarsch and Shearer, 1999, 2000, Haley, 2003, *among others*) and sharecropping (Ackerberg and Botticini, 1999, among others). Based directly on the first order conditions for contract optimality, this paper provides a set of conditions to identify moral hazard problem without relying on exogenous variation. As indicated by simulations and an empirical example, our

[24]For example, suppose we observe exogenous variation $\mathcal{Z}$, which affects variation of contract w. However, the agent effort should depend on $\mathcal{Z}$ and w, say $a(\mathcal{Z}, w)$. It is hard to separately identify the mechanism leading to the change of effort and evaluate the efficiency of contract.

approach seems to work well. Furthermore, this technique has broad applications to incentive design and allows more refined tests of contract theory (particularly moral hazard theory) using observable data. The data in this paper shows that piece rate contract can be rationalized by log utility very well in the sense that the lower bound of profit loss is very small, given that normality of output distribution is not rejected. Our approach can handle more complicated contracting situations such as CEO compensation, insurance co-pay plans, and taxation codes, though we use a relatively simple contract data set in this paper. We expect to further examine this approach in future empirical research.

This paper is the first paper taking a close look at the testing of the contract theory based on the first order approach, especially in a situation with moral hazard. However, there are a couple of authors also looking at the first order conditions of contract optimality in adverse selection (d'Haultfeuille and Fevrer, 2007). Their paper aims to identify contract optimality in pure adverse selection where the observed data are quantity of trades and transfers, but the agent's type, the distribution of types, the agent's cost function and the production function are not observed. These two independent papers have similar motivation, though our approach is very different.

There are several extensions of the current findings worth noting. The first one is to consider an environment in which an agent has privately-known productivity and there exist both moral hazard and adverse selection (see Myerson, 1982; Faynzilberg and Kumar, 2000). Limited here by length, we deal with this question in a separate paper which shows that our approach applies to this setting (Ke, 2008). The second extension is to examine algorithms for achieving contract optimality through repeated experiments (trial and error) without knowing the functions, mimicking how the principal behaves in the real world, and find evidence to support this type of behavior. The third extension is to extend the framework to a dynamic setting, where interactions due to learning between principal and agent is activated. All of these extensions seem to be challenging but interesting for future exploration.

1.8 Appendices

1.8.1 Proof of Propositon 1

Proof. By contradiction.

First, for a given contract $s(x)$ and output distribution $f(x,a)$, when the Lagrangian multiplier $\hat{\lambda}$ and $\hat{\mu}$ are fixed, there is only a unique utility $(\hat{u}(.), \hat{c}(a))$ rationalizing $s(x)$ since $\mathcal{U}$ is monotonically increasing and concave and $\hat{q} = \hat{\lambda} + \hat{\mu} l_a(x,a) > 0$. This is,

$$\hat{h}(\hat{s}) = \hat{\lambda} + \hat{\mu} l_a(x,a) \text{ or } \hat{s} = \hat{r}(\hat{q}).$$

Hence, if the underlying utility is different from $\hat{u}(.)$, the difference between u and $\hat{u}$ is up to an affine transformation in their inverse marginal utilities, i.e., $h(.) = k_1 \hat{h}(.) + k_2$. Therefore, the optimal contract under utility (u, c) should be:

$$h(s) = \lambda + \mu l_a(x,a) \text{ with } \lambda > 0 \text{ and } \mu > 0 \text{ or } s = r(q).$$

Since $h(.) = k_1 \hat{h}(.) + k_2$, it can be seen that $\hat{r}(\hat{q}) = r(\alpha_1 q + \alpha_2)$ with $\alpha_1 = \frac{\mu}{\hat{\mu}}$, $\alpha_2 = \lambda - \frac{\mu}{\hat{\mu}}\hat{\lambda}$. We want to prove $\alpha_1 = 1$ and $\alpha_2 = 0$.

Because $\hat{s}$ is the true contract offered to the agent, and $l_a(x,a)$ is generated by the agent's true effort a, therfore, keeping the agent's utility the same as that of contract $\hat{s}(x)$, we have

$$\begin{aligned} \int u(\hat{r}(\hat{q})) l_a(x) f(x,a) dx &= \int u(r(q)) l_a(x) f(x,a) dx \\ \text{and } \int u(\hat{r}(\hat{q})) f(x,a) dx &= \int u(r(q)) f(x,a) dx, \end{aligned}$$

yielding

$$\mathbb{E}[u(\hat{r}(\hat{q})) - u(r(q))]q = 0.$$

By the concavity of $u(.)$, if $\hat{r}(\hat{q}) \neq r(q)$ with positive probability, then,

$$\mathbb{E}(\hat{r}(\hat{q}) - r(q)) > 0.$$

At the same time, when $\hat{r}(\hat{q}) = r(\alpha_1 q + \alpha_2)$,

$$\mathbb{E}(\hat{r}(\hat{q}) - r(q)) < \mathbb{E}[u(r(\alpha_1 q + \alpha_2)) - u(r(q))](\alpha_1 q + \alpha_2) = 0,$$

we have $\hat{r}(\hat{q}) = r(q)$ pointwise. Q.E.D. ■

1.8.2 Proof of Proposition 2

Proof. The proof of theorem 3 is a special case of theorem 3, which is proved in Appendix 3.

(i) In this simpler situation, we have

$$\begin{aligned} nAVar(\hat{\rho}) &= \frac{1}{4QZ}\left[\begin{array}{c} 4\{Var((h-\lambda)l_a) + Z^{-1}[\mathbb{E}(h-\lambda)l_{aT}]^2\} + \frac{J^2}{Q^2}Var(\mathcal{Q}) \\ +(\frac{J}{Z})^2\left[Var(l_a^2) + 4Z^{-1}\left(\mathbb{E}l_a l_{aT}\right)^2\right] - 4(\frac{J}{Q})Cov((h-\lambda)l_a, \mathcal{Q}) + 2\frac{J^2}{QZ}Cov(\mathcal{Q}, l_a^2) \\ -4(\frac{J}{Z})Cov((h-\lambda)l_a, l_a^2) - \frac{8J}{Z^2}\left[\mathbb{E}(h-\lambda)l_{aT}\mathbb{E}l_a l_{aT}\right] \end{array}\right] \\ &= \frac{1}{4QZ}\left[\begin{array}{c} 4Var((h-\lambda)l_a) + \frac{J^2}{Q^2}Var(\mathcal{Q}) + (\frac{J}{Z})^2 Var(l_a^2) \\ -4(\frac{J}{Q})Cov((h-\lambda)l_a, \mathcal{Q}) + 2\frac{J^2}{QZ}Cov(\mathcal{Q}, l_a^2) - 4(\frac{J}{Z})Cov((h-\lambda)l_a, l_a^2) \end{array}\right] \\ &= \frac{1}{4QZ}Var(2(h-\lambda)l_a - \frac{J}{Q}\mathcal{Q} - \frac{J}{Z}l_a^2) \end{aligned}$$

(ii) Especially, under the null hypothesis, the asymptotic distribution of $\hat{\rho}(\hat{T}) - \rho$ here is a special case of theorem 3, and $\hat{T}$ here is a scalar and $\hat{\rho}_{TT} = 0$, therefore,

$$n[\hat{\rho}(\hat{T}) - 1] = -\frac{1}{2Z}n(\hat{\mathbb{E}}l_a)^2 \rightarrow^d -\frac{1}{2}\chi_1^2.$$

Q.E.D. ■

1.8.3 Proof of Proposition 3

Proof. (i) We want to estimate the profit loss due to the suboptimality of the contract. Suppose we observe a real payment w, and detect the effort level a and some exogenous parameter θ as well. Given the estimated γ, the potential optimal contract $s^*(x) = h^{-1}[\lambda^* + \mu^* l_{a^*}(x, a, \theta), \gamma]$ can be found by the following:

$$(a^*, \theta^*) = (a, \theta);$$

$$\int u(s^*(x),\gamma)l_a(x,T)f(x,T)dx = \int u(w,\gamma)l_a(x_i,T)f(x,T)dx;$$

$$\int u(s^*(x),\gamma)f(x,T)dx = \int u(w,\gamma)f(x,T)dx.$$

Among the above conditions, the first line means that the comparison is based on the same effort a as we estimated; the second line means that the IC constraint under the potential optimal contract will be binding, thus, by the monotonicity of $c'(a)$, both $s^*(x)$ and w_i will implement the same effort. The third line means the agent will get the same utility under two different contracts. Plugging $a^* = a$ into the second and the third equation and combining them, we can solve λ^* and μ^*. Thus, the profit distance (per capita) to the potential optimal contract is the following:

$$\begin{aligned}
\Delta\Pi &= \Pi(s^*(x)) - \Pi(s(x)) \\
&= \int wf(x,T)dx - \int s^*(x)f(x,T)dx \\
&= \int wf(x,T)dx - \int h^{-1}[\lambda^* + \mu^* l_a(x,T^*),\gamma]f(x,T^*)dx
\end{aligned}$$

In our problem, true effort level a and true contract $s(x)$ are unknown. We thus formulate an estimator of profit loss in terms of sample analogues, which is solve $(\hat{\lambda}^*, \hat{\mu}^*, T^*)$ solving the following equations:

$$\begin{aligned}
\hat{T}^* &= \hat{T} \\
\hat{\mathbb{E}}u(h^{-1}[\hat{\lambda}^* + \hat{\lambda}^* l_a(x_i,\hat{T}),\gamma],\hat{\gamma})l_a(x_i,\hat{T}) &= \frac{1}{n}\sum_{i=1}^{n} u(w_i,\hat{\gamma})l_a(x_i,\hat{T}) \\
\hat{\mathbb{E}}u(h^{-1}[\hat{\lambda}^* + \hat{\lambda}^* l_a(x_i,\hat{T}),\hat{\gamma}],\hat{\gamma}) &= \frac{1}{n}\sum_{i=1}^{n} u(w_i,\hat{\gamma}),
\end{aligned}$$

where $T = (a,\theta')'$ and $\hat{\gamma} \in \arg\min_\gamma \hat{\Lambda}(\gamma)$.

(ii)-(iii). The rest parts of proof are techical, so we put them on the technical supplement (TS1). Q.E.D. ■

1.8.4 Proof of Corollary 1

Proof. (i) It is obvious to see that w^{fb} is the full insurance contract through which the principal implements the same effort, keeping the agent's utility the same. (ii) Note that

$$\hat{w}^{fb} - w^{fb} = \frac{\partial u^{-1}(u(w^{fb},\gamma),\gamma)}{\partial \gamma'}(\hat{\gamma}-\gamma) + \frac{1}{u'(w^{fb})}(\hat{\mathbb{E}}u(w_i,\hat{\gamma}) - \mathbb{E}u(w_i,\gamma)) + h.o.$$

$$\hat{w}^{fb} - w^{fb} = \frac{du^{-1}(\hat{\mathbb{E}}u(w_i,\gamma),\gamma)}{d\gamma'}(\hat{\gamma}-\gamma) + h.o.$$

therefore,

$$nAVar(\hat{w}^{fb}) = \frac{du^{-1}(\mathbb{E}u(w_i,\gamma),\gamma)}{d\gamma'}\mathbf{\Gamma}\frac{du^{-1}(\mathbb{E}u(w_i,\gamma),\gamma)}{d\gamma}$$

since $ACov(\hat{\mathbb{E}}u(w_i),\hat{\gamma}) = 0$. And we have

$$\begin{aligned} nAVar(\hat{w}^{fb}) &= \frac{\partial u^{-1}(u(w^{fb},\gamma),\gamma)}{\partial \gamma'}\mathbf{\Gamma}\frac{\partial u^{-1}(u(w^{fb},\gamma),\gamma)}{\partial \gamma} \\ &\quad + h(w^{fb})^2[Var(u(w_i)) + \mathbb{E}u_{\gamma'}(w_i)\mathbf{\Gamma}\mathbb{E}u_{\gamma}(w_i)] + 2\frac{\partial u^{-1}(u(w^{fb},\gamma),\gamma)}{\partial \gamma'}\mathbf{\Gamma}\mathbb{E}u_{\gamma}(w_i) \end{aligned}$$

The rest parts of proof are technical, so we put them in technical supplement 2. Q.E.D. ■

1.8.5 Proof of Theorem 3

Proof. (i) First, we need to check the consistence of $\hat{\rho}$ under the null hypothesis. Using the sample analogue $\hat{\Lambda}(\gamma,\mathbf{T})$ as an approximation of $\Lambda(\gamma,\mathbf{T})$, for the uniform convergence to apply, we need to confirm

$$\sup_{(\gamma,\mathbf{T})}\left|\hat{\Lambda}(\gamma,\mathbf{T}) - \Lambda(\gamma,\mathbf{T})\right| \to^{p} 0$$

Note that,

$$\begin{aligned}
& \sup_{(\gamma,\mathbf{T})} \left|\hat{\Lambda}(\gamma,\mathbf{T}) - \Lambda(\gamma,\mathbf{T})\right| \\
= & \sup_{(\gamma,\mathbf{T})} \left|\frac{1}{n}\sum\left[\varepsilon_i(\gamma,\mathbf{T}) - (\bar{J}-J)\frac{h(w_i,\gamma)-\bar{h}(w,\gamma)}{J^{*2}} + (\bar{Z}-Z)\frac{l_a(x_i,\mathbf{T})}{Z^{*2}}\right]^2 - \Lambda(\gamma,\mathbf{T})\right| \\
\leq & \; 3\sup_{(\gamma,\mathbf{T})}\left|\frac{1}{n}\sum\varepsilon_i(\gamma,\mathbf{T})^2 - \Lambda(\gamma,\mathbf{T})\right| + 3\sup_{(\gamma,\mathbf{T})}(\bar{J}-J)^2\frac{1}{n}\sum\left(\frac{h(w_i,\gamma)-\bar{h}(w,\gamma)}{J^{*2}}\right)^2 \\
& +3\sup_{(\gamma,\mathbf{T})}(\bar{Z}-Z)^2\frac{1}{n}\sum\left(\frac{l_a(x_i,\mathbf{T})}{Z^{*2}}\right)^2
\end{aligned}$$

Note also that $J^* \to^p J$, $\hat{J} \to^p J$, $Z^* \to^p Z$, $\hat{Z} \to Z$, $\frac{1}{n}\sum\left(h(w_i,\gamma)-\bar{h}(w,\gamma)\right)^2 \to Q < \infty$, $\frac{1}{n}\sum l_a^2(x_i,\hat{\mathbf{T}}) \to Z < \infty$. The only item left is to show that $\sup_\gamma \left|\frac{1}{n}\sum\varepsilon_i(\gamma,\mathbf{T})^2 - \Lambda(\gamma,\mathbf{T})\right| \to^p 0$, which is implied by the uniform law of large numbers, because of

$$\begin{aligned}
& \mathbb{E}[\sup_{(\gamma,\mathbf{T})}\left(\frac{h(w,\gamma)-\mathbb{E}h(w,\gamma)}{\mathbb{E}h(w,\gamma)l_a(x,\mathbf{T})} - \frac{l_a(x,\mathbf{T})}{\mathbb{E}l_a(x,\mathbf{T})^2}\right)^2] \\
\leq & \; 2\mathbb{E}[\sup_{(\gamma,\mathbf{T})}\left(\frac{h(w,\gamma)-\mathbb{E}h(w,\gamma)}{\mathbb{E}h(w,\gamma)l_a(x,\mathbf{T})}\right)^2 + \left(\frac{l_a(x,\mathbf{T})}{\mathbb{E}l_a(x,\mathbf{T})^2}\right)^2] \\
= & \; 2\mathbb{E}[\sup_{(\gamma,\mathbf{T})}\left(\frac{h(w,\gamma)-\mathbb{E}h(w,\gamma)}{\mathbb{E}h(w,\gamma)l_a(x,\mathbf{T})}\right)^2] + 2\mathbb{E}\sup_{\mathbf{T}}\left(\frac{l_a(x,\mathbf{T})}{\mathbb{E}l_a(x,\mathbf{T})^2}\right)^2 \\
\leq & \; 2\mathbb{E}[\sup_{\gamma\in\mathbf{\Gamma}}(h(w,\gamma)-\mathbb{E}h(w,\gamma))^2]\sup_{(\gamma,\mathbf{T})}\frac{1}{[\mathbb{E}h(w,\gamma)l_a(x,\mathbf{T})]^2} + \frac{2}{\mathbb{E}l_a(x,\mathbf{T}^*)^2} \\
< & \; \infty
\end{aligned}$$

where $\mathbf{T}^* = \arg\max_{\mathbf{T}}\left(\frac{l_a(x,\mathbf{T})}{\mathbb{E}l_a(x,\mathbf{T})^2}\right)^2$. And the last step comes from $\mathbb{E}h(w,\gamma)l_a(x,\mathbf{T}) > 0$ and $\mathbb{E}l_a(x,\mathbf{T}^*)^2 > 0$. Thus, $\sup_\gamma\left|\hat{\Lambda}(\gamma,\mathbf{T}) - \Lambda(\gamma,\mathbf{T})\right| \to^p 0$.

Based on the above conditions, the extremum estimator of $(\hat{\gamma},\hat{\mathbf{T}}) \in \arg\min_{(\delta,\lambda,\mu)}\hat{\Lambda}(\gamma,\mathbf{T})$ is a consistent estimator of the minimizer of $\Lambda(\gamma,\mathbf{T})$.

(ii) We can derive the asymptotic distribution of $\hat{\rho}(\hat{\boldsymbol{\delta}})$ based on the asymptotic distrbiution of $\hat{\boldsymbol{\delta}}$. With the additional assumptions that

$$\sup_\gamma\left\|\frac{\partial^2\hat{\Lambda}(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'} - \frac{\partial^2\Lambda(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}\right\| \to^p 0 \text{ and } \sqrt{n}\hat{\Lambda}_{\boldsymbol{\delta}}(\boldsymbol{\delta}) \to^d \mathcal{N}(0,\Omega)$$

and $\frac{\partial^2 \Lambda(\boldsymbol{\delta})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'}$ is non-singular, the asymptotic normality of δ can be derived by the following (since we do not apply a two-step estimation):

$$0 = \hat{\Lambda}_{\boldsymbol{\delta}}(\boldsymbol{\delta}) = \hat{\Lambda}_{\boldsymbol{\delta}}(\boldsymbol{\delta}) + \hat{\Lambda}_{\boldsymbol{\delta}\boldsymbol{\delta}'}(\bar{\gamma}, \bar{\mathbf{T}})(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})$$

Therefore, we have,

$$\sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) = -[\hat{\Lambda}_{\boldsymbol{\delta}\boldsymbol{\delta}'}(\bar{\gamma}, \bar{\mathbf{T}})]^{-1}[\sqrt{n}\hat{\Lambda}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}\boldsymbol{\delta})]$$

Note that

$$\begin{aligned}
\hat{\rho}(\hat{\boldsymbol{\delta}}) - \rho(\delta) &= \hat{\rho}(\hat{\boldsymbol{\delta}}) - \hat{\rho}(\boldsymbol{\delta}) + \hat{\rho}(\boldsymbol{\delta}) - \rho(\delta) \\
&= \rho_{\boldsymbol{\delta}'}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) + \frac{1}{2}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})'\rho_{\boldsymbol{\delta}\boldsymbol{\delta}'}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) + \hat{\rho}(\boldsymbol{\delta}) - \rho(\delta) + h.o.
\end{aligned}$$

We find that under the null hypothesis,

$$\begin{aligned}
\hat{\rho}_{\gamma} &= \frac{1}{\sqrt{\hat{Z}}}\left(\frac{\hat{J}_{\gamma}(\boldsymbol{\delta})}{\sqrt{\hat{Q}(\gamma)}} - \frac{1}{2}\hat{Q}(\gamma)^{-\frac{3}{2}}\hat{J}(\boldsymbol{\delta})\hat{Q}_{\gamma}(\gamma)\right) \\
&= \frac{\hat{J}_{\gamma}(\boldsymbol{\delta})}{\mu\hat{Z}} - \frac{1}{2}\frac{\hat{J}_{\gamma}(\gamma)}{\mu^2\hat{Z}} + o(\frac{1}{n}) = o(\frac{1}{n})
\end{aligned}$$

and

$$\begin{aligned}
\hat{\rho}_{\mathbf{T}} &= \frac{1}{\sqrt{\hat{Q}}}\left(\frac{\hat{J}_{T}(\boldsymbol{\delta})}{\sqrt{\hat{Z}(T)}} - \frac{1}{2}\hat{Z}(T)^{-\frac{3}{2}}\hat{J}(\boldsymbol{\delta})\hat{Z}_{T}(T)\right) \\
&= \frac{1}{\sqrt{\hat{Q}}}\left(\frac{\frac{1}{2}\mu\hat{Z}_{T}(T) - \mu\hat{\mathbb{E}}l_a\hat{\mathbb{E}}l_{aT}}{\sqrt{\hat{Z}(T)}} - \frac{\mu}{2\sqrt{\hat{Z}(T)}}\hat{Z}_{T}(T) + o(\frac{1}{n})\right) \\
&= o(\frac{1}{n}),
\end{aligned}$$

and

$$
\begin{aligned}
&\hat{\rho}(\boldsymbol{\delta}) - \rho(\delta) \\
= \;& \frac{1}{J}\mu[\hat{Z} - Z - (\hat{\mathbb{E}}l_a)^2] - \frac{1}{2Q}\mu^2[\hat{Z} - Z - (\hat{\mathbb{E}}l_a)^2] - \frac{1}{2Z}(\hat{Z} - Z) \\
& - \frac{1}{2JQ}\mu^3[\hat{Z} - Z - (\hat{\mathbb{E}}l_a)^2]^2 - \frac{1}{2ZJ}\mu[\hat{Z} - Z - (\hat{\mathbb{E}}l_a)^2](\hat{Z} - Z) \\
& + \frac{1}{4J^2}\mu^2[\hat{Z} - Z - (\hat{\mathbb{E}}l_a)^2](\hat{Z} - Z) + \frac{3}{8}Q^{-2}\mu^4[\hat{Z} - Z - (\hat{\mathbb{E}}l_a)^2]^2 + \frac{3}{8}Z^{-2}(\hat{Z} - Z)^2 + h.o. \\
= \;& -\frac{1}{2Z}(\hat{\mathbb{E}}l_a)^2 - \frac{3}{8Z^2}[\hat{Z} - Z - (\hat{\mathbb{E}}l_a)^2]^2 - \frac{1}{4Z^2}[\hat{Z} - Z - (\hat{\mathbb{E}}l_a)^2](\hat{\mathbb{E}}l_a)^2 + \frac{3}{8}Z^{-2}(\hat{Z} - Z)^2 \\
= \;& -\frac{1}{2Z}(\hat{\mathbb{E}}l_a)^2 + \frac{1}{2Z^2}(\hat{Z} - Z)(\hat{\mathbb{E}}l_a)^2 + \frac{5}{8Z^2}(\hat{\mathbb{E}}l_a)^4 + o(n^{-\frac{3}{2}}),
\end{aligned}
$$

therefore,

$$
\begin{aligned}
n[\hat{\rho}(\hat{\boldsymbol{\delta}}) - \rho(\delta)] &= n[\hat{\rho}(\hat{\boldsymbol{\delta}}) - \hat{\rho}(\boldsymbol{\delta}) + \hat{\rho}(\boldsymbol{\delta}) - \rho(\delta)] \\
&= \frac{1}{2}\sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})'\rho_{\boldsymbol{\delta}\boldsymbol{\delta}'}\sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) - \frac{1}{2Z}(\sqrt{n}\hat{\mathbb{E}}l_a)^2 + h.o.
\end{aligned}
$$

Note that both $\sqrt{n}\hat{\mathbb{E}}l_a$ and $\sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})$ are asymptotically normal, then $n[\hat{\rho}(\hat{\boldsymbol{\delta}}) - \rho(\delta)]$ is quadratic form of a multivariate normal. We can rewrite it as the expression in theorem3. The typical value of quadratic form can be simulated or approximated. Q.E.D. ■

1.8.6 Proof of Proposition 5

Proof. In this case, we can run a two-step estimation.

Step 1:

Because $a(\boldsymbol{\theta})$ is endogenized by $\boldsymbol{\theta}$, for $(a, \boldsymbol{\theta})$, we have

$$
\int [\frac{\partial a}{\partial \theta_j} f_a(x, a(\boldsymbol{\theta}), \boldsymbol{\theta}) + f_{\theta_j}(x, a(\boldsymbol{\theta}), \boldsymbol{\theta})]dx = 0 \text{ for } j = 1, ..., t
$$

Note that a is a parameter, so $\int f_a(x, a(\boldsymbol{\theta}), \boldsymbol{\theta})dx = 0$ still holds. Therefore, together with $\int f_{\theta_j}(x, a(\boldsymbol{\theta}), \boldsymbol{\theta})dx = 0$, there are $t + 1$ moment conditions to identify $(a, \boldsymbol{\theta})$ under A6. Let $\mathbf{T} = (a, \boldsymbol{\theta})'$, then the parameter $\mathbf{T}$ can be estimated through MLE or GMM. The standard result of MLE follows:

$$
\sqrt{n}(\hat{\mathbf{T}} - \mathbf{T}) \to^d \mathcal{N}(0, \mathbf{Z}^{-1})
$$

where $\mathbf{Z} = -\mathbb{E}\frac{\partial^2 \ln f(x,\mathbf{T})}{\partial \mathbf{T} \partial \mathbf{T}'}$ is information quantity.

Step 2:

We can justify the applicability of the uniform convergence, i.e., $\sup_\gamma \left|\hat{\Lambda}(\gamma, \hat{\mathbf{T}}) - \Lambda(\gamma, \mathbf{T})\right| \to^p 0$ by the same technique used in theorem 3. With the additional assumptions A7, the asymptotic normality of $\hat{\gamma}$ can be derived as follows. By the first order condition

$$0 = \hat{\Lambda}_\gamma(\hat{\gamma}, \hat{\mathbf{T}}) = \hat{\Lambda}_\gamma(\gamma, \mathbf{T}) + \hat{\Lambda}_{\gamma\gamma'}(\bar{\gamma}, \bar{\mathbf{T}})(\hat{\gamma} - \gamma) + \hat{\Lambda}_{\gamma\mathbf{T}'}(\bar{\gamma}, \bar{\mathbf{T}})(\hat{\mathbf{T}} - \mathbf{T})$$

we have

$$\sqrt{n}(\hat{\gamma} - \gamma) = -[\hat{\Lambda}_{\gamma\gamma'}(\bar{\gamma}, \bar{\mathbf{T}})]^{-1}[\sqrt{n}\hat{\Lambda}_\gamma(\gamma, \mathbf{T}) + \hat{\Lambda}_{\gamma\mathbf{T}'}(\bar{\gamma}, \bar{\mathbf{T}})\sqrt{n}(\hat{\mathbf{T}} - \mathbf{T})]$$

Note that $\hat{\Lambda}_{\gamma\mathbf{T}'}(\bar{\gamma}, \bar{\mathbf{T}}) \to^p \Lambda_{\gamma\mathbf{T}'}$, $\hat{\Lambda}_\gamma(\gamma, \mathbf{T}) \to^p \Lambda_\gamma$, $\hat{\Lambda}_{\gamma\gamma'}(\bar{\gamma}, \bar{\mathbf{T}}) \to^p \Lambda_{\gamma\gamma'}$, therefore, the asymptotic distribution of $(\hat{\gamma} - \gamma)$ is:

$$\sqrt{n}(\hat{\gamma} - \gamma_0) \to^d \mathcal{N}(0, \Lambda_{\gamma\gamma'}^{-1}[\Omega + \Lambda_{\gamma\mathbf{T}'}\mathbf{Z}^{-1}\Lambda_{\gamma'\mathbf{T}}]\Lambda_{\gamma\gamma'}^{-1}).$$

For convenience, write $\mathbf{\Gamma} = \Lambda_{\gamma\gamma'}^{-1}[\Omega + \Lambda_{\gamma\mathbf{T}'}\mathbf{Z}^{-1}\Lambda_{\gamma'\mathbf{T}}]\Lambda_{\gamma\gamma'}^{-1}$. Particulary, under the null hypothesis, $\Omega = 0$, and $nAVar(\hat{\Lambda}_{\gamma\mathbf{T}'}(\bar{\gamma}, \bar{\mathbf{T}})) = Var(\varepsilon_{\gamma\mathbf{T}'})$, we have

$$n(\hat{\gamma} - \gamma) \to^d \mathcal{N}(0, \mathbb{E}(\varepsilon_{\gamma\mathbf{T}'} - \mathbb{E}\varepsilon_{\gamma\mathbf{T}'}))Z^{-1}\mathbb{E}(\varepsilon_{\gamma\mathbf{T}'} - \mathbb{E}\varepsilon_{\gamma\mathbf{T}'}))]$$

which converges at rate $\frac{1}{n}$ rather than $\frac{1}{\sqrt{n}}$. Given this effect, according the proof in A5, $\sqrt{n}(\hat{\gamma} - \gamma)$ converge faster than $\sqrt{n}(\hat{\mathbf{T}} - \mathbf{T})$, so we can neglect them in the the asymptotic distribution of $\hat{\rho}(\hat{\gamma}, \hat{\mathbf{T}})$.

The asymptotic distribution of $\hat{M}$ and $\hat{K}$ can also be derived similarly. For estimator $\hat{A} = \hat{\mathbb{E}}u(w_i, \hat{\gamma})g(x_i, \hat{\mathbf{T}})$, its asymptotic distribution is

$$\sqrt{n}(\hat{A} - A) \to^d \mathcal{N}(0, Var(ug) + (\mathbb{E}ug_{T'})\,\mathbf{Z}^{-1}\,(\mathbb{E}ug_T) + \left(\mathbb{E}u_{\gamma'}g\right)\mathbf{\Gamma}\,(\mathbb{E}u_\gamma g) + 2\left(\mathbb{E}u_{\gamma'}g\right)\sigma_{\gamma T}\mathbb{E}ug_T)$$

where $g = l_a$ for $\hat{M}$ and $g = l_{aa} + l_a^2$ for $\hat{K}$. Q.E.D. ∎

1.8.7 Proof of lemma 1

Proof. Note that if γ is maximizer of value function $V(\gamma)$,

$$\begin{aligned} V(\gamma) &= \int v(x - s(x,\gamma))f(x,a^*)dx \\ &+\lambda[\int [u(s(x,\gamma),\gamma) - c(a^*)]f(x,a^*)dx - \underline{U}] \\ &+\mu[\int u(s(x,\gamma),\gamma)f_a(x,a^*)dx - c'(a^*)], \end{aligned}$$

and by the envelop theorem,

$$\begin{aligned} V_\gamma^* &= \frac{dV^*}{d\gamma} = \frac{\partial V}{\partial a}\frac{\partial a^*}{\partial \gamma} + \frac{\delta V}{\delta s}\frac{\partial s}{\partial \gamma} + \lambda \int \frac{\partial}{\partial \gamma} u(s(x,\gamma),\gamma)f(x,a^*(\gamma))dx \\ &+\mu \int \frac{\partial}{\partial \gamma} u(s(x,\gamma),\gamma)f_a(x,a^*(\gamma))dx + \frac{\partial \lambda}{\partial \gamma}[\int [u(s(x,\gamma),\gamma) - c(a^*)]f(x,a^*)dx - \underline{U}] \\ &+\frac{\partial \mu}{\partial \gamma}[\int u(s(x,\gamma),\gamma)f_a(x,a^*)dx - c'(a^*)] \\ &= \lambda \int \frac{\partial}{\partial \gamma} u(s(x,\gamma),\gamma)f(x,a^*)dx + \mu \int \frac{\partial}{\partial \gamma} u(s(x,\gamma),\gamma)f_a(x,a^*(\gamma))dx \\ &= 0. \end{aligned}$$

Here we utilize $\frac{\partial \underline{U}}{\partial \gamma} = 0$. The reason that $\frac{\partial \underline{U}}{\partial \gamma} = 0$ can have different interpretations. The obvious one is that the reservation utility is kept the same cross different types of agent. The second one is that the principal assures that every agent got the same utility since IR constraint will be binding for any type. $\underline{U}$ does not vary means that any type of agent will have the same utility, which could be regarded as market equilibrium (otherwise, some agent will switch). In terms of statistics, $\underline{U}$ need not to be exactly the same, as long as the error happens to $\underline{U}$ (or the agent's equilibrium utility) is mean independent of the type, the above result still holds.

Note that $\mu f_a(x,a^*) = h(s^*,\gamma) - \lambda$ when $s^*(x) \geq \underline{s}$, and $s^*(x)$ is uniformly bounded by $\underline{s}$ from below, therefore we obtain the condition $V_\gamma^* = \frac{\partial u(\underline{s},\gamma)}{\partial \gamma} \int_{w<\underline{s}} (\lambda + \mu l_a(x,a^*))f(x,a)dx + \int_{w \geq \underline{s}} \frac{\frac{\partial}{\partial \gamma} u(s^*(x),\gamma)}{\frac{\partial}{\partial w} u(s^*(x),\gamma)} f(x,a^*)dx = 0$. Q.E.D. ∎

1.8.8 Proof of Proposition 9

Proof. (i) if $\varpi(x,\mathbf{T})$ does not contain a, given the score function $l_a(x,a) = \frac{1}{\eta(\mathbf{T})}(\varpi(x,\mathbf{T}) - \mathbb{E}\varpi(x_i,\mathbf{T}))$, thus

$$l_{aa}(x,a) = -\frac{\frac{\partial}{\partial a}\eta(\mathbf{T})}{\eta(\mathbf{T})^2}(\varpi(x,\mathbf{T}) - \mathbb{E}\varpi(x_i,\mathbf{T})) - \frac{1}{\eta(\mathbf{T})^2}\mathbb{E}\varpi(x_i,\mathbf{T})(\varpi(x_i,\mathbf{T}) - \mathbb{E}\varpi(x_i,\mathbf{T}))$$

Note that $\mathbb{E}l_a^2 = -\mathbb{E}l_{aa}$, thus

$$\begin{aligned} K &= \frac{\frac{\partial}{\partial a}\eta(\mathbf{T})}{\eta(\mathbf{T})^2}\mathbb{E}[u(w_i)(\varpi(x_i,\mathbf{T}) - \mathbb{E}\varpi(x_i,\mathbf{T}))] - \frac{1}{\eta(\mathbf{T})^2}\mathbb{E}[u(w_i)(\varpi(x_i,\mathbf{T}) - \mathbb{E}\varpi(x_i,\mathbf{T}))^2] \\ &\quad + \frac{1}{\eta(\mathbf{T})^2}\mathbb{E}\varpi(x_i,\mathbf{T})(\varpi(x_i,\mathbf{T}) - \mathbb{E}\varpi(x_i,\mathbf{T}))\mathbb{E}u(w_i) \end{aligned}$$

Meanwhile, note that

$$\begin{aligned} \frac{\partial^2 \mathbb{E}x_i}{\partial a^2} &= -\frac{\frac{\partial}{\partial a}\eta(\mathbf{T})}{\eta(\mathbf{T})^2}\int x(\varpi(x,\mathbf{T}) - \mathbb{E}\varpi(x_i,\mathbf{T}))f(x,a)dx + \frac{1}{\eta(\mathbf{T})^2}\mathbb{E}[x_i(\varpi(x_i,\mathbf{T}) - \mathbb{E}\varpi(x_i,\mathbf{T}))^2] \\ &\quad - \frac{1}{\eta(\mathbf{T})^2}\mathbb{E}\varpi(x_i,\mathbf{T})(\varpi(x_i,\mathbf{T}) - \mathbb{E}\varpi(x_i,\mathbf{T}))\mathbb{E}x_i \\ &< 0. \end{aligned}$$

And note that $\frac{\partial}{\partial a}\eta(\mathbf{T}) > 0$ iff $Cov(x_i, \varpi(x_i,\mathbf{T})) > 0$. When $Cov(x_i, \varpi(x_i,\mathbf{T})) > 0$ is the case, we have $\frac{\partial}{\partial a}\eta(\mathbf{T}) > 0$ and $\mathbb{E}[u(w_i)(\varpi(x_i,\mathbf{T}) - \mathbb{E}\varpi(x_i,\mathbf{T}))] > 0$, therefore, the sufficient condition for $K > 0$ is,

$$\frac{Cov(x_i, (\varpi(x_i,\mathbf{T}) - \mathbb{E}\varpi(x_i,\mathbf{T}))^2)}{Cov(x, \varpi(x_i,\mathbf{T}))}Cov(u(w_i), \varpi(x_i,\mathbf{T})) > Cov(u(w_i), (\varpi(x_i,\mathbf{T}) - \mathbb{E}\varpi(x_i,\mathbf{T}))^2). \tag{1.31}$$

When $Cov(x_i, \varpi(x_i,\mathbf{T})) < 0$, then $\frac{\partial}{\partial a}\eta(\mathbf{T}) < 0$ and $\mathbb{E}[u(w_i)(\varpi(x_i,\mathbf{T}) - \mathbb{E}\varpi(x_i,\mathbf{T}))] < 0$, the sufficient condition will be the same as inequality (23).

(ii) When $\varpi(x,\mathbf{T})$ contains a as an explicit parameter, thus, we normalize $\eta(\mathbf{T})$ to a constant independence of a, therefore,

$$l_{aa}(x,a) = \frac{1}{\eta(\mathbf{T})}\left(\frac{\partial}{\partial a}\varpi(x,\mathbf{T}) - \mathbb{E}\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T})\right) - \frac{1}{\eta(\mathbf{T})^2}\mathbb{E}\varpi(x_i,\mathbf{T})(\varpi(x_i,\mathbf{T}) - \mathbb{E}\varpi(x_i,\mathbf{T})).$$

Note that

$$\begin{aligned} K &= -\mathbb{E}u(w_i)(l_{aa}(x_i,a)+l_a^2(x_i,a)) \\ &= -\frac{1}{\eta(\mathbf{T})}\mathbb{E}[u(w_i)(\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T})-\mathbb{E}\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T}))]-\frac{1}{\eta(\mathbf{T})^2}Cov(u(w_i),(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))^2) \end{aligned}$$

Meanwhile, note that,

$$\begin{aligned} \frac{\partial^2\mathbb{E}x_i}{\partial a^2} &= \frac{1}{\eta(\mathbf{T})}\mathbb{E}[x_i(\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T})-\mathbb{E}\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T}))]+\frac{1}{\eta(\mathbf{T})^2}Cov(x_i,(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))^2) \\ &< 0 \end{aligned}$$

Therefore, when $\mathbb{E}[x_i(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))]>0$, we have $\frac{1}{\eta(\mathbf{T})}>0$, and

$$\frac{1}{\eta(\mathbf{T})}Cov(x_i,(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))^2)<-\mathbb{E}[x_i(\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T})-\mathbb{E}\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T}))]$$

Moreover, when the sign of $Cov(x_i,(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))^2)$ agrees with that of $Cov(u(w_i),(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))^2)$, then, to test $K>0$, it suffices to test,

$$\begin{aligned} & -\mathbb{E}[u(w_i)(\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T})-\mathbb{E}\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T}))]-\frac{1}{\eta(\mathbf{T})}Cov(u(w_i),(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))^2) \\ > & -Cov(u(w_i),\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T}))+Cov(x_i,\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T}))\frac{Cov(u(w_i),(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))^2)}{Cov(x_i,(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))^2)} \\ > & \ 0 \end{aligned}$$

If $\mathbb{E}[x_i(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))]<0$, we have $\frac{1}{\eta(\mathbf{T})}<0$, and

$$\frac{1}{\eta(\mathbf{T})}Cov(x_i,(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))^2)>-\mathbb{E}[x_i(\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T})-\mathbb{E}\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T}))]$$

To test $K>0$, it suffices to test,

$$\begin{aligned} & -\mathbb{E}[u(w_i)(\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T})-\mathbb{E}\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T}))]-\frac{1}{\eta(\mathbf{T})}Cov(u(w_i),(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))^2) \\ < & -Cov(u(w_i),\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T}))+Cov(x_i,\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T}))\frac{Cov(u(w_i),(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))^2)}{Cov(x_i,(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))^2)} \\ < & \ 0 \end{aligned}$$

Put these two cases together, we need to test that the sign of $\mathbb{E}[x_i(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))]$ agrees with that of $Cov(x_i,\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T}))\frac{Cov(u(w_i),(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))^2)}{Cov(x_i,(\varpi(x_i,\mathbf{T})-\mathbb{E}\varpi(x_i,\mathbf{T}))^2)} - Cov(u(w_i),\frac{\partial}{\partial a}\varpi(x_i,\mathbf{T}))$. Use the sample analogue of the objects, we can test $K>0$. Q.E.D. ■

1.8.9 Proof of Theorem 4

Proof. (i) For convenience, we use $K_2=\int\varphi^2K(\varphi)d\varphi$ as our notation. We have,

$$\begin{aligned}\mathbb{E}\hat{l}_a-l_a &= \frac{1}{f(\varepsilon)}\frac{1}{6}b^2f'''(\varepsilon)\int\varphi^3K'(\varphi)d\varphi+\frac{f'(\varepsilon)}{f(\varepsilon)^2}\frac{1}{2}f''(\varepsilon)b^2\int\varphi^2K(\varphi)d\varphi \\ &= \frac{1}{2}b^2A(\varepsilon)K_2\end{aligned}$$

where $A(\varepsilon)=\frac{[f'(\varepsilon)f''(\varepsilon)-f(\varepsilon)f'''(\varepsilon)]}{f(\varepsilon)^2}$. And we have

$$\begin{aligned}MSE(\hat{l}_a) &= Var(\hat{l}_a)+(\mathbb{E}\hat{l}_a-l_a)^2 \\ &= \frac{1}{nb^3}\frac{1}{f(\varepsilon)}\int[K'(\varphi)]^2d\varphi+b^4\left(\frac{1}{2}A(\varepsilon)\int\varphi^2K(\varphi)d\varphi\right)^2\end{aligned}$$

Choose $b\propto n^{-\frac{1}{7}-\alpha}$ $(\alpha>0)$, thus $\lim_{n\to\infty}(\mathbb{E}\hat{l}_a-l_a)^2\to 0$, and $\lim_{n\to\infty}Var(\hat{l}_a)=0$. So we have

$$\sqrt{nb^3}(\hat{l}_a-l_a)\to\mathcal{N}(0,\frac{1}{f(\varepsilon)}\int[K'(\varphi)]^2d\varphi)$$

We can proceed to estimate $h(w)$ with monotonicity constraint.

We use Delette et al (2006) procedure to estimate a monotone function $h(w_i)$. We have regression

$$\hat{\zeta}_i=h(w_i)+\sigma(x_i)\epsilon_i$$

where ϵ_i comes from the first stage estimation error. The first step is estimate $h(w_i)$ without monotonicity constraint, such as

$$\hat{h}_N(w)=\frac{\frac{1}{nb_h}\sum_{i=1}^nK_r(\frac{w_i-w}{b_h})\hat{\zeta}_i}{\frac{1}{nb_h}\sum_{i=1}^nK_r(\frac{w_i-w}{b_h})}.$$

We have

$$
\begin{aligned}
Var(\hat{h}_N) &= \frac{1}{(\mathbb{E}\hat{N}_2)^2}Var(\hat{N}_1) + \frac{(\mathbb{E}\hat{N}_1)^2}{(\mathbb{E}\hat{N}_2)^4}Var(\hat{N}_2) - 2\frac{\mathbb{E}\hat{N}_1}{(\mathbb{E}\hat{N}_2)^3}Cov(\hat{N}_1, \hat{N}_2) \\
&= \frac{1}{nb_h}\frac{1}{\phi(w)}\frac{\int [K'(\varphi)]^2 d\varphi}{nb^3 f(\varepsilon)}\int K_r(\varphi)^2 d\varphi
\end{aligned}
$$

where we utilize $\int \varphi K_r(\varphi)^2 d\varphi = 0$ and $Var(\hat{N}_1) = \mathbb{E}Var_x(\hat{N}_1) + Var(\mathbb{E}_x(\hat{N}_1))$ and $\phi(w) = \frac{f(\varepsilon)}{s'(x)}$ is the density of w under the null hypothesis.

To minimize the AIMSE of $\hat{h}$, we can choose $b_h \propto (n^2b^3)^{-\frac{1}{5}} = o(n^{-\frac{1}{3}})$, then the biasness term asymptotically disappears, so we have,

$$
n\sqrt{b^3 b_h}(\hat{h}(w) - h(w)) \to^d \mathcal{N}\left(0, \frac{1}{\phi(w)f(\varepsilon)}\int [K'(\varphi)]^2 d\varphi \int K_r(\varphi)^2 d\varphi\right).
$$

The convergence rate is faster than routine one.

The second step is to derive the asymptotic distribution of $\hat{h}(w)$ with monotonicity constraint. Our results is based on Dette et al (2006). Their approach is to inverse the estimated density of $\hat{h}_N(w)$. The second step estimator of $h(w)$ is

$$
\hat{h}^{-1}(t) = \frac{1}{Nb_d}\sum_{i=1}^{N}\int_{-\infty}^{t} K_d\left(\frac{\hat{h}_N(\frac{i}{N}) - \tau}{b_d}\right) d\tau.
$$

and the third step is to reflect $\hat{h}^{-1}(t)$ on the axis $y = x$, we have the asymptotic distribution of $\hat{h}_I(t)$ as following:

$$
\begin{aligned}
& n\sqrt{b^3 b_h}\left(\hat{h}(w) - h(w) - \frac{1}{2}b_h^2\frac{h''(w)\phi(w) + 2h'(w)\phi'(w)}{\phi(w)}\int_{-1}^{1}\varphi^2 K_r(\varphi)d\varphi\right) \\
\to & \ {}^d\mathcal{N}\left(0, \frac{1}{\phi(w)f(\varepsilon)}\int [K'(\varphi)]^2 d\varphi \int K_r(\varphi)^2 d\varphi\right),
\end{aligned}
$$

where $\lim_{n\to\infty}\frac{b_h}{b_d} = \infty$. $b_h^2 = o(n^{-\frac{2}{3}})$ as we set before.

We turn to derive the asymptotic distribution of $\hat{\rho}$. First of all, the asymptotic distribution of $\hat{Z} = \hat{\mathbb{E}}\hat{l}_a^2$ can be determined as follows,

$$\begin{aligned}
& \lim_{n\to\infty} n\mathbb{E}(\hat{Z} - \mathbb{E}\hat{Z})^2 \\
= & \lim_{n\to\infty} n\mathbb{E}(\hat{Z} - Z)^2 - \lim_{n\to\infty} n\mathbb{E}(Z - \mathbb{E}\hat{Z})^2 \\
= & \lim_{n\to\infty} n\mathbb{E}[\hat{\mathbb{E}}(\hat{l}_a^2(x_i) - l_a^2(x_i)) + \hat{\mathbb{E}}l_a^2(x_i) - Z - (\hat{\mathbb{E}}\hat{l}_a(x_i))^2]^2 - \lim_{n\to\infty} n\mathbb{E}(Z - \mathbb{E}\hat{l}_a^2(x_i) + \mathbb{E}(\hat{\mathbb{E}}\hat{l}_a(x_i))^2)^2 \\
= & \lim_{n\to\infty} n\mathbb{E}[\hat{\mathbb{E}}(\hat{l}_a^2(x_i) - l_a^2(x_i)) + \hat{\mathbb{E}}l_a^2(x_i) - Z]^2 - 2\lim_{n\to\infty} n\mathbb{E}(\hat{\mathbb{E}}\hat{l}_a(x_i))^2[\hat{\mathbb{E}}(\hat{l}_a^2(x_i) - l_a^2(x_i)) + \hat{\mathbb{E}}l_a^2(x_i) - Z] \\
& + \lim_{n\to\infty} n\mathbb{E}(\hat{\mathbb{E}}\hat{l}_a(x_i))^4 - \lim_{n\to\infty} n\mathbb{E}[\mathbb{E}\hat{l}_a^2(x_i) - Z]^2 + 2\lim_{n\to\infty} n\mathbb{E}(\hat{\mathbb{E}}\hat{l}_a(x_i))^2\mathbb{E}[\mathbb{E}\hat{l}_a^2(x_i) - Z] - \lim_{n\to\infty} n(\mathbb{E}(\hat{\mathbb{E}}\hat{l}_a(x_i)^2
\end{aligned}$$

The first term is equal to,

$$\begin{aligned}
& \mathbb{E}[\hat{\mathbb{E}}(\hat{l}_a^2(x_i) - l_a^2(x_i))]^2 + 2\mathbb{E}[\hat{\mathbb{E}}(\hat{l}_a^2(x_i) - l_a^2(x_i))][\hat{\mathbb{E}}l_a^2(x_i) - Z] + \mathbb{E}(\hat{\mathbb{E}}l_a^2(x_i) - Z)^2 \\
= & 4\mathbb{E}l_a^2(x_i)(\hat{l}_a(x_i) - l_a(x_i))^2 + (n-1)4\mathbb{E}l_a(x_i)(\hat{l}_a(x_i) - l_a(x_i))l_a(x_j)(\hat{l}_a(x_j) - l_a(x_j)) \\
& +4\mathbb{E}[l_a(x_i)(\hat{l}_a(x_i) - l_a(x_i))(l_a^2(x_i) - Z)] + 4(n-1)\mathbb{E}[l_a(x_i)(\hat{l}_a(x_i) - l_a(x_i))(l_a^2(x_j) - Z)] \\
& +Var(l_a^2(x_i)) \\
= & 4\left(\frac{1}{nb^3}\left(\int [K'(\varphi)]^2 d\varphi\right)\mathbb{E}\frac{l_a^2(x_i)}{f(\varepsilon_i)} + \frac{1}{4}b^4K_2^2\mathbb{E}\left[A_i^2 l_a^2(x_i)\right]\right) + 4(n-1)\left(\frac{b^2}{2}K_2\right)^2[\mathbb{E}A_i l_a(x_i)]^2 \\
& +4\left(\frac{b^2}{2}K_2^2\right)\mathbb{E}A_i l_a(x_i)(l_a^2(x_i) - Z)] + 4(n-1)\left(\frac{b^2}{2}K_2^2\right)\mathbb{E}A_i l_a(x_i)\mathbb{E}(l_a^2(x_j) - Z)] + Var(l_a^2(x_i))
\end{aligned}$$

The sum of the second term and the fouth term is:

$$\begin{aligned}
& \lim_{n\to\infty} n\mathbb{E}(\hat{\mathbb{E}}\hat{l}_a(x_i))^2[\hat{\mathbb{E}}(\hat{l}_a^2(x_i) - l_a^2(x_i)) + \hat{\mathbb{E}}l_a^2(x_i) - Z] - \lim_{n\to\infty} n\mathbb{E}(\hat{\mathbb{E}}\hat{l}_a(x_i))^2\mathbb{E}[\mathbb{E}\hat{l}_a^2(x_i) - Z] \\
= & \frac{2(n-1)}{n}(\mathbb{E}\hat{l}_a(x_i))[\mathbb{E}\hat{l}_a(x_i)[(\hat{l}_a^2(x_i) - l_a^2(x_i)) + \hat{\mathbb{E}}l_a^2(x_i) - Z]] \\
& +\frac{(n-1)(n-2)}{n}(\mathbb{E}\hat{l}_a(x_i))^2[\mathbb{E}(\hat{l}_a^2(x_i) - Z] - n\mathbb{E}(\hat{\mathbb{E}}\hat{l}_a(x_i))^2\mathbb{E}[\hat{l}_a^2(x_i) - Z] \\
= & 2(\mathbb{E}\hat{l}_a(x_i))[\mathbb{E}\hat{l}_a(x_i)[(\hat{l}_a^2(x_i) - l_a^2(x_i)) + \hat{\mathbb{E}}l_a^2(x_i) - Z]] - \frac{2(n-1)}{n}(\mathbb{E}\hat{l}_a(x_i))^2\mathbb{E}[\hat{l}_a^2(x_i) - Z] \\
& -\mathbb{E}\hat{l}_a^2(x_i)\mathbb{E}[\hat{l}_a^2(x_i) - Z] + o(\frac{1}{n}) \\
= & 4\left(\frac{1}{2}K_2b^2\mathbb{E}A(\varepsilon_i)\right)\frac{1}{2}K_2b^2\mathbb{E}A(\varepsilon_i)l_a^2(x_i) - 4\left(\frac{1}{2}K_2b^2\right)^2\mathbb{E}A(\varepsilon_i)\mathbb{E}A(\varepsilon_i)l_a(x_i) \\
& -2Z\left(\frac{1}{2}K_2b^2\right)\mathbb{E}A(\varepsilon_i)l_a(x_i) + o(b^4)
\end{aligned}$$

The sum of the third term and the fouth term is equal to,

$$\lim_{n\to\infty} n\mathbb{E}(\hat{\mathbb{E}}\hat{l}_a(x_i))^4 - \lim_{n\to\infty} n(\mathbb{E}(\hat{\mathbb{E}}\hat{l}_a(x_i)^2)^2 = 4\mathbb{E}\hat{l}_a^2(x_i)(\mathbb{E}\hat{l}_a(x_i))^2 + o(b^4) = 4Z(\frac{1}{2}K_2 b^2 \mathbb{E}A(\varepsilon_i))^2.$$

Putting all terms together, we have,

$$\begin{aligned}
&\lim_{n\to\infty} n\mathbb{E}(\hat{Z} - \mathbb{E}\hat{Z})^2 \\
= \;& 4\left(\frac{1}{nb^3}\left(\int [K'(\varphi)]^2 d\varphi\right)\mathbb{E}\frac{l_a^2(x_i)}{f(\varepsilon_i)} + \frac{1}{4}b^4K_2^2\mathbb{E}\left[A_i^2 l_a^2(x_i)\right]\right) + 4(n-1)\left(\frac{b^2}{2}K_2\right)^2 [\mathbb{E}A_i l_a(x_i)]^2 \\
& +4\left(\frac{b^2}{2}K_2^2\right)\mathbb{E}A_i l_a(x_i)(l_a^2(x_i) - Z)] + 4(n-1)\left(\frac{b^2}{2}K_2^2\right)\mathbb{E}A_i l_a(x_i)\mathbb{E}(l_a^2(x_j) - Z)] + Var(l_a^2(x_i)) \\
& -2\left[\begin{array}{c} 4\left(\frac{1}{2}K_2b^2\mathbb{E}A(\varepsilon_i)\right)\frac{1}{2}K_2b^2\mathbb{E}A(\varepsilon_i)l_a^2(x_i) - 4\left(\frac{1}{2}K_2b^2\right)^2\mathbb{E}A(\varepsilon_i)\mathbb{E}A(\varepsilon_i)l_a(x_i) \\ -2Z\left(\frac{b^2}{2}K_2\right)\mathbb{E}A(\varepsilon_i)l_a(x_i) \end{array}\right] \\
= \;& 4\left(\frac{1}{nb^3}\left(\int [K'(\varphi)]^2 d\varphi\right)\mathbb{E}\frac{l_a^2(x_i)}{f(\varepsilon_i)} + \frac{1}{4}b^4K_2^2\mathbb{E}\left[A_i^2 l_a^2(x_i)\right]\right) + 4(n-1)\left(\frac{b^2}{2}K_2\right)^2 [\mathbb{E}A_i l_a(x_i)]^2 \\
& +4\left(\frac{b^2}{2}K_2^2\right)\mathbb{E}A_i l_a(x_i) l_a^2(x_i) + Var(l_a^2(x_i)) - 8\left(\frac{1}{2}K_2b^2\right)^2\mathbb{E}A(\varepsilon_i)\left[\mathbb{E}A(\varepsilon_i)l_a^2(x_i) - \mathbb{E}A(\varepsilon_i)l_a(x_i)\right]
\end{aligned}$$

As $nb^3 \to \infty$, $b \to 0$, we have

$$\sqrt{n}(\hat{Z} - Z) \to^d (\frac{1}{2}b^2K_2\mathbb{E}\left[A(\varepsilon_i)l_a(x_i)\right], Var(l_a^2(x_i))).$$

Note that $\hat{Z}$ converges to Z at routine rate. This is the case as $\hat{J}$ and $\hat{Q}$ since $\hat{h}(w)$ converge to $h(w)$ faster than $\hat{Z}$'s. Given this fact, we choose $\hat{J} = \hat{\mathbb{E}}(h - \bar{h})(\hat{l}_a - \hat{\mathbb{E}}\hat{l}_a)$, $\hat{Z} = \hat{\mathbb{E}}(\hat{l}_a - \hat{\mathbb{E}}\hat{l}_a)^2$ to estimate $\hat{\rho}$, and we have,

$$\begin{aligned}
&\frac{\hat{J}(\hat{l}_a)}{\sqrt{\hat{Q}\hat{Z}(\hat{l}_a)}} - \rho \\
= \;& \frac{1}{Z}\left(-\frac{1}{2}\hat{\mathbb{E}}(\hat{l}_a - \hat{\mathbb{E}}\hat{l}_a)(\hat{l}_a - l_a) + \frac{1}{2}\hat{\mathbb{E}}(\hat{l}_a - l_a)(l_a - \hat{\mathbb{E}}l_a)\right) + o(\frac{1}{nb^3\sqrt{nb_h}}) + o(\frac{1}{n^2b^3b_h}) \\
& +\frac{1}{2}\frac{1}{\sqrt{n}}o(\hat{\mathbb{E}}(\hat{l}_a - \hat{\mathbb{E}}\hat{l}_a)(\hat{l}_a - l_a)) + \frac{1}{8}\frac{1}{n}[O(\frac{1}{nb^3}) + O(b^4)] \\
= \;& -\frac{1}{2Z}\hat{\mathbb{E}}[(\hat{l}_a - l_a) - \hat{\mathbb{E}}(\hat{l}_a - l_a)]^2 + o(\frac{1}{\sqrt{n}})
\end{aligned}$$

This is due to the second order items is in the order of $\frac{1}{2}\frac{1}{\sqrt{n}}o(\frac{1}{\sqrt{nb^3}}) + o(\frac{1}{n})$. Note that

$$
\begin{aligned}
& \mathbb{E}[\hat{\mathbb{E}}[\hat{l}_a(x_i) - l_a(x_i) - \hat{\mathbb{E}}(\hat{l}_a(x_i) - l_a(x_i)]^2] \\
= \; & \mathbb{E}[\hat{l}_a(x_i) - l_a(x_i) - \hat{\mathbb{E}}(\hat{l}_a(x_i) - l_a(x_i)]^2 \\
= \; & \frac{n-1}{n}\left(\mathbb{E}(\hat{l}_a(x_i) - l_a(x_i))^2 - [\mathbb{E}(\hat{l}_a(x_i) - l_a(x_i))]^2\right) \\
= \; & \frac{1}{nb^3}\left(\int [K'(\varphi)]^2 d\varphi\right)\mathbb{E}\frac{1}{f(\varepsilon_i)} + \frac{1}{4}b^4K_2^2Var(A(\varepsilon_i))
\end{aligned}
$$

and

$$
\begin{aligned}
& \lim_{n\to\infty}\{\mathbb{E}[\hat{l}_a(x_i) - l_a(x_i) - \hat{\mathbb{E}}(\hat{l}_a(x_i) - l_a(x_i)]^4 - \left(\mathbb{E}[\hat{l}_a(x_i) - l_a(x_i) - \hat{\mathbb{E}}(\hat{l}_a(x_i) - l_a(x_i)]^2\right)^2\} \\
= \; & \lim_{n\to\infty}\{\mathbb{E}[\hat{l}_a(x_i) - l_a(x_i)]^4 - 4\mathbb{E}[\hat{l}_a(x_i) - l_a(x_i)]^3\mathbb{E}[\hat{l}_a(x_i) - l_a(x_i)] \\
& +8\mathbb{E}[\hat{l}_a(x_i) - l_a(x_i)]^2[\mathbb{E}(\hat{l}_a(x_i) - l_a(x_i))]^2 - [\mathbb{E}(\hat{l}_a(x_i) - l_a(x_i))^2]^2 - 4[\mathbb{E}(\hat{l}_a(x_i) - l_a(x_i))]^4\} \\
= \; & \lim_{n\to\infty}\mathbb{E}[\hat{l}_a(x_i) - l_a(x_i)]^4 + O(\mathbb{E}[\hat{l}_a(x_i) - l_a(x_i)]^3\mathbb{E}[\hat{l}_a(x_i) - l_a(x_i)) \\
= \; & \lim_{n\to\infty}\mathbb{E}[\frac{1}{f(\varepsilon_i)}[\hat{f}'(x_i) - f'(x_i)]]^4 + o(\mathbb{E}[\frac{1}{f(\varepsilon_i)}[\hat{f}'(x_i) - f'(x_i)]]^4) \\
= \; & \frac{24}{nb^3}\int K'(\varphi)^2 d\varphi\mathbb{E}\frac{l_a^2(x_i)}{f(\varepsilon_i)}
\end{aligned}
$$

Finally, we have,

$$
\begin{aligned}
& \hat{\rho} - 1 + \frac{1}{2Z}\left(\frac{1}{nb^3}\mathbb{E}\frac{1}{f(\varepsilon_i)}\int[K'(\varphi)]^2 d\varphi + \frac{1}{4}b^4K_2^2Var(A(\varepsilon_i))\right) \\
\to \; & {}^d\mathcal{N}\left(0, \frac{1}{4nZ^2}\left(\frac{24}{nb^3}\mathbb{E}\frac{l_a^2(x_i)}{f(\varepsilon_i)}\int[K'(\varphi)]^2 d\varphi] + O(\frac{1}{nb}) + o(b^4)\right)\right).
\end{aligned}
$$

■

1.8.10 Proof of theorem 5

Proof. Given $\hat{h}(\hat{s}) = \hat{\lambda} + \hat{\mu} l_a$, under the null hypothesis, $\hat{s}(x)$ is the optimal contract under utility $\hat{u}(.)$, therefore,

$$\begin{aligned} 0 &= \int \hat{u}(\hat{r}(\hat{q})) l_a(x) f(x,a) dx - \hat{c}'(a) \\ \underline{U} &= \int \hat{u}(\hat{r}(\hat{q})) f(x,a) dx - \hat{c}(a) \end{aligned}$$

if $\hat{u}(.)$ is the true primitive utility, then we done, $\hat{h}$ is identified by the asymptotic distribution of estimation of $\hat{h}$.

If $\hat{h}$ is not from the true utility $u(w)$, instead under the true utility $u(w)$, the optimal contract should be

$$\hat{h}(\hat{s}) = \lambda + \mu l_a, \text{ with } \lambda > 0 \text{ and } \mu > 0 \text{ as two unknown parameters,}$$

since true marginal incentive cost $h(w)$ is an affine transformation of estimated marginal utility $\hat{h}(w)$. Therefore, there must exist another contract $s(x) = r(q)$, and IC and IR constraints binding accordingly. Thus the following condition satisfied (using the notation as proof of theorem 1),

$$\begin{aligned} \mathbb{E}s(x) &\le \mathbb{E}\hat{s}(x); \\ \int u(r(q)) l_a(x) f(x,a) dx - c'(a) &= 0 = \int \hat{u}(\hat{r}(\hat{q})) l_a(x) f(x,a) dx - \hat{c}'(a); \\ \int u(r(q)) f(x,a) dx - c(a) &= \underline{U} = \int \hat{u}(\hat{r}(\hat{q})) f(x,a) dx - \hat{c}(a). \end{aligned}$$

At the same time, since the true utility is $u(.)$, then we also observe the following fact to hold:

$$\begin{aligned} \int u(\hat{r}(\hat{q})) l_a(x) f(x,a) dx - c'(a) &= 0 = \int \hat{u}(\hat{r}(\hat{q})) l_a(x) f(x,a) dx - \hat{c}'(a) \\ \int u(\hat{r}(\hat{q})) f(x,a) dx - c(a) &= \underline{U} = \int \hat{u}(\hat{r}(\hat{q})) f(x,a) dx - \hat{c}(a). \end{aligned}$$

Combining all the above conditions, we have

$$\mathbb{E}[u(\hat{r}(\hat{q})) - u(r(q))]q = 0.$$

These two conditions are sufficient for theorem 1 to apply. So estimated $\hat{q}$ is consistent with the true q. The asymptotic distribution of $\hat{h}(w)$ is in A9. Q.E.D. ■

1.8.11 Proof of Theorem 6

Proof. Similarly, we can estimate the profit loss, based on

$$\begin{aligned} L^*(\hat{h}, \lambda^*, \mu^*) &= \int [w - h^{-1}(\lambda^* + \mu^* \hat{l}_a]) f(x,a)dx \\ &+ \lambda^* [\int u(h^{-1}(\lambda^* + \mu^* \hat{l}_a), \gamma) f(x,a)dx - \int u(w,\gamma) f(x,a)dx] \\ &+ \mu^* \left[\int u(h^{-1}(\lambda^* + \mu^* \hat{l}_a), \gamma) \hat{l}_a f(x,a)dx - \int u(w,\gamma) \hat{l}_a f(x,a)dx \right]. \end{aligned}$$

Therefore, we can find a lower bound and a upper bound of the profit loss by choosing $\hat{h}$. That is $(\hat{h}_{\min}, \lambda^*(\hat{h}_{\min}), \mu^*(\hat{h}_{\min})) \in \arg\min_{h \in \mathcal{H}, h_w > 0, \lambda > 0, \mu > 0} L^*(h, \lambda^*, \mu^*)$ s.t. IC and IR constraints. Therefore the lower bound of profit loss is,

$$\Delta\Pi(h)^{lb} = \Delta\Pi(h_{\min}) = \mathbb{E}w_i - \mathbb{E}s^*(x_i, h_{\min}); \tag{1.32}$$

and the upper bound of profit loss is

$$\Delta\Pi(h)^{ub} = \Delta\Pi(h_{\max}) = \mathbb{E}w_i - \mathbb{E}s^*(x_i, h_{\max}). \tag{1.33}$$

Q.E.D. ■

1.8.12 Proof of Theorem 7

Proof. We prove (i) and (ii) together. Use the sample analogue $\hat{\Lambda}(\gamma, \mathbf{T})$ as an approximation of $\Lambda(\gamma, \mathbf{T})$, for the uniform convergence to apply, we need to confirm

$$\sup_{(\gamma,\mathbf{T})} \left|\hat{\Lambda}(\gamma, \mathbf{T}) - \Lambda(\gamma, \mathbf{T})\right| \rightarrow^p 0$$

Note that,

$$\begin{aligned}
&\sup_{\Upsilon} \left|\hat{\Lambda}(\gamma, \mathbf{T}) - \Lambda(\gamma, \mathbf{T})\right| \\
= &\sup_{\Upsilon} \left|\sum_{\mathcal{Z}} \frac{1}{n} \sum_{i=1}^{n_{\mathcal{Z}}} \left[\varepsilon_i(\Upsilon\,|\mathcal{Z})^2 - (\bar{J}(\mathcal{Z}) - J(\mathcal{Z}))\frac{h(w_i, \gamma\,|\mathcal{Z}) - \bar{h}(\gamma\,|\mathcal{Z})}{J(\mathcal{Z})^{*2}} + (\bar{Z}(\mathcal{Z}) - Z(\mathcal{Z}))\frac{l_a(x_i, \mathcal{Z}, \mathbf{T})}{Z(\mathcal{Z})^{*2}}\right]^2 - \Lambda\right. \\
\leq &\; 3\sup_{\Upsilon} \left|\frac{n_{\mathcal{Z}}}{n} \sum_{\mathcal{Z}} \left(\frac{1}{n_{\mathcal{Z}}} \sum_{i=1}^{n_{\mathcal{Z}}} \varepsilon_i(\Upsilon\,|\mathcal{Z})^2 - \mathbb{E}_x[\varepsilon(\Upsilon/\mathcal{Z})^2\,|\mathcal{Z}]\right)\right| \\
&+3\sup_{\Upsilon} \sum_{\mathcal{Z}} \frac{n_{\mathcal{Z}}}{n} (\bar{J}(\mathcal{Z}) - J(\mathcal{Z}))^2 \frac{1}{n_{\mathcal{Z}}} \sum_{i=1}^{n_{\mathcal{Z}}} \left(\frac{h(w_i, \gamma\,|\mathcal{Z}) - \bar{h}(\gamma\,|\mathcal{Z})}{J(\mathcal{Z})^{*2}}\right)^2 \\
&+3\sup_{\Upsilon} \sum_{\mathcal{Z}} \frac{n_{\mathcal{Z}}}{n} (\bar{Z}(\mathcal{Z}) - Z(\mathcal{Z}))^2 \frac{1}{n_{\mathcal{Z}}} \sum_{i=1}^{n_{\mathcal{Z}}} \left(\frac{l_a(x_i, \mathbf{T}, \mathcal{Z})}{Z(\mathcal{Z})^{*2}}\right)^2
\end{aligned}$$

Note also that as $n \rightarrow \infty$, $\lim_{n\rightarrow\infty} \frac{1}{n}|\mathcal{Z}| = 0$, this means $\sup_{\mathcal{Z}} n_{\mathcal{Z}} \rightarrow \infty$, for any $\mathcal{Z}\in$Z such that $\frac{n_{\mathcal{Z}}}{n} = k > 0$, $J^* \rightarrow^p J$, $\bar{J} \rightarrow^p J$, $Z^* \rightarrow^p Z$, $\bar{Z} \rightarrow Z$, $\frac{1}{n_{\mathcal{Z}}} \sum \left(h(w_i, \gamma\,|\mathcal{Z}) - \bar{h}(w, \gamma\,|\mathcal{Z})\right)^2 \rightarrow Q_{\mathcal{Z}} < \infty$, $\frac{1}{n_{\mathcal{Z}}} \sum l_a^2(x_i,\mathcal{Z}, \hat{\mathbf{T}}) \rightarrow Z_{\mathcal{Z}} < \infty$. For any $\mathcal{Z}\in$Z such that $\frac{n_{\mathcal{Z}}}{n} < 1$, we have $\frac{n_{\mathcal{Z}}}{n}(\bar{Z}(\mathcal{Z}) - Z(\mathcal{Z}))^2 = 0$, $\frac{n_{\mathcal{Z}}}{n}(\bar{J}(\mathcal{Z}) - J(\mathcal{Z}))^2 = 0$. Thus, the only item left is to show that

$$\sup_{\Upsilon} \left|\frac{n_{\mathcal{Z}}}{n} \sum_{\mathcal{Z}} \left(\frac{1}{n_{\mathcal{Z}}} \sum_{i=1}^{n_{\mathcal{Z}}} \varepsilon_i(\Upsilon/\mathcal{Z})^2 - \mathbb{E}_x \varepsilon(\Upsilon/\mathcal{Z})^2\right)\right| \rightarrow^p 0.$$

This is implied by the uniform law of large numbers, because for a given $\mathcal{Z}$,

$$\mathbb{E}[\sup_{(\gamma,\mathbf{T})} \left(\frac{h(w, \gamma\,|\mathcal{Z}) - \mathbb{E}_x[h(w, \gamma)\,|\mathcal{Z}]}{\mathbb{E}_x[h(w, \gamma\,|\mathcal{Z})l_a(x, \mathcal{Z}, \mathbf{T})\,|\mathcal{Z}]} - \frac{l_a(x, \mathcal{Z}, \mathbf{T})}{\mathbb{E}_x[l_a^2(x, \mathcal{Z}, \mathbf{T})\,|\mathcal{Z}]}\right)^2] < \infty$$

similar to the proof of in previous section (see theorem 3). Thus $\sup_{\boldsymbol{\delta}} \left|\hat{\Lambda}(\boldsymbol{\delta}) - \Lambda(\boldsymbol{\delta})\right| \rightarrow^p 0$.

Based on the above conditions, the extremum estimator of $(\hat{\gamma}, \hat{\mathbf{T}}) \in \arg\min_{(\boldsymbol{\delta},\lambda,\mu)} \hat{\Lambda}(\gamma, \mathbf{T})$

is a consistent estimator of the minimizer of $\Lambda(\gamma, \mathbf{T})$. The rest part of proof is similar to the proof of theorem 3. Under the null hypothesis, the asymptotic variance of $\widehat{\bar{\rho}}$ is regarded as asymptotic variance of weighted sum of $\hat{\rho}_{\mathcal{Z}}(\hat{\boldsymbol{\delta}})$, this is,

$$n_s AVar(\frac{1}{n_s}\sum_{j=1}^{n}\hat{\rho}_j) = \frac{\mathbf{1}'_{n_s} nACov(\hat{\rho}, \hat{\rho}')\mathbf{1}_{n_s}}{n_s},$$

where $\mathbf{1}_{n_s}$ is a vector of ones. Q.E.D. ■

1.9 Technical Supplements

1.9.1 TS1. Techical Supplement to Proof of Proposition 3

Proof. (ii) To derive the asymptotic distribution of $\widehat{\Delta\Pi}$ based on $\hat{\gamma}$, we rewrite the above equations as follows:

$$
\begin{aligned}
& \hat{\mathbb{E}}u(h^{-1}[\hat{\lambda}^* + \hat{\lambda}^* l_a(x_i, \hat{T}), \gamma], \hat{\gamma}) l_a(x_i, \hat{T}) - \hat{\mathbb{E}}u(h^{-1}[\lambda^* + \lambda^* l_a(x, \mathbf{T}^*), \gamma], \gamma) l_a(x_i, T^*) \\
& + \hat{\mathbb{E}}u(h^{-1}[\lambda^* + \lambda^* l_a(x, \mathbf{T}^*), \gamma], \gamma) l_a(x_i, T^*) - \int u(h^{-1}[\lambda^* + \lambda^* l_a(x, T^*), \gamma], \gamma) f_a(x, T^*) dx \\
= \; & \frac{1}{n} \sum_{i=1}^{n} u(w_i, \hat{\gamma}) l_a(x_i, \hat{T}) - \int u(w, \gamma) l_a(x_i, \mathbf{T}^*) f(x, a) dx
\end{aligned}
$$

and

$$
\begin{aligned}
& \hat{\mathbb{E}}u(h^{-1}[\hat{\lambda}^* + \hat{\lambda}^* l_a(x_i, \hat{T}), \gamma], \hat{\gamma}) - \hat{\mathbb{E}}u(h^{-1}[\lambda^* + \lambda^* l_a(x, \mathbf{T}^*), \gamma], \gamma) \\
& + \hat{\mathbb{E}}u(h^{-1}[\lambda^* + \lambda^* l_a(x, \mathbf{T}^*), \gamma], \gamma) - \int u(h^{-1}[\lambda^* + \lambda^* l_a(x, T^*), \gamma], \gamma) f(x, T^*) dx \\
= \; & \frac{1}{n} \sum_{i=1}^{n} u(w_i, \hat{\gamma}) - \int u(w, \gamma) f(x, a) dx
\end{aligned}
$$

For notational convenience, we compress notation γ when there is no confusion in expression $u(s^*, \gamma)$ or $u(w, \gamma)$ where γ is not the estimated value.

Note that,

$$
\begin{aligned}
& u(\hat{s}^*(x), \hat{\gamma}) - u(s^*(x), \gamma) \\
= \; & \left(h^{-1}[\hat{\lambda}^* + \hat{\mu}^* l_a(x, \hat{T}^*)] - h^{-1}[\lambda^* + \mu^* l_a(x, \mathbf{T}^*)] \right) + \left(u_w(s^*(x)) \frac{\partial s^*}{\partial \gamma} + \frac{\partial u(s^*(x))}{\partial \gamma} \right) (\hat{\gamma} - \gamma) \\
= \; & \frac{u_w(s^*(x))}{h_w(s^*(x))} \left(\hat{\lambda}^* - \lambda^* + (\hat{\mu}^* - \mu^*) l_a(x, \hat{T}^*) + \mu^* (l_a(x, \hat{T}^*) - l_a(x, \mathbf{T}^*)) + h.o. \right) \\
& + \left(u_w(s^*(x)) \frac{\partial s^*}{\partial \gamma'} + \frac{\partial u(s^*(x))}{\partial \gamma'} \right) (\hat{\gamma} - \gamma),
\end{aligned}
$$

therefore,

$$\begin{aligned}
& \hat{\mathbb{E}}\frac{u_w(s^*(x_i))}{h_w(s^*(x_i))}\left(\hat{\lambda}^* - \lambda^* + (\hat{\mu}^* - \mu^*)l_a(x_i,\hat{T}^*) + \mu^*(l_a(x_i,\hat{\mathbf{T}}^*) - l_a(x_i,\mathbf{T}^*)) + h.o.\right) l_a(x_i,\hat{T}^*) \\
& +\hat{\mathbb{E}}\left(u_w(s^*(x_i))\frac{\partial s^*}{\partial \gamma'} + \frac{\partial u(s^*(x_i))}{\partial \gamma'} + h.o.\right) l_a(\hat{\gamma}-\gamma) \\
& +\hat{\mathbb{E}}u(s^*(x_i))(l_a(x_i,\hat{T}^*) - l_a(x,\mathbf{T}^*)) + \hat{\mathbb{E}}u(s^*(x_i))l_a(x_i,\mathbf{T}^*) - \mathbb{E}u(s^*(x_i))l_a(x_i,\mathbf{T}^*) \\
= & \ \hat{\mathbb{E}}u(w_i,\hat{\gamma})l_a(x_i,\hat{T}) - \mathbb{E}u(w_i)l_a(x_i,\mathbf{T}^*)
\end{aligned}$$

and

$$\begin{aligned}
& \hat{\mathbb{E}}\frac{u_w(s^*(x_i))}{h_w(s^*(x_i))}\left(\hat{\lambda}^* - \lambda^* + (\hat{\mu}^* - \mu^*)l_a(x_i,\hat{T}^*) + \mu^*(l_a(x_i,\hat{T}^*) - l_a(x_i,\mathbf{T}^*)) + h.o.\right) \\
& +\hat{\mathbb{E}}\left(u_w(s^*(x_i))\frac{\partial s^*}{\partial \gamma'} + \frac{\partial u(s^*(x_i))}{\partial \gamma'} + h.o.\right)(\hat{\gamma}-\gamma) + \hat{\mathbb{E}}u(s^*(x_i)) - \mathbb{E}u(s^*(x_i)) \\
= & \ \hat{\mathbb{E}}u(w_i,\hat{\gamma}) - \mathbb{E}u(w_i)
\end{aligned}$$

For convenience, denote $\hat{\mathbb{E}}\frac{u_w(s^*(x_i))}{h_w(s^*(x_i))}l_a(x_i,\hat{T}^*) = \hat{k}_1$, $\hat{\mathbb{E}}\frac{u_w(s^*(x_i))}{h_w(s^*(x_i))}l_a^2(x_i,\hat{T}^*) = \hat{k}_2$, $\hat{\mathbb{E}}\frac{u_w(s^*(x_i))}{h_w(s^*(x_i))} = \hat{k}_3$, therefore,

$$\begin{cases} \hat{\lambda}^* - \lambda^* = \frac{\hat{k}_1\hat{A} - \hat{k}_2\hat{B}}{\hat{k}_1^2 - \hat{k}_2\hat{k}_3} \\ \hat{\mu}^* - \mu^* = \frac{\hat{k}_3\hat{A} - \hat{k}_1\hat{B}}{\hat{k}_2\hat{k}_3 - \hat{k}_1^2} \end{cases},$$

where

$$\begin{aligned}
\hat{A} = & \ \hat{\mathbb{E}}u(w_i,\hat{\gamma})l_a(x_i,\hat{T}) - \mathbb{E}u(w_i,\gamma)l_a(x_i,\mathbf{T}^*) - \hat{\mathbb{E}}\frac{u_w(s^*(x_i))}{h_w(s^*(x_i))}[\mu^*(l_a(x_i,\hat{T}^*) - l_a(x_i,\mathbf{T}^*)) + h.o.]l_a(x_i,\hat{T}^*) \\
& -\{\hat{\mathbb{E}}u(s^*(x_i))(l_a(x_i,\hat{T}^*) - l_a(x,\mathbf{T}^*)) + \hat{\mathbb{E}}u(s^*(x_i))l_a(x_i,\mathbf{T}^*) - \mathbb{E}u(s^*(x_i))l_a(x_i,\mathbf{T}^*)\} \\
& -\hat{\mathbb{E}}\left(u_w(s^*)\frac{\partial s^*}{\partial \gamma'} + \frac{\partial u(s^*)}{\partial \gamma'} + h.o.\right) l_a(\hat{\gamma}-\gamma)
\end{aligned}$$

and

$$\begin{aligned}
\hat{B} = & \ \hat{\mathbb{E}}u(w_i,\hat{\gamma}) - \mathbb{E}u(w_i,\gamma) - \hat{\mathbb{E}}\frac{u_w(s^*(x_i))}{h_w(s^*(x_i))}[\mu^*(l_a(x_i,\hat{T}^*) - l_a(x_i,\mathbf{T}^*)) + h.o.] \\
& -\{\hat{\mathbb{E}}u(s^*(x_i)) - \mathbb{E}u(s^*(x_i))\} - \hat{\mathbb{E}}\left(u_w(s^*)\frac{\partial s^*}{\partial \gamma'} + \frac{\partial u(s^*)}{\partial \gamma'} + h.o.\right)(\hat{\gamma}-\gamma)
\end{aligned}$$

Now we show the plim of $\hat{A}$ and $\hat{B}$ first. Let, $p\lim k_i = k_i$ (i=1,2,3), we have

$$p\lim_{n\to\infty}\sqrt{n}\hat{A} = 0, \text{ and } p\lim_{n\to\infty}\sqrt{n}\hat{B} = 0.$$

And by the delta method,

$$\begin{aligned}
& AVar(\sqrt{n}\hat{A}) \\
= \; & Var(u(w_i)l_a) + [\mathbb{E}u(w_i)l_{a\mathbf{T}'}]\mathbf{Z}^{-1}\mathbb{E}u(w_i)l_{a\mathbf{T}} + [\mathbb{E}u_{\gamma'}(w_i)l_a]\mathbf{\Gamma}[\mathbb{E}u_{\gamma}(w_i)l_a] + [\mathbb{E}u_{\gamma'}(w_i)l_a]\sigma_{\gamma T'}\mathbb{E}u(w_i)l_{a\mathbf{T}} \\
& +\mu^{*2}[\mathbb{E}\frac{u_w(s^*)}{h_w(s^*))}l_a l_{a\mathbf{T}'}]\mathbf{Z}^{-1}\mathbb{E}\frac{u_w(s^*)}{h_w(s^*))}l_a l_{a\mathbf{T}} + [\mathbb{E}u(s^*)l_{a\mathbf{T}'}]\mathbf{Z}^{-1}\mathbb{E}u(s^*)l_{a\mathbf{T}} + Var(u(s^*)l_a) \\
& -2[\mathbb{E}u(w_i)l_{a\mathbf{T}'}]\mathbf{Z}^{-1}\mathbb{E}\frac{u_w(s^*)}{h_w(s^*)}\mu^* l_a l_{a\mathbf{T}} - 2\mathbb{E}u(w_i)l_{a\mathbf{T}'}\mathbf{Z}^{-1}\mathbb{E}u(s^*)l_{a\mathbf{T}} + 2[\mathbb{E}u(s^*)l_{a\mathbf{T}'}]\mathbf{Z}^{-1}\mathbb{E}\frac{u_w(s^*)}{h_w(s^*)}\mu^* l_a l_{a\mathbf{T}} \\
& +[\mathbb{E}(u_w(s^*)s^*_{\gamma'} + u_{\gamma'}(s^*))l_a]\mathbf{\Gamma}[\mathbb{E}(u_w(s^*)s^*_{\gamma} + u_{\gamma}(s^*))l_a] - 2[\mathbb{E}u_{\gamma'}(w_i)l_a]\mathbf{\Gamma}[\mathbb{E}(u_w(s^*)s^*_{\gamma} + u_{\gamma}(s^*))l_a] \\
& -2[\mathbb{E}(u_w(s^*)s^*_{\gamma'} + u_{\gamma'}(s^*))l_a]\sigma_{\gamma T'}\mathbb{E}u(w_i)l_{a\mathbf{T}} + 2\mu^*[\mathbb{E}(u_w(s^*)s^*_{\gamma'} + u_{\gamma'}(s^*))l_a]\sigma_{\gamma T'}\mathbb{E}\frac{u_w(s^*)}{h_w(s^*))}l_a l_{a\mathbf{T}} \\
& +2[\mathbb{E}(u_w(s^*)s^*_{\gamma'} + u_{\gamma'}(s^*))l_a]\sigma_{\gamma T'}\mathbb{E}u(s^*)l_{a\mathbf{T}};
\end{aligned}$$

$$\begin{aligned}
& AVar(\sqrt{n}\hat{B}) \\
= \; & Var(u(w_i)) + [\mathbb{E}u_{\gamma'}(w_i)]\mathbf{\Gamma}[\mathbb{E}u_{\gamma}(w_i)] + \mu^{*2}[\mathbb{E}\frac{u_w(s^*)}{h_w(s^*)}l_{a\mathbf{T}}]\mathbf{Z}^{-1}\mathbb{E}\frac{u_w(s^*)}{h_w(s^*)}l_{a\mathbf{T}'} + Var(u(s^*)) \\
& +[\mathbb{E}(u_w(s^*)s^*_{\gamma'} + u_{\gamma'}(s^*))]\mathbf{\Gamma}[\mathbb{E}(u_w(s^*)s^*_{\gamma} + u_{\gamma}(s^*))] - 2[\mathbb{E}u_{\gamma'}(w_i)]\mathbf{\Gamma}[\mathbb{E}(u_w(s^*)s^*_{\gamma} + u_{\gamma}(s^*))] \\
& -2[\mathbb{E}u_{\gamma'}(w_i)]\boldsymbol{\sigma}_{\gamma T'}[\mathbb{E}\frac{u_w(s^*)}{h_w(s^*)}\mu^* l_{a\mathbf{T}} + 2[\mathbb{E}(u_w(s^*)s^*_{\gamma'} + u_{\gamma'}(s^*))]\boldsymbol{\sigma}_{\gamma T'}\mathbb{E}\frac{u_w(s^*)}{h_w(s^*)}\mu^* l_{a\mathbf{T}};
\end{aligned}$$

and,

$$\begin{aligned}
& ACov(\sqrt{n}\hat{A}, \sqrt{n}\hat{B}) \\
= \; & \left([\mathbb{E}[u_{\gamma'}(w_i) - u_w(s^*)s^*_{\gamma'} - u_{\gamma'}(s^*)]l_a]\mathbf{\Gamma} + [\mathbb{E}[u(w_i) - \mu^*\frac{u_w(s^*)}{h_w(s^*)}l_a - u(s^*)]l_{aT'}]\sigma_{\gamma T'}\right)\mathbb{E}u_{\gamma}(w_i) \\
& +\mu^*\left([\mathbb{E}[u_{\gamma'}(w_i) - u_w(s^*)s^*_{\gamma'} - u_{\gamma'}(s^*)]l_a]\sigma_{\gamma T'} + [\mathbb{E}[u(w_i) - \mu^*\frac{u_w(s^*)}{h_w(s^*)}l_a - u(s^*)]l_{aT'}]\mathbf{Z}^{-1}\right)\mathbb{E}l_{a\mathbf{T}}\frac{u_w(s^*)}{h_w(s^*)} \\
& -\left([\mathbb{E}[u_{\gamma'}(w_i) - u_w(s^*)s^*_{\gamma'} - u_{\gamma'}(s^*)]l_a]\mathbf{\Gamma} + [\mathbb{E}[u(w_i) - \mu^*\frac{u_w(s^*)}{h_w(s^*)}l_a - u(s^*)]l_{aT'}]\sigma_{\gamma T'}\right)\mathbb{E}(u_w(s^*)s^*_{\gamma} + u_{\gamma'}
\end{aligned}$$

By the continuous mapping theorem,

$$p\lim_{n\to\infty}\sqrt{n}(\hat{\lambda}^*-\lambda^*)=0,\text{ and } p\lim_{n\to\infty}\sqrt{n}(\hat{\mu}^*-\mu^*)=0.$$

and

$$nAVar(\hat{\lambda}^*)=AVar[\sqrt{n}(\hat{\lambda}^*-\lambda^*)]=\frac{k_1^2AVar(\sqrt{n}\hat{A})+k_2^2AVar(\sqrt{n}\hat{B})-2k_1k_2ACov(\sqrt{n}\hat{A},\sqrt{n}\hat{B})}{[k_1^2-k_2k_3]^2}$$

$$nAVar(\hat{\mu}^*)=AVar[\sqrt{n}(\hat{\mu}^*-\mu^*)]=\frac{k_3^2AVar(\sqrt{n}\hat{A})+k_1^2AVar(\sqrt{n}\hat{B})-2k_1k_3ACov(\sqrt{n}\hat{A},\sqrt{n}\hat{B})}{[k_1^2-k_2k_3]^2}$$

$$\begin{aligned} nACov(\hat{A},\hat{T}^{*\prime}) &= \lim_{n\to\infty} Cov(\sqrt{n}\hat{A},\sqrt{n}(\hat{T}^*-\mathbf{T}^*)') \\ &= [\mathbb{E}u_{\gamma'}(w_i)l_a-\mathbb{E}(u_w(s^*)s^*_{\gamma'}+u_{\gamma'}(s^*))l_a]\sigma_{\gamma T'}+[\mathbb{E}(u(w_i)-\mu^*\frac{u_w(s^*)}{h_w(s^*)}l_a-u(s^*)]l_{aT'})\mathbf{Z}^{-1} \end{aligned}$$

$$\begin{aligned} nACov(\hat{A},\hat{\gamma}') &= \lim_{n\to\infty} Cov(\sqrt{n}\hat{A},\sqrt{n}(\hat{\gamma}-\gamma)') \\ &= [\mathbb{E}u_{\gamma'}(w_i)l_a-\mathbb{E}(u_w(s^*)s^*_{\gamma'}+u_{\gamma'}(s^*))l_a]\mathbf{\Gamma}+[\mathbb{E}(u(w_i)-\mu^*\frac{u_w(s^*)}{h_w(s^*)}l_a-u(s^*)]l_{aT'})\boldsymbol{\sigma}'_{\gamma T'} \end{aligned}$$

$$\begin{aligned} nACov(\hat{B},\hat{T}^{*\prime}) &= \lim_{n\to\infty} Cov(\sqrt{n}\hat{B},\sqrt{n}(\hat{T}^*-\mathbf{T}^*)') \\ &= [\mathbb{E}u_{\gamma'}(w_i)-\mathbb{E}(u_w(s^*)s^*_{\gamma'}+u_{\gamma'}(s^*))]\sigma_{\gamma T'}+\mu^*[\mathbb{E}\frac{u_w(s^*)}{h_w(s^*)}l_{aT'})]\mathbf{Z}^{-1} \end{aligned}$$

$$\begin{aligned} nACov(\hat{B},\hat{\gamma}') &= \lim_{n\to\infty} Cov(\sqrt{n}\hat{B},\sqrt{n}(\hat{\gamma}-\gamma)') \\ &= [\mathbb{E}u_{\gamma'}(w_i)-\mathbb{E}(u_w(s^*)s^*_{\gamma'}+u_{\gamma'}(s^*))]\mathbf{\Gamma}+\mu^*[\mathbb{E}\frac{u_w(s^*)}{h_w(s^*)}l_{aT'})]\boldsymbol{\sigma}'_{\gamma T'} \end{aligned}$$

Therefore,

$$\begin{aligned} nACov(\hat{\lambda}^*,\hat{\mu}^*) &= \lim_{n\to\infty} Cov(\sqrt{n}(\hat{\lambda}^*-\lambda^*),\sqrt{n}(\hat{\mu}^*-\mu^*)) \\ &= \frac{-k_1k_3AVar(\sqrt{n}\hat{A})-k_1k_2AVar(\sqrt{n}\hat{B})+(k_1^2+k_2k_3)ACov(\sqrt{n}\hat{A},\sqrt{n}\hat{B})}{[k_1^2-k_2k_3]^2} \end{aligned}$$

$$
\begin{aligned}
nACov(\hat{\lambda}^*, \hat{T}^*) &= \lim_{n\to\infty} Cov(\sqrt{n}(\hat{\lambda}^* - \lambda^*), \sqrt{n}(\hat{T}^* - \mathbf{T}^*)) \\
&= \frac{1}{k_1^2 - k_2k_3}(k_1 nACov(\hat{A}, \hat{T}^{*\prime})' - k_2 nACov(\hat{B}, \hat{T}^{*\prime})') \\
&= \frac{\mathbf{Z}^{-1}}{k_1^2 - k_2k_3}\left(k_1\mathbb{E}\left(u(w_i) - u(s^*) - \frac{u'(s^*)\mu^* l_a}{h_w(s^*)}\right) l_{a\mathbf{T}} + k_2\mathbb{E}\frac{u'(s^*)\mu^* l_{a\mathbf{T}}}{h_w(s^*)}\right) \\
&\quad + \frac{\sigma'_{\gamma T'}}{k_1^2 - k_2k_3}\left(k_1\mathbb{E}[u_{\gamma'}(w_i) - u_w(s^*)s^*_{\gamma'} - u_{\gamma'}(s^*)]l_a - k_2\mathbb{E}[u_{\gamma'}(w_i) - u_w(s^*)s^*_{\gamma'} - u_{\gamma'}(s^*)]\right)
\end{aligned}
$$

$$
\begin{aligned}
nACov(\hat{\mu}^*, \hat{T}^*) &= \lim_{n\to\infty} Cov(\sqrt{n}(\hat{\mu}^* - \mu^*), \sqrt{n}(\hat{T}^* - \mathbf{T}^*)) \\
&= -\frac{1}{k_1^2 - k_2k_3}(k_3 nACov(\hat{A}, \hat{T}^{*\prime})' - k_1 nACov(\hat{B}, \hat{T}^{*\prime})') \\
&= -\frac{\mathbf{Z}^{-1}}{k_1^2 - k_2k_3}\left(k_3\mathbb{E}\left(u(w_i) - u(s^*) - \frac{u'(s^*)\mu^* l_a}{h_w(s^*)}\right) l_{a\mathbf{T}} + k_1\mathbb{E}\frac{u'(s^*)\mu^* l_{a\mathbf{T}}}{h_w(s^*)}\right) \\
&\quad - \frac{\sigma'_{\gamma T'}}{k_1^2 - k_2k_3}\left(k_3\mathbb{E}[u_{\gamma'}(w_i) - u_w(s^*)s^*_{\gamma'} - u_{\gamma'}(s^*)]l_a - k_1\mathbb{E}[u_{\gamma'}(w_i) - u_w(s^*)s^*_{\gamma'} - u_{\gamma'}(s^*)]\right)
\end{aligned}
$$

$$
\begin{aligned}
nACov(\hat{\lambda}^*, \hat{\gamma}) &= \frac{1}{k_1^2 - k_2k_3}(k_1 nACov(\hat{A}, \hat{\gamma}')' - k_2 nACov(\hat{B}, \hat{\gamma}')') \\
nACov(\hat{\mu}^*, \hat{\gamma}) &= -\frac{1}{k_1^2 - k_2k_3}(k_3 nACov(\hat{A}, \hat{\gamma}')' - k_1 nACov(\hat{B}, \hat{\gamma}')')
\end{aligned}
$$

Based on the above formulas, for every given x,

$$
\begin{aligned}
\sqrt{n}[\hat{s}^*(x) - s^*(x)] &= \sqrt{n}\left(h^{-1}[\hat{\lambda}^* + \hat{\mu}^* l_a(x, \hat{T}^*), \hat{\gamma}] - h^{-1}[\lambda^* + \mu^* l_a(x, \mathbf{T}^*), \gamma]\right) \\
&= \sqrt{n}\frac{1}{h_w(s^*(x))}\left(\hat{\lambda}^* - \lambda^* + (\hat{\mu}^* - \mu^*)l_a(x, \hat{\mathbf{T}}^*) + \mu^*(l_a(x, \hat{T}^*) - l_a(x, \mathbf{T}^*)) + h.o.\right) \\
&\quad + \sqrt{n}(s^*_{\gamma'}(x)(\hat{\gamma} - \gamma) + h.o.)
\end{aligned}
$$

then,

$$nAVar(\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i))) = (\mathbb{E}\frac{1}{h_w(s^*)})^2 nAVar(\hat{\lambda}^*) + (\mathbb{E}\frac{l_a}{h_w(s^*)})^2 nAVar(\hat{\mu}^*)$$
$$+2nACov(\hat{\lambda}^*, \mu^*)(\mathbb{E}\frac{l_a}{h_w(s^*)})(\mathbb{E}\frac{1}{h_w(s^*)}) + \mu^{*2}(\mathbb{E}\frac{l_{a\mathbf{T}'}}{h_w(s^*)})\mathbf{Z}^{-1}(\mathbb{E}\frac{l_{a\mathbf{T}}}{h_w(s^*)})$$
$$+2(\mathbb{E}\frac{1}{h_w(s^*)})\mathbb{E}\frac{\mu^* l_{a\mathbf{T}'}}{h_w(s^*)} nACov(\hat{\lambda}^*, \hat{\mathbf{T}}^*) + 2(\mathbb{E}\frac{l_a}{h_w(s^*)})\mathbb{E}\frac{\mu^* l_{a\mathbf{T}'}}{h_w(s^*)} nACov(\hat{\mu}^*, \hat{\mathbf{T}}^*)$$
$$+\mathbb{E}s^*_{\gamma'}(x_i)\{\mathbf{\Gamma}\mathbb{E}s^*_{\gamma}(x_i) + 2\mathbb{E}\frac{1}{h_w(s^*)} nACov(\hat{\lambda}^*, \hat{\gamma}) + 2\mathbb{E}\frac{l_a}{h_w(s^*)} nACov(\hat{\mu}^*, \hat{\gamma}) + 2\sigma_{\gamma T'}\mathbb{E}\frac{\mu^* l_{aT}}{h_w(s^*)}\}.$$

Additionally,

$$\begin{aligned} nACov(\hat{A}, \hat{\mathbb{E}}w_i - \mathbb{E}w_i) &= \lim_{n\to\infty} Cov(\sqrt{n}\hat{A}, \sqrt{n}[\hat{\mathbb{E}}w_i - \mathbb{E}w_i]) \\ &= Cov(u(w_i)l_a, w_i) - Cov(u(s^*)l_a, w_i) = \mathbb{E}(u(w_i) - u(s^*))l_a w_i \end{aligned}$$

and

$$\begin{aligned} nACov(\hat{B}, \hat{\mathbb{E}}w_i - \mathbb{E}w_i) &= \lim_{n\to\infty} Cov(\sqrt{n}\hat{B}, \sqrt{n}[\hat{\mathbb{E}}w_i - \mathbb{E}w_i]) \\ &= Cov(u(w_i), w_i) - Cov(u(s^*), w_i) = \mathbb{E}(u(w_i) - u(s^*))w_i, \end{aligned}$$

therefore,

$$nACov(\hat{\lambda}^*, \hat{\mathbb{E}}w_i - \mathbb{E}w_i) = \frac{1}{k_1^2 - k_2k_3}[k_1\mathbb{E}(u(w_i) - u(s^*))l_a w_i - k_2\mathbb{E}(u(w_i) - u(s^*))w_i]$$

and

$$nACov(\hat{\mu}^*, \hat{\mathbb{E}}w_i - \mathbb{E}w_i) = -\frac{1}{k_1^2 - k_2k_3}[k_3\mathbb{E}(u(w_i) - u(s^*))l_a w_i - k_1\mathbb{E}(u(w_i) - u(s^*))w_i],$$

therefore, we have

$$\begin{aligned}
& nACov(\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i)), \hat{\mathbb{E}}w_i - \mathbb{E}w_i) \\
= & \lim_{n\to\infty} Cov(\sqrt{n}\hat{\mathbb{E}}[\hat{s}^*(x_i) - s^*(x_i)], \sqrt{n}[\hat{\mathbb{E}}w_i - \mathbb{E}w_i]) \\
= & nACov(\hat{\lambda}^*, \hat{\mathbb{E}}w_i)\mathbb{E}\frac{1}{h_w(s^*)} + nACov(\hat{\mu}^*, \hat{\mathbb{E}}w_i)\mathbb{E}\frac{l_a}{h_w(s^*)} \\
= & \frac{\mathbb{E}(u(w_i) - u(s^*))l_a w_i}{k_1^2 - k_2k_3}[k_1\mathbb{E}\frac{1}{h_w(s^*)} - k_3\mathbb{E}\frac{l_a}{h_w(s^*)}] \\
& -\frac{\mathbb{E}(u(w_i) - u(s^*))w_i}{k_1^2 - k_2k_3}[k_2\mathbb{E}\frac{1}{h_w(s^*)} - k_1\mathbb{E}\frac{l_a}{h_w(s^*)}],
\end{aligned}$$

similarly,

$$\begin{aligned}
& nACov(\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i)), \hat{\mathbb{E}}s^*(x_i) - \mathbb{E}s^*(x_i)) \\
= & \frac{\mathbb{E}(u(w_i) - u(s^*))l_a s^*}{k_1^2 - k_2k_3}[k_1\mathbb{E}\frac{1}{h_w(s^*)} - k_3\mathbb{E}\frac{l_a}{h_w(s^*)}] \\
& -\frac{\mathbb{E}(u(w_i) - u(s^*))s^*}{k_1^2 - k_2k_3}[k_2\mathbb{E}\frac{1}{h_w(s^*)} - k_1\mathbb{E}\frac{l_a}{h_w(s^*)}]
\end{aligned}$$

It is straightforward to see,

$$p\lim\sqrt{n}(\widehat{\Delta\Pi} - \Delta\Pi) = p\lim\sqrt{n}[\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i)) + \hat{\mathbb{E}}s^*(x_i) - \mathbb{E}s^*(x)] - p\lim\sqrt{n}[\hat{\mathbb{E}}w_i - \mathbb{E}w_i] = 0.$$

As a result,

$$\begin{aligned}
& nAVar(\widehat{\Delta\Pi}) = Var(w_i) + nAVar(\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i))) + Var(s^*(x_i)) \\
& -2nCov(\hat{\mathbb{E}}w_i - \mathbb{E}w_i, \hat{\mathbb{E}}s^*(x_i) - \mathbb{E}s^*(x_i)) - 2nACov(\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i)), \hat{\mathbb{E}}w_i - \mathbb{E}w_i) \\
& +2nACov(\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i)), \hat{\mathbb{E}}s^*(x_i) - \mathbb{E}s^*(x_i)) \\
= & Var(w_i) + nAVar(\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i))) + Var(s^*(x_i)) - 2Cov(w_i, s^*(x_i)) \\
& -2\left(nACov(\hat{\lambda}^*, \hat{\mathbb{E}}w_i)\mathbb{E}\frac{1}{h_w(s^*)} + nACov(\hat{\mu}^*, \hat{\mathbb{E}}w_i)\mathbb{E}\frac{l_a}{h_w(s^*)}\right) \\
& +2\left(nACov(\hat{\lambda}^*, \hat{\mathbb{E}}s^*(x_i))\mathbb{E}\frac{1}{h_w(s^*)} + nACov(\hat{\mu}^*, \hat{\mathbb{E}}s^*(x_i))\mathbb{E}\frac{l_a}{h_w(s^*)}\right)
\end{aligned}$$

Using the sample analogues of the objects appear in the above formula, we can form $\widehat{nAVar(\widehat{\Delta\Pi})}$, a consistent estimator of $nAVar(\widehat{\Delta\Pi})$.

Particularly, under the null hypothesis that the w_i generated based on some optimal contract $s^*(x_i)$, and $s^*(x_i)$ can be estimated by $\hat{s}^*(x_i) = h^{-1}(\hat{\lambda}^* + \widehat{\mu^*} l_a(x_i, \hat{T}), \hat{\gamma})$, where $\hat{\lambda}^*$, $\widehat{\mu^*}$ and $\hat{\gamma}$ need not to be solved by equations IC and IR constraints, instead of directly plugging in $\hat{\lambda}^* = \hat{\mathbb{E}} h(w_i, \hat{\gamma}),$, $\widehat{\mu^*} = \frac{\widehat{Cov}(h(w_i,\hat{\gamma}), l_a(x_i,\hat{T}))}{\hat{\mathbb{E}} l_a^2(x_i,\hat{T})}$. In this case,

$$
\begin{aligned}
\widehat{\Delta\Pi} &= \hat{\mathbb{E}}(x_i - w_i) - \hat{\mathbb{E}}(x_i - \hat{s}^*(x_i)) = \hat{\mathbb{E}}(\hat{s}^*(x_i) - w_i) = \hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i)) \\
&= \hat{\mathbb{E}} \frac{1}{h_w(s^*(x_i))} \left(\hat{\lambda}^* - \lambda^* + (\hat{\mu}^* - \mu^*) l_a(x_i, \hat{T}^*) + \mu^*(l_a(x_i, \hat{T}^*) - l_a(x_i, \mathbf{T}^*)) + h.o. \right) + \hat{\mathbb{E}} s^*_{\gamma'}(x_i)(\hat{\gamma} - \gamma)
\end{aligned}
$$

Note that for random sample value A_i,

$$
\begin{aligned}
& nAVar[(\hat{\lambda}^* - \lambda^*)\hat{\mathbb{E}} A_i] \\
=& \lim_{n\to\infty} n\mathbb{E}((\hat{\lambda}^* - \lambda^*)\hat{\mathbb{E}} A_i)^2 - n \lim_{n\to\infty} (\mathbb{E}(\hat{\lambda}^* - \lambda^*)\hat{\mathbb{E}} A_i)^2 \\
=& \lim_{n\to\infty} \mathbb{E}(\hat{\lambda}^* - \lambda^*)^2 \frac{1}{n} \sum\sum A_j A_i - \lim_{n\to\infty} (\mathbb{E} \sum (\hat{\lambda}^* - \lambda^*) A_i)^2 \\
=& AVar(\hat{\lambda}^*) Var(A_i) - \frac{1}{n} Cov(A_i, \hat{\lambda}^*)^2 \\
=& \frac{1}{n} AVar(\sqrt{n}\hat{\lambda}^*) Var(A_i) - \frac{1}{n} [\mathbb{E}(h(w_i) - \lambda) A_i]^2
\end{aligned}
$$

Therefore,

$$nAVar(\widehat{\Delta\Pi}) = nAVar(\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i)))$$

where,

$$
\begin{aligned}
& AVar(\sqrt{n}\hat{A}) = 2Var(u(w_i) l_a) + [\mathbb{E} u_{\gamma'}(w_i) l_a] \sigma_{\gamma T'} \mathbb{E} u(w_i) l_{a\mathbf{T}} + \mu^{*2} [\mathbb{E} \frac{u_w(s^*)}{h_w(s^*))} l_a l_{a\mathbf{T}'}] \mathbf{Z}^{-1} \mathbb{E} \frac{u_w(s^*)}{h_w(s^*))} l_a l_{a\mathbf{T}} \\
& + [\mathbb{E}(u_w(s^*) s^*_{\gamma'}] \mathbf{\Gamma} [\mathbb{E}(u_w(s^*) s^*_{\gamma}] + 2\mu^* [\mathbb{E}(u_w(s^*) s^*_{\gamma'} + u_{\gamma'}(s^*)) l_a] \sigma_{\gamma T'} \mathbb{E} \frac{u_w(s^*)}{h_w(s^*))} l_a l_{a\mathbf{T}}; \\
& AVar(\sqrt{n}\hat{B}) = 2Var(u(w_i)) + \mu^{*2} [\mathbb{E} \frac{u'(s^*)}{h_w(s^*)} l_{a\mathbf{T}}] \mathbf{Z}^{-1} \mathbb{E} \frac{u'(s^*)}{h_w(s^*)} l_{a\mathbf{T}'} + [\mathbb{E}(u_w(s^*) s^*_{\gamma'}] \mathbf{\Gamma} [\mathbb{E}(u_w(s^*) s^*_{\gamma}] \\
& + 2[\mathbb{E}(u_w(s^*) s^*_{\gamma'}] \boldsymbol{\sigma}_{\gamma T'} [\mathbb{E} \frac{u_w(s^*)}{h_w(s^*)} \mu^* l_{a\mathbf{T}}].
\end{aligned}
$$

$$ACov(\sqrt{n}\hat{A},\sqrt{n}\hat{B}) = -\mu^*\left([\mathbb{E}u_w(s^*)s^*_{\gamma'}l_a]\sigma_{\gamma T'} - [\mathbb{E}\mu^*\frac{u_w(s^*)}{h_w(s^*)}l_a l_{aT'}]\mathbf{Z}^{-1}\right)\mathbb{E}l_{a\mathbf{T}}\frac{u_w(s^*)}{h_w(s^*)}$$
$$+\left([\mathbb{E}u_w(s^*)s^*_{\gamma'}l_a]\mathbf{\Gamma} + [\mathbb{E}\mu^*\frac{u_w(s^*)}{h_w(s^*)}l_a l_{aT'}]\sigma_{\gamma T'}\right)\mathbb{E}(u_w(s^*)s^*_{\gamma})$$

then $\frac{(\hat{\mathbb{E}}w_i - \hat{\mathbb{E}}\hat{s}^*)^2}{\widehat{AVar(\Delta\Pi)}} \to^d \chi^2_1$.

At times, if the function of $(\hat{\lambda}^*, \hat{\mu}^*)$ is too complicated, then we may reduce the intensity of computation by approximating. For example, if taking the first order approximation of $u(h^{-1}[\lambda^* + \mu^* l_a(x,\hat{T})])$ around observed contract $h(w)$, we have

$$\begin{aligned}
\hat{\mathbb{E}}\left[u(w_i) - \frac{u'(w_i)^3}{u''(w_i)}(\lambda^* + \mu^* l_a(x_i,\hat{T}) - h(w_i))\right] l_a(x_i,\hat{T}) &\cong \frac{1}{n}\sum_{i=1}^n u(w_i) l_a(x_i,\hat{T}) \\
\hat{\mathbb{E}}\left[u(w_i) - \frac{u'(w_i)^3}{u''(w_i)}(\lambda^* + \mu^* l_a(x_i,\hat{T}) - h(w_i))\right] &\cong \frac{1}{n}\sum_{i=1}^n u(w_i)
\end{aligned}$$

Use the same notation as in theorem 3, as $n \to \infty$, we obtain:

$$\begin{aligned}
\hat{k}_1\lambda^* + \hat{k}_2\mu^* + \hat{\mathbb{E}}\frac{u'(w_i)^2}{u''(w_i)} l_a(x_i,\hat{\mathbf{T}}) &= 0 \\
\hat{k}_3\lambda^* + \hat{k}_1\mu^* + \hat{\mathbb{E}}\frac{u'(w_i)^2}{u''(w_i)} &= 0.
\end{aligned}$$

(iii) Based on (ii), we can derive the asymptotic distribution of $\hat{M}^*$ and $\hat{K}^*$. The asymptotic consistency follows by the continuous mapping theorem. The asymptotic variance is based on the delta method. For a continuous function $g(x,a)$, the asymptotic variance of $\hat{\mathbb{E}}u(\hat{s}^*(x_i),\hat{\gamma})g(x_i,\hat{\mathbf{T}})$ can be found by follows.

$$\begin{aligned}
& nAVar(\hat{\mathbb{E}}u(\hat{s}^*(x_i),\hat{\gamma})g(x_i,\hat{\mathbf{T}})) \\
= \;& Var(u(s^*)g) + [\mathbb{E}u(s^*)g_{\mathbf{T}'}]\mathbf{Z}^{-1}[\mathbb{E}u(s^*)g_{\mathbf{T}}] + nAVar(\hat{\mathbb{E}}u'(s^*)(\hat{s}^* - s^*)g) \\
& + \mathbb{E}[(u_w(s^*)s^*_{\gamma'} + u_{\gamma'}(s^*))g]\mathbf{\Gamma}\mathbb{E}[(u_w(s^*)s^*_{\gamma} + u_{\gamma}(s^*))g] \\
& + 2\mathbb{E}[(u_w(s^*)s^*_{\gamma'} + u_{\gamma'}(s^*))g]nACov(\hat{\mathbb{E}}u_w(s^*)(\hat{s}^*(x_i) - s^*(x_i)),\hat{\gamma}) \\
& + 2\mathbb{E}[(u_w(s^*)s^*_{\gamma'} + u_{\gamma'}(s^*))g]\sigma_{\gamma T'}\mathbb{E}u(s^*)g_{\mathbf{T}} \\
& + 2[\mathbb{E}u(s^*)g_{\mathbf{T}'}]nACov(\hat{\mathbb{E}}u_w(s^*)(\hat{s}^*(x_i) - s^*(x_i)),\hat{T}),
\end{aligned}$$

where

$$
\begin{aligned}
& nAVar(\hat{\mathbb{E}}u_w(s^*)(\hat{s}^* - s^*)g) \\
= \quad & (\mathbb{E}\frac{u_w(s^*)g}{h_w(s^*)})^2 nAVar(\hat{\lambda}^*) + (\mathbb{E}\frac{u_w(s^*)gl_a}{h_w(s^*)})^2 nAVar(\hat{\mu}^*) \\
& + \mu^{*2}\mathbb{E}\left(\frac{u_w(s^*)gl_{a\mathbf{T}'}}{h_w(s^*)}\right)\mathbf{Z}^{-1}\mathbb{E}\left(\frac{u_w(s^*)gl_{a\mathbf{T}}}{h_w(s^*)}\right) + 2nACov(\hat{\lambda}^*, \hat{\mu}^*)(\mathbb{E}\frac{u_w(s^*)g}{h_w(s^*)})(\mathbb{E}\frac{u_w(s^*)gl_a}{h_w(s^*)}) \\
& + 2(\mu^*\mathbb{E}\frac{u_w(s^*)g}{h_w(s^*)}\mathbb{E}\frac{u_w(s^*)g}{h_w(s^*)}l_{a\mathbf{T}'})nACov(\hat{\lambda}^*, \hat{\mathbf{T}}^*) + 2(\mu^*\mathbb{E}\frac{u_w(s^*)gl_a}{h_w(s^*)}\mathbb{E}\frac{u_w(s^*)gl_a}{h_w(s^*)}l_{a\mathbf{T}'})nACov(\hat{\mu}^*, \hat{\mathbf{T}}^*)
\end{aligned}
$$

and

$$
\begin{aligned}
& nACov(\hat{\mathbb{E}}u_w(s^*)(\hat{s}^*(x_i) - s^*(x_i)), \hat{\gamma}) \\
= \quad & [\mathbb{E}\frac{u_w(s^*)}{h_w(s^*)}]nACov(\hat{\lambda}, \hat{\gamma}) + [\mathbb{E}\frac{u_w(s^*)}{h_w(s^*)}l_a]nACov(\hat{\mu}, \hat{\gamma}) + \mathbb{E}\frac{u_w(s^*)\mu^* l_{aT'}}{h_w(s^*)}nACov(\hat{\gamma}, \hat{T})
\end{aligned}
$$

and

$$
\begin{aligned}
& nACov(\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i)), \hat{T}) \\
= \quad & [\mathbb{E}\frac{u_w(s^*)}{h_w(s^*)}]nACov(\hat{\lambda}, \hat{T}) + [\mathbb{E}\frac{u_w(s^*)}{h_w(s^*)}l_a]nACov(\hat{\mu}, \hat{T}) + \mathbb{E}\frac{u_w(s^*)\mu^* l_{aT'}}{h_w(s^*)}\mathbf{Z}^{-1}
\end{aligned}
$$

Therefore, $nAVar(\hat{M}^*(\hat{\mathbf{T}}))$ can be computed by sending $g(x_i, \hat{\mathbf{T}}) = l_a(x_i, \hat{\mathbf{T}})$, and $nAVar(\hat{M}^*(\hat{\mathbf{T}}))$ can be computed by sending $g(x_i, \hat{\mathbf{T}}) = l_a^2(x_i, \hat{\mathbf{T}}) + l_{aa}(x_i, \hat{\mathbf{T}})$. Q.E.D. ■

1.9.2 TS2. Technical Supplement to Proof of Corollary 1

Proof. The asymptotic variance of AC can be derived as follows:

$$
\sqrt{n}(\widehat{AC} - AC) = \sqrt{n}\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i)) + \sqrt{n}(\hat{\mathbb{E}}s^*(x_i) - \mathbb{E}s^*(x_i)) - \sqrt{n}(\hat{w}^{fb} - w^{fb}) + h.o \to^p 0
$$

and

$$
\begin{aligned}
nAVar(\widehat{AC}) &= nAVar(\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i))) + Var(s^*(x_i)) - 2Cov(s^*(x_i), u(w_i))h(w^{fb}) \\
&\quad -2nACov(\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i)), \hat{\mathbb{E}}u(w_i) - \mathbb{E}u(w_i))h(w^{fb}) \\
&\quad +2nACov(\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i)), \hat{\mathbb{E}}s^*(x_i) - \mathbb{E}s^*(x_i)) \\
&\quad +h(w^{fb})^2[Var(u(w_i)) + \mathbb{E}u_{\gamma'}(w_i)\mathbf{\Gamma}\mathbb{E}u_{\gamma}(w_i)] + 2\frac{\partial u^{-1}(u(w^{fb},\gamma),\gamma)}{\partial \gamma'}\mathbf{\Gamma}\mathbb{E}u_{\gamma}(w_i) \\
&\quad +\frac{\partial u^{-1}(u(w^{fb},\gamma),\gamma)}{\partial \gamma'}\mathbf{\Gamma}\frac{\partial u^{-1}(u(w^{fb},\gamma),\gamma)}{\partial \gamma} \\
&\quad -2[\frac{\partial u^{-1}(u(w^{fb},\gamma),\gamma)}{\partial \gamma'} + h(w^{fb})\mathbb{E}u_{\gamma'}(w_i)]nACov(\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i)), \hat{\gamma}),
\end{aligned}
$$

where,

$$
\begin{aligned}
& nACov(\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i)), \hat{\gamma}) \\
= \; & [\mathbb{E}\frac{1}{h_w(s^*)}]nACov(\hat{\lambda}, \hat{\gamma}) + [\mathbb{E}\frac{1}{h_w(s^*)}l_a]nACov(\hat{\mu}, \hat{\gamma}) + \mathbb{E}\frac{\mu^* l_{aT'}}{h_w(s^*)}nACov(\hat{\gamma}, \hat{T})
\end{aligned}
$$

and

$$
\begin{aligned}
& nACov(\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i)), \hat{\mathbb{E}}u(w_i) - \mathbb{E}u(w_i)) \\
= \; & \frac{\mathbb{E}(u(w_i) - u(s^*))l_a u(w_i)}{k_1^2 - k_2k_3}[k_1\mathbb{E}\frac{1}{h_w(s^*)} - k_3\mathbb{E}\frac{l_a}{h_w(s^*)}] - \frac{\mathbb{E}(u(w_i) - u(s^*))u(w_i)}{k_1^2 - k_2k_3}[k_2\mathbb{E}\frac{1}{h_w(s^*)} - k_1\mathbb{E}\frac{l_a}{h_w(s^*)}]
\end{aligned}
$$

[$nAVar(\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i)))$ and $nACov(\hat{\mathbb{E}}(\hat{s}^*(x_i) - s^*(x_i)), \hat{\mathbb{E}}s^*(x_i) - \mathbb{E}s^*(x_i))$ are derived in TS1].

For the total profit loss, to keep the agent's utility the same as $\underline{U}$, a fixed payment w^{fb} needs to solve $u(w^{fb}) = \int u(w)f(x,a)dx$. The asymptotic distribution of w^{fb} is:

$$
\begin{aligned}
nAVar(\widehat{TL}) &= n \lim_{n\to\infty} \mathbb{E}\left[\hat{\mathbb{E}}w_i - \mathbb{E}w_i - \left[u^{-1}\left(\hat{\mathbb{E}}u(w_i)\right) - u^{-1}(\mathbb{E}u(w_i)))\right]\right]^2 \\
&\quad -n\left[\mathbb{E}\lim_{n\to\infty}\left(\hat{\mathbb{E}}w_i - \mathbb{E}w_i - \left[u^{-1}\left(\hat{\mathbb{E}}u(w_i)\right) - u^{-1}(\mathbb{E}u(w_i)))\right]\right)\right]^2 \\
&= Var(w_i) + Var(u(w_i))h(w^{fb})^2 - 2Cov(w_i, u(w_i))h(w^{fb}).
\end{aligned}
$$

Q.E.D. ■

1.9.3 TS3. Analytic derivation of example 1

Proof. The contract B is a piece-rate contract. We can solve the potential contract by

$$\begin{cases} \hat{\lambda}^* = \frac{1}{n}\sum_{i=1}^n \sqrt{w_i} \\ \hat{\mu}^* = \frac{\frac{1}{n}\sum_{i=1}^n \sqrt{w_i} l_a(x_i,\hat{a})}{\mathbb{E}l_{\hat{a}}^2(x_i,\hat{a})} \end{cases}$$

Therefore, the estimated profit loss per capita is

$$\widehat{\Delta\Pi} = \frac{1}{n}\sum_{i=1}^n w_i - \left(\frac{1}{n}\sum_{i=1}^n \sqrt{w_i}\right)^2 - \frac{[\frac{1}{n}\sum_{i=1}^n \sqrt{w_i} l_a(x_i,\hat{a})]^2}{\mathbb{E}l_{\hat{a}}^2(x_i,\hat{a})}$$

and its asymptotic variance is

$$\begin{aligned} nAVar(\widehat{\Delta\Pi}) &= Var(w_i) + nAVar((\frac{1}{n}\sum_{i=1}^n \sqrt{w_i})^2) + nAVar(\frac{\hat{J}^2}{\hat{Z}}) + 2nACov(\frac{\hat{J}^2}{\hat{Z}}, (\frac{1}{n}\sum_{i=1}^n \sqrt{w_i})^2) \\ &\quad -2nACov(\hat{\mathbb{E}}w_i, (\frac{1}{n}\sum_{i=1}^n \sqrt{w_i})^2) - 2nACov(\hat{\mathbb{E}}w_i, \frac{\hat{J}^2}{\hat{Z}}) \\ &= Var(w_i) + 4\lambda^2 nAVar(\hat{\lambda}^*) + \frac{4J^2}{Z^2} nAVar(\hat{J}) - 4\frac{J^3}{Z^3} nACov(\hat{Z},\hat{J}) + \frac{J^4}{Z^4} nAVar(\hat{Z}) \\ &\quad +8\lambda\frac{J}{Z} nACov(\hat{J},\hat{\lambda}^*) - 4\lambda\frac{J^2}{Z^2} nACov(\hat{Z},\hat{\lambda}^*) - 4\frac{J}{Z} nACov(\hat{\mathbb{E}}w_i,\hat{J}) + 2\frac{J^2}{Z^2} nACov(\hat{\mathbb{E}}w_i,\hat{Z}) \\ &= Var(w_i) + 4\lambda^2 Var(\sqrt{w_i}) + \frac{4J^2}{Z^2}(Var((h-\lambda)l_a) + (\mathbb{E}(h-\lambda)l_{aT})^2 Z^{-1}) \\ &\quad -4\frac{J^3}{Z^3}[Cov((h-\lambda)l_a, l_a^2) + 2Z^{-1}\mathbb{E}(h-\lambda)l_{aT}\mathbb{E}l_a l_{aT}] + \frac{J^4}{Z^4} Var(l_a^2) \\ &\quad +8\lambda\frac{J}{Z} Cov((h-\lambda)l_a, h) - 4\lambda\frac{J^2}{Z^2} Cov(l_a^2, h) - 4\frac{J}{Z} Cov(w,(h-\lambda)l_a) + 2\frac{J^2}{Z^2} Cov(w, l_a^2) \end{aligned}$$

Under the null hypothesis,

$$\begin{aligned} nAVar(\widehat{\Delta\Pi}) &= Var(w_i) + 4\lambda^2 Var(\sqrt{w_i}) + \frac{4J^2}{Z^2}(Var((h-\lambda)l_a)) - 4\frac{J^3}{Z^3} Z^{-1}\mu\mathbb{E}l_a l_{aT}\mathbb{E}l_a l_{aT} \\ &\quad +4\lambda\frac{J^2}{Z^2} Cov(l_a^2, h) - 2\frac{J}{Z} Cov(w_i,(h-\lambda)l_a) \\ &= Var(w_i) + 4\lambda^2 Var(\sqrt{w_i}) + \frac{4J^2}{Z^2} Var((h-\lambda)l_a) \\ &\quad +4\lambda\frac{J^2}{Z^2} Cov(l_a^2, h) - 2\frac{J}{Z} Cov(w_i,(h-\lambda)l_a) \end{aligned}$$

$$
\begin{aligned}
nAVar(\widehat{\Delta\Pi}) &= 4[\mathbb{E}\sqrt{w_i}]^2 Var(\sqrt{w_i}) + 4[\mathbb{E}l_{\hat{a}}^2(x_i,\hat{a})]^2 Var(\sqrt{w_i}l_a(x_i,\hat{a})) \\
&\quad +4[\mathbb{E}\sqrt{w_i}]\mathbb{E}l_{\hat{a}}^2(x_i,\hat{a})Cov(\sqrt{w_i}l_a(x_i,\hat{a}),\sqrt{w_i})
\end{aligned}
$$

Analytically, in this case, relative to the piece-rate contract, the potentially optimal contract is given by,

$$
\lambda = \frac{\sqrt{\beta\pi a}}{2}; \mu = \frac{a^2\sqrt{\beta\pi}}{4\sqrt{a}}
$$

thus

$$
\Delta\Pi = \beta a(1 - \frac{1}{4}\pi - \frac{1}{16}\pi) \cong 0.018\beta a \cong 0.09
$$

the relative loss

$$
\frac{\Delta\Pi}{\Pi^*} = \frac{\beta(1-\frac{1}{4}\pi - \frac{1}{16}\pi)}{1-\frac{1}{4}\pi\beta - \frac{1}{16}\pi\beta}
$$

For $\beta < 0.9$, the relative loss is less than 15%. Q.E.D. ■

1.9.4 TS4. Proof of Proposition 4

Proof. The sample analogue of the Lagrange becomes:

$$
\begin{aligned}
&\hat{L}^*(\boldsymbol{\delta},\lambda^*,\mu^*) \\
= \; &\hat{\mathbb{E}}[w_i - h^{-1}(\lambda^* + \mu^*[\omega_a(x_i,\mathbf{T}) - \hat{\mathbb{E}}\omega_a(x_i,\mathbf{T})])] \\
&+\lambda^*[\hat{\mathbb{E}}u(h^{-1}(\lambda^* + \mu^*[\omega_a(x_i,\mathbf{T}) - \hat{\mathbb{E}}\omega_a(x_i,\mathbf{T})]),\gamma) - \hat{\mathbb{E}}u(w_i,\gamma)] \\
&+\mu^*\{\hat{\mathbb{E}}u(h^{-1}(\lambda^* + \mu^*[\omega_a(x_i,\mathbf{T}) - \hat{\mathbb{E}}\omega_a(x_i,\mathbf{T})]),\gamma)\omega_a(x_i,\mathbf{T}) - \hat{\mathbb{E}}u(w,\gamma)[\omega_a(x_i,\mathbf{T}) - \hat{\mathbb{E}}\omega_a(x_i,\mathbf{T})]\},
\end{aligned}
$$

associtated with two constraints:

$$
\begin{aligned}
\hat{\mathbb{E}}u(h^{-1}(\lambda^* + \mu^*[\omega_a(x_i,\mathbf{T}) - \hat{\mathbb{E}}\omega_a(x_i,\mathbf{T})]),\gamma) - \hat{\mathbb{E}}u(w_i,\gamma) &= 0 \\
\hat{\mathbb{E}}u(h^{-1}(\lambda^* + \mu^*[\omega_a(x_i,\mathbf{T}) - \hat{\mathbb{E}}\omega_a(x_i,\mathbf{T})]),\gamma)\omega_a(x_i,\mathbf{T}) - \hat{\mathbb{E}}u(w,\gamma)\omega_a(x_i,\mathbf{T}) &= 0.
\end{aligned}
$$

and the first order conditions can be replace by the empirical moment conditions.

Suppose γ should solve the first order condition:

$$\begin{aligned}\frac{dL^*(\gamma,\lambda^*,\mu^*)}{d\gamma} &= \frac{\delta L}{\delta s}\frac{\partial s^*}{\partial \delta} + \lambda^*(\gamma)[\int_{w\geq \underline{s}} \frac{\partial}{\partial\gamma}u(s^*(x,\gamma),\gamma)f(x,a)dx - \int_{w\geq \underline{s}}\frac{\partial}{\partial\gamma}u(w,\gamma)f(x,a)dx] \\ &+\mu^*(\gamma)[\int_{w\geq \underline{s}} \frac{\partial}{\partial\gamma}u(s^*(x,\gamma),\gamma)f_a(x,a^*(\gamma))dx - \int_{w\geq \underline{s}}\frac{\partial}{\partial\gamma}u(w,\gamma)f_a(x,a)dx] \\ &+\frac{\partial\lambda^*(\gamma)}{\partial\gamma}[\int_{w\geq \underline{s}} u(s^*(x,\gamma),\gamma)f(x,a)dx - \int_{w\geq \underline{s}}u(w,\gamma)f(x,a)dx] \\ &+\frac{\partial\mu^*(\gamma)}{\partial\gamma}[\int_{w\geq \underline{s}} u(s^*(x,\gamma),\gamma)f_a(x,a^*(\gamma))dx - \int_{w\geq \underline{s}}u(w,\gamma)f_a(x,a)dx] \\ &= 0,\end{aligned}$$

by the envelop theorem, and note that $\lambda^*(\gamma)+\mu^*(\gamma)\frac{f_a(x,a^*(\gamma))}{f(x,a^*(\gamma))} = \frac{1}{u'(s^*)}$, which results in,

$$\begin{aligned}0 &= \lambda^*(\gamma)[\int_{w\geq \underline{s}} \frac{\partial}{\partial\gamma}u(s^*(x,\gamma),\gamma)f(x,a)dx - \int_{w\geq \underline{s}}\frac{\partial}{\partial\gamma}u(w,\gamma)f(x,a)dx] \\ &+\mu^*(\gamma)[\int_{w\geq \underline{s}} \frac{\partial}{\partial\gamma}u(s^*(x,\gamma),\gamma)f_a(x,a^*(\gamma))dx - \int_{w\geq \underline{s}}\frac{\partial}{\partial\gamma}u(w,\gamma)f_a(x,a)dx] \\ &= \int_{w\geq \underline{s}} \frac{[\frac{\partial}{\partial\gamma}u(s^*,\gamma) - \frac{\partial}{\partial\gamma}u(w,\gamma)]}{\frac{\partial}{\partial s}u(s^*,\gamma)}f(x,a)dx. \end{aligned} \tag{1.34}$$

Together with the constraint (15) and (16), we have $r+2$ equations to solve the same number of unknowns. For convenience, denote $\zeta = (\lambda,\mu,\boldsymbol{\delta}')'$ and,

$$\Psi(\zeta,\mathbf{T}) = \begin{bmatrix} \int_{w\geq \underline{s}} u(s^*(x,\gamma),\gamma)l_a(x,\mathbf{T})f(x,a)dx - \int_{w\geq \underline{s}} u(w,\gamma)l_a(x,\mathbf{T})f(x,a)dx \\ \int_{w\geq \underline{s}} u(s^*(x,\gamma),\gamma)f(x,a)dx - \int_{w\geq \underline{s}} u(w,\gamma)f(x,a)dx \\ \int_{w\geq \underline{s}} \frac{[\frac{\partial}{\partial\gamma}u(s^*(x,\gamma),\gamma) - \frac{\partial}{\partial\gamma}u(w,\gamma)]}{\frac{\partial}{\partial s}u(s^*(x,\gamma),\gamma)}f(x,a)dx \\ \int_{w\geq \underline{s}} u(h^{-1}(\lambda^*+\mu^*[\omega_a(x_i,\mathbf{T}) - \hat{\mathbb{E}}\omega_a(x_i,\mathbf{T})]),\gamma)\omega_a(x_i,\mathbf{T}) - \int_{w\geq \underline{s}} u(w,\gamma)\omega_a(x_i,\mathbf{T}) \end{bmatrix}$$

Assume that $\sup_{\gamma\in\boldsymbol{\Gamma},\lambda>0,\mu>0}\left|\hat{\Psi}(\zeta,\mathbf{T}) - \Psi(\zeta,\mathbf{T})\right| \to^p 0$, therefore,

$$\sqrt{n}(\hat{\Psi}(\zeta,\mathbf{T}) - \Psi(\zeta,\mathbf{T})) \to^d (0,\boldsymbol{\Phi}),$$

where the i-th row and j-th column element of upper triangle of symmetric matrix $\boldsymbol{\Phi}$ can be found accordingly. Based on the above formula, we have

$$\sqrt{n}(\hat{\zeta} - \zeta) = -\sqrt{n}[\Psi_{\zeta'}(\bar{\zeta}, \bar{\mathbf{T}})]^{-1}(\hat{\Psi}(\zeta, \mathbf{T}) + \Psi_{\mathbf{T}'}(\bar{\zeta}, \bar{\mathbf{T}})(\hat{\mathbf{T}} - \mathbf{T})),$$

therefore, we have

$$\sqrt{n}(\hat{\zeta} - \zeta) \rightarrow^d (0, \Psi_{\zeta'}^{-1}\mathbf{\Phi}\Psi_{\zeta}^{-1} + \Psi_{\zeta'}^{-1}\Psi_{\mathbf{T}'}\mathbf{Z}^{-1}\Psi_{\mathbf{T}}\Psi_{\zeta}^{-1}).$$

Therefore, for every fixed x,

$$\sqrt{n}(\hat{s}^*(x, \hat{\gamma}) - s^*(x, \gamma)) \rightarrow^d \mathcal{N}(0, nAVar(\hat{s}^*(x, \hat{\gamma})))$$

where

$$\begin{aligned} & nAVar(\hat{s}^*(x, \hat{\gamma})) \\ = \; & \frac{\partial s^*(x, \gamma)}{\partial \zeta'}[\Psi_{\zeta'}^{-1}\mathbf{\Phi}\Psi_{\zeta}^{-1} + \Psi_{\zeta'}^{-1}\Psi_{\mathbf{T}'}\mathbf{Z}^{-1}\Psi_{\mathbf{T}}\Psi_{\zeta}^{-1}]\frac{\partial s^*(x, \gamma)}{\partial \zeta} + \mu^2\mathbb{E}[\frac{1}{h_s(s^*(x, \gamma))}l_{a\mathbf{T}'}]\mathbf{Z}^{-1}\mathbb{E}[\frac{1}{h_s(s^*(x, \gamma))}l_{a\mathbf{T}}]. \end{aligned}$$

Finally, we have

$$\sqrt{n}(\widehat{\Delta\Pi^k} - \Delta\Pi^k) \rightarrow^d \mathcal{N}\left(0, nAVar(\widehat{\Delta\Pi^k})\right),$$

where

$$nAVar(\widehat{\Delta\Pi^k}) = Var(w_i) + Var(s^*(x_i, \gamma)) + \mathbb{E}\frac{\partial s^*(x_i, \gamma)}{\partial \zeta'}nAVar(\hat{\zeta})\mathbb{E}\frac{\partial s^*(x_i, \gamma)}{\partial \zeta}.$$

To justify $M > 0$ and $K > 0$, the asymptotic variance of $\hat{\mathbb{E}}u(\hat{s}^*(x_i))g(x_i, \hat{\mathbf{T}})$ can be found by follows.

$$nAVar(\hat{\mathbb{E}}u(\hat{s}^*(x_i))g(x_i, \hat{\mathbf{T}})) = Var(u(s^*)g) + [\mathbb{E}u(s^*)g_{T'}]\mathbf{Z}^{-1}[\mathbb{E}u(s^*)g_T] + nAVar(\hat{\mathbb{E}}u'(s^*)(\hat{s}^* - s^*)g),$$

where

$$
\begin{aligned}
& nAVar(\hat{\mathbb{E}}u_w(s^*)(\hat{s}^* - s^*)g) \\
= \; & (\mathbb{E}\frac{u_w(s^*)g}{h_w(s^*)})^2 nAVar(\hat{\lambda}^*) + (\mathbb{E}\frac{u_w(s^*)gl_a}{h_w(s^*)})^2 nAVar(\hat{\mu}^*) \\
& + \mu^{*2}\mathbb{E}\left(\frac{u_w(s^*)gl_{a\mathbf{T}'}}{h_w(s^*)}\right)\mathbf{Z}^{-1}\mathbb{E}\left(\frac{u_w(s^*)gl_{a\mathbf{T}}}{h_w(s^*)}\right) + 2nACov(\hat{\lambda}^*, \hat{\mu}^*)(\mathbb{E}\frac{u_w(s^*)g}{h_w(s^*)})(\mathbb{E}\frac{u_w(s^*)gl_a}{h_w(s^*)}) \\
& + 2(\mu^*\mathbb{E}\frac{u_w(s^*)g}{h_w(s^*)}\mathbb{E}\frac{u_w(s^*)g}{h_w(s^*)}l_{a\mathbf{T}'})nACov(\hat{\lambda}^*, \hat{\mathbf{T}}^*) + 2(\mu^*\mathbb{E}\frac{u_w(s^*)gl_a}{h_w(s^*)}\mathbb{E}\frac{u_w(s^*)gl_a}{h_w(s^*)}l_{a\mathbf{T}'})nACov(\hat{\mu}^*, \hat{\mathbf{T}}^*)
\end{aligned}
$$

Therefore, $nAVar(\hat{M}^*(\hat{\mathbf{T}}))$ can be computed by sending $g(x_i, \hat{\mathbf{T}}) = l_a(x_i, \hat{\mathbf{T}})$, and $nAVar(\hat{M}^*(\hat{\mathbf{T}}))$ can be computed by sending $g(x_i, \hat{\mathbf{T}}) = l_a^2(x_i, \hat{\mathbf{T}}) + l_{aa}(x_i, \hat{\mathbf{T}})$. ■

1.9.5 TS5. Proof of Proposition 6

Proof. There are two types of test we can do. If γ_0 is not the boundary point, then we can directly test the first order condition based on γ_0, which is the following hypothesis test:

$$H_{0\gamma} : V_\gamma^*(\gamma_0) = 0$$

where $V_\gamma^*(\gamma_0)$ can be estimated by the empirical analogue

$$\widehat{V_\gamma^*}(\hat{\gamma}_0) = \Pr(w \geq \underline{s})\hat{\mathbb{E}}_{w \geq \underline{s}}\frac{\frac{\partial}{\partial\gamma}u(w_i, \hat{\gamma}_0)}{\frac{\partial}{\partial w}u(w_i, \hat{\gamma}_0)} + \Pr(w < \underline{s})\frac{\partial u(\underline{s}, \hat{\gamma}_0)}{\partial\hat{\gamma}_0}\hat{\mathbb{E}}_{w<\underline{s}}(\hat{\lambda} + \hat{\mu}l_a(x, \hat{a}))$$

with $\Pr(w \geq \underline{s}) = 1(w \geq \underline{s})$ and $\Pr(w < \underline{s}) = 1(w < \underline{s})$.

The asymptotic distribution of $\widehat{V_\gamma^*}(\hat{\gamma}_0)$ can be calculated analytically or by bootstrapping. When $f(x, T)$ is functionally specified, we can estimate $\Pr(w \geq \underline{s})$ by integration

$$\Pr(w \geq \underline{s}) = F(s^{*-1}(\underline{s}), T) = \int_{s^{*-1}(\underline{s})}^{\infty} f(x, T)dx$$

For convenience, denote $x_0 = s^{*-1}(\underline{s})$, $\widehat{\mathcal{M}(\hat{\lambda},\hat{\mu},\hat{\gamma}_0,\hat{a})} = \frac{\partial u(\underline{s},\hat{\gamma}_0)}{\partial\hat{\gamma}_0}\hat{\mathbb{E}}_{w<\underline{s}}(\hat{\lambda}+\hat{\mu}l_a(x,\hat{a}))$, therefore,

$$\begin{aligned} nAVar(\widehat{V_\gamma^*}(\hat{\gamma}_0)) &= nAVar((1-F(x_0,\hat{T}))\hat{\mathbb{E}}_{w\geq \underline{s}}\frac{\frac{\partial}{\partial\gamma}u(w_i,\hat{\gamma}_0)}{\frac{\partial}{\partial w}u(w_i,\hat{\gamma}_0)}) + nAVar(F(x_0,\hat{T})\widehat{\mathcal{M}(\hat{\lambda},\hat{\mu},\hat{\gamma}_0,\hat{a})}) \\ &\quad +2nACov((1-F(x_0,\hat{T}))\hat{\mathbb{E}}_{w\geq \underline{s}}\frac{\frac{\partial}{\partial\gamma}u(w_i,\hat{\gamma}_0)}{\frac{\partial}{\partial w}u(w_i,\hat{\gamma}_0)}, F(x_0,\hat{T})\widehat{\mathcal{M}(\hat{\lambda},\hat{\mu},\hat{\gamma}_0,\hat{a})}) \end{aligned}$$

Note that

$$\begin{aligned} & nAVar((1-F(x_0,\hat{T}))\hat{\mathbb{E}}_{w\geq \underline{s}}\frac{\frac{\partial}{\partial\gamma}u(w_i,\hat{\gamma}_0)}{\frac{\partial}{\partial w}u(w_i,\hat{\gamma}_0)}) \\ =\; & (1-F)^2 Var_{w\geq \underline{s}}(\frac{u_\gamma(w_i,\gamma_0)}{u_w(w_i,\gamma_0)}) + (1-F)^2\mathbb{E}_{w\geq \underline{s}}\frac{\partial}{\partial\gamma'}\left(\frac{u_\gamma(w_i,\gamma_0)}{u_w(w_i,\gamma_0)}\right)\Gamma\mathbb{E}_{w\geq \underline{s}}\frac{\partial}{\partial\gamma}\left(\frac{u_\gamma(w_i,\gamma_0)}{u_w(w_i,\gamma_0)}\right) \\ & -2(1-F)\mathbb{E}_{w\geq \underline{s}}\frac{u_\gamma(w_i,\gamma_0)}{u_w(w_i,\gamma_0)}\mathbb{E}_{w\geq \underline{s}}\frac{\partial}{\partial\gamma'}\left(\frac{u_\gamma(w_i,\gamma_0)}{u_w(w_i,\gamma_0)}\right)\sigma_{\gamma T'}F_T(x_0,T) \\ & +F_{T'}\mathbf{Z}^{-1}F_T\mathbb{E}_{w\geq \underline{s}}\frac{u_\gamma(w_i,\gamma_0)}{u_w(w_i,\gamma_0)}\mathbb{E}_{w\geq \underline{s}}\frac{u_{\gamma'}(w_i,\gamma_0)}{u_w(w_i,\gamma_0)} \end{aligned}$$

and

$$\begin{aligned} & nAVar(F(x_0,\hat{T})\widehat{\mathcal{M}(\hat{\lambda},\hat{\mu},\hat{\gamma}_0,\hat{a})}) \\ =\; & F^2 u_\gamma(\underline{s},\gamma_0)u_{\gamma'}(\underline{s},\gamma_0)\left[\begin{array}{c} Var_{w<\underline{s}}(\lambda+\mu l_a)+\sigma_\lambda^2+\mu^2\mathbb{E}_{w<\underline{s}}l_{aT'}\mathbf{Z}^{-1}\mathbb{E}_{w<\underline{s}}l_{aT}+\left(\mathbb{E}_{w<\underline{s}}l_a\right)^2\sigma_\mu^2 \\ +2\sigma_{\lambda T}\mathbb{E}_{w<\underline{s}}l_{aT'}+2\sigma_{\lambda\mu}\mathbb{E}_{w<\underline{s}}l_a+2\mu\sigma_{\mu T}\mathbb{E}_{w<\underline{s}}l_{aT'}\mathbb{E}_{w<\underline{s}}l_a \end{array}\right] \\ & +2F\mathbb{E}_{w<\underline{s}}(\lambda+\mu l_a)u_\gamma(\underline{s},\gamma_0)\left[\begin{array}{c} \sigma_{\lambda T}F_{T'}u_\gamma(\underline{s},\gamma_0)+u_{\gamma\gamma'}(\underline{s},\gamma_0)\sigma_{\lambda\gamma}F+\mu\mathbb{E}_{w<\underline{s}}l_{aT'}\mathbf{Z}^{-1}F_{T'}u_\gamma(\underline{s},\gamma_0) \\ \mu F u_{\gamma\gamma'}(\underline{s},\gamma_0)\sigma_{\gamma T'}\mathbb{E}_{w<\underline{s}}l_{aT}+\sigma_{\mu T}F_{T'}\mathbb{E}_{w<\underline{s}}l_a+u_{\gamma\gamma'}(\underline{s},\gamma_0)\sigma_{\mu\gamma}F\mathbb{E}_{w<\underline{s}}l_a \end{array}\right] \\ & +(\lambda+\mu\mathbb{E}_{w<\underline{s}}l_a)^2[F_{T'}\mathbf{Z}^{-1}F_T u_\gamma(\underline{s},\gamma_0)u_{\gamma'}(\underline{s},\gamma_0)+F^2u_{\gamma\gamma'}(\underline{s},\gamma_0)\Gamma u_{\gamma'\gamma}(\underline{s},\gamma_0)+2Fu_{\gamma\gamma'}(\underline{s},\gamma_0)\sigma_{\gamma T'}F_{T'}] \end{aligned}$$

and

$$\begin{aligned} & 2nACov((1-F(x_0,\hat{T}))\hat{\mathbb{E}}_{w\geq \underline{s}}\frac{\frac{\partial}{\partial\gamma}u(w_i,\hat{\gamma}_0)}{\frac{\partial}{\partial w}u(w_i,\hat{\gamma}_0)}, F(x_0,\hat{T})\widehat{\mathcal{M}(\hat{\lambda},\hat{\mu},\hat{\gamma}_0,\hat{a})}) \\ =\; & 2F(1-F)u_\gamma(\underline{s},\gamma_0)\left[\begin{array}{c} \sigma_{\lambda T}F_{T'}u_\gamma(\underline{s},\gamma_0)+u_{\gamma\gamma'}(\underline{s},\gamma_0)\sigma_{\lambda\gamma}F+\mu\mathbb{E}_{w<\underline{s}}l_{aT'}\mathbf{Z}^{-1}F_{T'}u_\gamma(\underline{s},\gamma_0) \\ \mu F u_{\gamma\gamma'}(\underline{s},\gamma_0)\sigma_{\gamma T'}\mathbb{E}_{w<\underline{s}}l_{aT}+\sigma_{\mu T}F_{T'}\mathbb{E}_{w<\underline{s}}l_a+u_{\gamma\gamma'}(\underline{s},\gamma_0)\sigma_{\mu\gamma}F\mathbb{E}_{w<\underline{s}}l_a \end{array}\right] \end{aligned}$$

Particularly, when $\underline{s}$ is not effective, then the second item $\Pr(w<\underline{s})\frac{\partial u(\underline{s},\hat{\gamma}_0)}{\partial\hat{\gamma}_0}\hat{\mathbb{E}}_{w<\underline{s}}(\hat{\lambda}+\hat{\mu}l_a(x,\hat{a}))$

can be neglected, the asymptotic distribution is much simpler, which is the following:

$$nAVar(\widehat{V_\gamma^*}(\hat{\gamma}_0)) = Var_{w\geq s}(\frac{u_\gamma(w_i,\gamma_0)}{u_w(w_i,\gamma_0)} + \mathbb{E}_{w\geq s}\frac{\partial}{\partial\gamma'}\left(\frac{u_\gamma(w_i,\gamma_0)}{u_w(w_i,\gamma_0)}\right)\Gamma\mathbb{E}_{w\geq s}\frac{\partial}{\partial\gamma}\left(\frac{u_\gamma(w_i,\gamma_0)}{u_w(w_i,\gamma_0)}\right)$$

Because both $\widehat{V_\gamma^*}(\hat{\gamma}_0)$ and $nAVar(\widehat{V_\gamma^*}(\hat{\gamma}_0))$ do not exist[25], we can use alternative approach to test whether $\hat{\rho}(\hat{\gamma}_0)$ is significantly close to 1 under $\hat{\gamma}_0$, solved by $\widehat{V_\gamma^*}(\hat{\gamma}_0) = 0$ (For example, for CRRA utility, we solve $\hat{\gamma} = \frac{\hat{\mathbb{E}}w_i}{\hat{\mathbb{E}}w_i \log w_i}$).

First of all, there exist only a unique γ such that $V(\gamma)$ is maximized. To see that, suppose that there is another $\tilde{\gamma}$ also maximizing $V(\gamma)$, then both γ and $\tilde{\gamma}$ will satisfying the first order condition (1), this contradicts with the uniqueness of optimal contract.

Now suppose that the solution γ^* solves the first order condition (20). Assume that

$$\sup_{\gamma\in\Gamma}\left|\hat{\mathbb{E}}\frac{\frac{\partial}{\partial\gamma}u(s^*(x_i),\gamma)}{\frac{\partial}{\partial w}u(s^*(x_i),\gamma)} - \mathbb{E}\frac{\frac{\partial}{\partial\gamma}u(s^*(x_i),\gamma)}{\frac{\partial}{\partial w}u(s^*(x_i),\gamma)}\right| \to^p 0$$

therefore,

$$\sqrt{n}(\hat{\gamma}^* - \gamma^*) = -\sqrt{n}\left[\hat{\mathbb{E}}(\frac{\partial}{\partial\gamma'}\frac{u_\gamma(w_i,\bar{\gamma})}{u_w(w_i,\bar{\gamma})})\right]^{-1}\hat{\mathbb{E}}\frac{u_\gamma}{u_w}$$

therefore,

$$\sqrt{n}(\hat{\gamma}^* - \gamma^*) \to^d N(0,\left[\mathbb{E}(\frac{\partial}{\partial\gamma'}\frac{u_\gamma}{u_w})\right]^{-1} Var(\frac{u_\gamma}{u_w})\left[\mathbb{E}(\frac{\partial}{\partial\gamma}\frac{u_\gamma}{u_w})\right]^{-1}).$$

Based on the above formula, $\sqrt{n}(\hat{\rho}(\hat{\gamma}^*) - \rho(\gamma^*)) \to^d \mathcal{N}(0, nAVar(\hat{\rho}(\hat{\gamma}^*)))$, and the asymptotic distribution of $\hat{\rho}(\hat{\gamma}^*)$ under the null hypothesis is,

$$nAVar(\hat{\rho}) = \frac{1}{4QZ}\left[\begin{array}{c} 4AVar(\hat{J}) + \frac{J^2}{Q^2}AVar(\hat{Q}) + (\frac{J}{Z})^2 AVar(\hat{Z}) \\ -4\frac{J}{Q}ACov(\hat{J}-J,\hat{Q}-Q) - 4\frac{J}{Z}ACov(\hat{J}-J,\hat{Z}-Z) \\ +2\frac{J^2}{QZ}ACov(\hat{Q}-Q,\hat{Z}-Z)\end{array}\right];$$

[25] For example when $u(w) = \frac{1}{\gamma}w^\gamma$, we can test whether $\widehat{V_\gamma^*}(\hat{\gamma}_0) = \frac{1}{\hat{\gamma}_0}(\frac{1}{\hat{\gamma}_0}\hat{\mathbb{E}}w_i + \hat{\mathbb{E}}w_i \log w_i) = 0$. The asymptotic distribution of $\sqrt{n}[\frac{1}{\hat{\gamma}_0}\hat{\mathbb{E}}w_i + \hat{\mathbb{E}}w_i \log w_i]$ is

$$nAVar(\frac{1}{\hat{\gamma}_0}\hat{\mathbb{E}}w_i + \hat{\mathbb{E}}w_i \log w_i) = \frac{1}{\gamma_0^2}(Var(w_i) + (\mathbb{E}w_i)^2\Gamma) + Var(w_i \log(w_i)) - \frac{2}{\gamma_0}Cov(w_i, w_i \log w_i)$$

Although when $\gamma_0 \to 0$, $\widehat{V_\gamma^*}(\hat{\gamma}_0) \to \infty$ and $nAVar(\frac{1}{\hat{\gamma}_0}\hat{\mathbb{E}}w_i + \hat{\mathbb{E}}w_i \log w_i) \to \infty$, we have $\frac{\widehat{V_\gamma^*}(\hat{\gamma}_0)}{\sqrt{nAVar(\widehat{V_\gamma^*}(\hat{\gamma}_0))}} \to^p \frac{\mathbb{E}w_i}{\sqrt{Var(w_i)+(\mathbb{E}w_i)^2\Gamma}}$ as z-value.

or

$$\frac{n\hat{Q}\hat{Z}(1-\hat{\rho})^2}{\left[4\hat{\lambda}^2\left(\hat{\mathbb{E}}l_{a\mathbf{T}'}\hat{\mathbf{Z}}^{-1}\hat{\mathbb{E}}l_{a\mathbf{T}}+\hat{Z}\right)+2\hat{Z}\hat{\mathbb{E}}h_{\gamma'}\hat{\Gamma}\hat{\mathbb{E}}h_{\gamma}-4\hat{\mathbb{E}}l_a h_{\gamma'}\hat{\Gamma}\hat{\mathbb{E}}h_{\gamma}l_a\right]} \to^d \chi^2_{r+t+1} \quad \text{if } \rho = 1;$$

where $\hat{\Gamma} = \left[\mathbb{E}(\frac{\partial}{\partial\gamma'}\frac{u_\gamma}{u_w})\right]^{-1} Var(\frac{u_\gamma}{u_w}) \left[\mathbb{E}(\frac{\partial}{\partial\gamma}\frac{u_\gamma}{u_w})\right]^{-1}$. ■

1.9.6 TS6. Proof of Proposition 7

Proof. The sample analogue of the first order condition is,

$$\hat{V}^*_\theta = \hat{\mathbb{E}}(x_i - s^*(x_i))l_\theta(x_i,\hat{\mathbf{T}}) + \hat{\lambda}\hat{\mathbb{E}}u(s^*(x_i),\hat{\gamma}_0)l_\theta(x_i,\hat{\mathbf{T}}) + \hat{\mu}\hat{\mathbb{E}}u(s^*(x_i),\hat{\gamma}_0)[l_{\theta a}(x_i,\hat{\mathbf{T}}) + l_a(x_i,\hat{\mathbf{T}})l_\theta(x_i,\hat{\mathbf{T}})].$$

Note that

$$\begin{aligned} & nAVar(\hat{\lambda}\hat{\mathbb{E}}u(s^*(x_i),\hat{\gamma}_0)l_\theta(x_i,\hat{\mathbf{T}})) \\ = \; & \lambda^2 Var(ul_\theta) + \lambda^2\{\mathbb{E}u_{\gamma'}l_\theta\Gamma\mathbb{E}u_\gamma l_\theta + \mathbb{E}ul_{\theta T'}\mathbf{Z}^{-1}\mathbb{E}ul_{\theta T} + 2[\mathbb{E}u_{\gamma'}l_\theta]\sigma_{\gamma T'}\mathbb{E}ul_{\theta T}\} \\ & +2\lambda\mathbb{E}ul_\theta\{\mathbb{E}u_{\gamma'}l_\theta\sigma_{\lambda\gamma} + \mathbb{E}ul_{\theta T'}\sigma_{\lambda T}\} + (\mathbb{E}ul_{\theta'})\, nAVar(\hat{\lambda})\,(\mathbb{E}ul_\theta) \end{aligned}$$

where $\sigma_{\lambda T} = \mathbb{E}h_{\gamma'}\sigma_{\gamma T'}$.

And,

$$\begin{aligned} & nAVar(\hat{\mu}\hat{\mathbb{E}}u(s^*(x_i),\hat{\gamma}_0)[l_{\theta a}(x_i,\hat{\mathbf{T}}) + l_a(x_i,\hat{\mathbf{T}})l_\theta(x_i,\hat{\mathbf{T}})]) \\ = \; & \mu^2 Var(u(l_{\theta a}+l_a l_\theta)) + \mu^2 \begin{pmatrix} \mathbb{E}u_{\gamma'}(l_{\theta a}+l_a l_\theta)\Gamma\mathbb{E}u_\gamma(l_{\theta a}+l_a l_\theta) \\ +\mathbb{E}u(l_{\theta aT'}+l_{aT'}l_\theta+l_a l_{\theta T'})\mathbf{Z}^{-1}\mathbb{E}u(l_{\theta aT'}+l_{aT'}l_\theta+l_a l_{\theta T'}) \\ +2[\mathbb{E}u_{\gamma'}(l_{\theta a}+l_a l_\theta)]\sigma_{\gamma T'}\mathbb{E}u(l_{\theta aT'}+l_{aT'}l_\theta+l_a l_{\theta T'}) \end{pmatrix} \\ & +2\mu\mathbb{E}u(l_{\theta a}+l_a l_\theta)\{\mathbb{E}u_{\gamma'}(l_{\theta a}+l_a l_\theta)\sigma_{\mu\gamma} + \mathbb{E}u(l_{\theta aT'}+l_{aT'}l_\theta+l_a l_{\theta T'})\sigma_{\mu T}\} \\ & +(\mathbb{E}u(l_{\theta a}+l_a l_\theta))\, nAVar(\hat{\mu})\,(\mathbb{E}u(l_{\theta a}+l_a l_\theta))' \end{aligned}$$

where

$$\sigma_{\mu\gamma} = nACov(\hat{\mu},\hat{\gamma}') = \frac{1}{Z}\Gamma\mathbb{E}h_\gamma l_a + \frac{1}{Z}\sigma_{\gamma T'}\mathbb{E}hl_{aT} - 2\frac{J}{Z^2}\sigma_{\gamma T'}\mathbb{E}l_a l_{aT}$$

and

$$\sigma_{\mu T'} = nACov(\hat{T}, \hat{\mu}) = \frac{1}{Z}[\mathbb{E}u_{\gamma'}(w_i, \gamma)l_a\sigma_{\gamma T'} + \mathbb{E}u(w_i, \gamma)l_{aT'}\mathbf{Z}^{-1}] - 2\frac{J}{Z^2}\mathbb{E}l_a l_{aT'}\mathbf{Z}^{-1}.$$

and

$$\begin{aligned} & nACov(\hat{\mathbb{E}}(x_i - s^*(x_i))l_\theta(x_i, \hat{\mathbf{T}}), \hat{\lambda}\hat{\mathbb{E}}u(s^*(x_i), \hat{\gamma}_0)l_\theta(x_i, \hat{\mathbf{T}})) \\ = \; & \lambda Cov((x_i - s^*(x_i))l_\theta, ul_\theta) + \lambda\mathbb{E}u_{\gamma'}l_\theta\sigma_{\gamma T'}\mathbb{E}(x_i - s^*(x_i))l_{\theta T} + \lambda\mathbb{E}(x_i - s^*(x_i))l_{\theta T'}\mathbf{Z}^{-1}\mathbb{E}ul_{\theta T} \\ & + \mathbb{E}(x_i - s^*(x_i))l_{\theta T'}\sigma_{\lambda T}\mathbb{E}ul_\theta \end{aligned}$$

and

$$\begin{aligned} & nACov(\hat{\mathbb{E}}(x_i - s^*(x_i))l_\theta(x_i, \hat{\mathbf{T}}), \hat{\mu}\hat{\mathbb{E}}u(s^*(x_i), \hat{\gamma}_0)[l_{\theta a}(x_i, \hat{\mathbf{T}}) + l_a(x_i, \hat{\mathbf{T}})l_\theta(x_i, \hat{\mathbf{T}})) \\ = \; & \mu Cov((x_i - s^*(x_i))l_\theta, u(l_{\theta a} + l_a l_\theta)) + \mu\mathbb{E}u_{\gamma'}(l_{\theta a} + l_a l_\theta)\sigma_{\gamma T'}\mathbb{E}(x_i - s^*(x_i))l_{\theta T} \\ & + \mu\mathbb{E}(x_i - s^*(x_i))l_{\theta T'}\mathbf{Z}^{-1}\mathbb{E}u(l_{\theta aT'} + l_{aT'}l_\theta + l_a l_{\theta T'}) + \mathbb{E}(x_i - s^*(x_i))l_{\theta T'}\sigma_{\mu T}\mathbb{E}u(l_{\theta a} + l_a l_\theta) \end{aligned}$$

and

$$\begin{aligned} & nACov(\hat{\lambda}\hat{\mathbb{E}}u(s^*(x_i), \hat{\gamma}_0)l_\theta(x_i, \hat{\mathbf{T}}), \hat{\mu}\hat{\mathbb{E}}u(s^*(x_i), \hat{\gamma}_0)[l_{\theta a}(x_i, \hat{\mathbf{T}}) + l_a(x_i, \hat{\mathbf{T}})l_\theta(x_i, \hat{\mathbf{T}})) \\ = \; & \lambda\mu Cov(ul_\theta, u(l_{\theta a} + l_a l_\theta)) + \lambda\mu\mathbb{E}(u_{\gamma'}l_\theta)\Gamma\mathbb{E}u_\gamma(l_{\theta a} + l_a l_\theta) + \lambda\mu[\mathbb{E}u_{\gamma'}l_\theta]\sigma_{\gamma T'}\mathbb{E}u(l_{\theta aT'} + l_{aT'}l_\theta + l_a l_{\theta T'}) \\ & + \lambda\mu\mathbb{E}u_\gamma(l_{\theta a} + l_a l_\theta)\sigma_{\gamma T'}\mathbb{E}ul_{\theta T'} + \lambda\mu\mathbb{E}ul_{\theta T'}\mathbf{Z}^{-1}\mathbb{E}u(l_{\theta aT} + l_{aT}l_\theta + l_a l_{\theta T}) \\ & + \mu\mathbb{E}ul_\theta[\mathbb{E}u_{\gamma'}(l_{\theta a} + l_a l_\theta)\sigma_{\lambda\gamma} + \mathbb{E}u(l_{\theta aT'} + l_{aT'}l_\theta + l_a l_{\theta T'})\sigma_{\lambda T}] \\ & + \lambda\mathbb{E}u(l_{\theta a} + l_a l_\theta)[\mathbb{E}(u_{\gamma'}l_\theta)\sigma_{\mu\gamma} + \mathbb{E}ul_{\theta T'}\sigma_{\mu T}] + \sigma_{\lambda\mu}\mathbb{E}ul_\theta\mathbb{E}u(l_{\theta a} + l_a l_\theta) \end{aligned}$$

where

$$\sigma_{\mu\lambda} = nACov(\hat{\mu}, \hat{\lambda}) = \frac{1}{Z}[Cov(h, hl_a) + \mathbb{E}h_{\gamma'}\Gamma\mathbb{E}h_\gamma l_a + \mathbb{E}h_{\gamma'}\sigma_{\gamma T'}\mathbb{E}hl_{aT}] - \frac{J}{Z^2}Cov(h, l_a^2) - 2\frac{J}{Z^2}\mathbb{E}h_{\gamma'}\sigma_{\gamma T'}\mathbb{E}l_a l_{aT}$$

Putting together, $\hat{V}_\theta^*$ has asymptotic distribution:

$$
\begin{aligned}
& nAVar(\hat{V}_\theta^*) \\
= \;& \lim_{n\to\infty} n\mathbb{E}(\hat{V}_\theta^* - \lim_{n\to\infty}\mathbb{E}\hat{V}_\theta^*)^2 \\
= \;& Var((x_i - s^*(x_i))l_\theta) + \mathbb{E}(x_i - s^*(x_i))l_{\theta T'}\mathbf{Z}^{-1}\mathbb{E}(x_i - s^*(x_i))l_{\theta T} + nAVar(\hat{\lambda}\hat{\mathbb{E}}u(s^*(x_i),\hat{\gamma}_0)l_\theta(x_i,\hat{\mathbf{T}})) \\
& + nAVar(\hat{\mu}\hat{\mathbb{E}}u(s^*(x_i),\hat{\gamma}_0)[l_{\theta a}(x_i,\hat{\mathbf{T}}) + l_a(x_i,\hat{\mathbf{T}})l_\theta(x_i,\hat{\mathbf{T}})]) \\
& + 2nACov(\hat{\mathbb{E}}(x_i - s^*(x_i))l_\theta(x_i,\hat{\mathbf{T}}), \hat{\lambda}\hat{\mathbb{E}}u(s^*(x_i),\hat{\gamma}_0)l_\theta(x_i,\hat{\mathbf{T}})) \\
& + 2nACov(\hat{\mathbb{E}}(x_i - s^*(x_i))l_\theta(x_i,\hat{\mathbf{T}}), \hat{\mu}\hat{\mathbb{E}}u(s^*(x_i),\hat{\gamma}_0)[l_{\theta a}(x_i,\hat{\mathbf{T}}) + l_a(x_i,\hat{\mathbf{T}})l_\theta(x_i,\hat{\mathbf{T}})) \\
& + 2nACov(\hat{\lambda}\hat{\mathbb{E}}u(s^*(x_i),\hat{\gamma}_0)l_\theta(x_i,\hat{\mathbf{T}}), \hat{\mu}\hat{\mathbb{E}}u(s^*(x_i),\hat{\gamma}_0)[l_{\theta a}(x_i,\hat{\mathbf{T}}) + l_a(x_i,\hat{\mathbf{T}})l_\theta(x_i,\hat{\mathbf{T}}))
\end{aligned}
$$

Q.E.D. ■

1.9.7 TS7. Testing the Adjoint Equation

Proof. To test the adjoint equation,

$$\widehat{AE} = \hat{\mathbb{E}}(x_i - w_i)l_a(x_i,\hat{T}) + \hat{\mu}\hat{\mathbb{E}}u(w_i,\hat{\gamma})(l_a^2(x_i,\hat{T}) + l_{aa}(x_i,\hat{T}))$$

note that

$$
\begin{aligned}
& nAVar((\hat{\mu}-\mu)\hat{\mathbb{E}}u(w_i,\hat{\gamma})(l_a^2(x_i,\hat{T}) + l_{aa}(x_i,\hat{T}))) \\
= \;& \lim_{n\to\infty} n\mathbb{E}((\hat{\mu}-\mu)\hat{\mathbb{E}}u(w_i,\hat{\gamma})(l_a^2(x_i,\hat{T}) + l_{aa}(x_i,\hat{T}))^2 - \lim_{n\to\infty}(\mathbb{E}(\hat{\mu}-\mu)\hat{\mathbb{E}}u(w_i,\hat{\gamma})(l_a^2(x_i,\hat{T}) + l_{aa}(x_i,\hat{T})))^2 \\
= \;& nAVar(\hat{\mu})\left(\mathbb{E}u(l_a^2 + l_{aa})\right)^2
\end{aligned}
$$

and,

$$\begin{aligned}
& nACov(\hat{\mathbb{E}}(x_i - w_i)l_a(x_i,\hat{T}), (\hat{\mu}-\mu)\hat{\mathbb{E}}u(w_i,\hat{\gamma})(l_a^2(x_i,\hat{T}) + l_{aa}(x_i,\hat{T}))) \\
= & \lim_{n\to\infty} n\mathbb{E}[\hat{\mathbb{E}}(x_i - w_i)l_a(x_i,\hat{T})(\hat{\mu}-\mu)\hat{\mathbb{E}}u(w_i,\hat{\gamma})(l_a^2(x_i,\hat{T}) + l_{aa}(x_i,\hat{T})))] \\
& - \lim_{n\to\infty} n\mathbb{E}(x_i - w_i)l_a(x_i,\hat{T})\mathbb{E}(\hat{\mu}-\mu)u(w_i,\hat{\gamma})(l_a^2(x_i,\hat{T}) + l_{aa}(x_i,\hat{T}))) \\
= & \lim_{n\to\infty} (n-1)\mathbb{E}(x_i - w_i)l_{aT'}(\hat{T}-T)(\hat{\mu}-\mu)\mathbb{E}u(w_j,\gamma)(l_a^2 + l_{aa}) \\
& + \lim_{n\to\infty} (n-1)\mathbb{E}(x_i - w_i)l_a\mathbb{E}(\hat{\mu}-\mu)(\hat{\gamma}-\gamma)'u_\gamma(w_j,\gamma)(l_a^2 + l_{aa}) \\
& + \lim_{n\to\infty} (n-1)\mathbb{E}(x_i - w_i)l_a\mathbb{E}(\hat{\mu}-\mu)(2l_al_{aT'} + l_{aaT'})(\hat{T}-T)u(w_j,\gamma)) \\
= & \mathbb{E}(x_i - w_i)l_{aT'}\sigma_{\mu T}\mathbb{E}u(w_j,\gamma)(l_a^2 + l_{aa}) + \mathbb{E}(x_i - w_i)l_a\mathbb{E}\sigma_{\mu\gamma'}u_\gamma(w_j,\gamma)(l_a^2 + l_{aa}) \\
& + \mathbb{E}(x_i - w_i)l_a\mathbb{E}u(w_j,\gamma)(2l_al_{aT'} + l_{aaT'}))\sigma_{\mu T}
\end{aligned}$$

where

$$\sigma_{\mu\gamma} = nACov(\hat{\mu},\hat{\gamma}') = \frac{1}{Z}\Gamma\mathbb{E}h_\gamma l_a + \frac{1}{Z}\sigma_{\gamma T'}\mathbb{E}hl_{aT} - 2\frac{J}{Z^2}\sigma_{\gamma T'}\mathbb{E}l_al_{aT}$$

and

$$\sigma_{\mu T'} = nACov(\hat{T},\hat{\mu}) = \frac{1}{Z}[\mathbb{E}u_{\gamma'}(w_i,\gamma)l_a\sigma_{\gamma T'} + \mathbb{E}u(w_i,\gamma)l_{aT'}\mathbf{Z}^{-1}] - 2\frac{J}{Z^2}\mathbb{E}l_al_{aT'}\mathbf{Z}^{-1}.$$

And

$$\begin{aligned}
& \mu nACov(\hat{\mathbb{E}}(x_i - w_i)l_a(x_i,\hat{T}), \hat{\mathbb{E}}u(w_i,\hat{\gamma})(l_a^2(x_i,\hat{T}) + l_{aa}(x_i,\hat{T})) - \mathbb{E}u(w_i,\gamma)(l_a^2 + l_{aa})) \\
= & \mu Cov((x_i - w_i)l_a, u(w_i,\gamma)(l_a^2 + l_{aa})) + \mu\mathbb{E}u_{\gamma'}(w_i,\gamma)(l_a^2 + l_{aa})\sigma_{\gamma T'}\mathbb{E}[(x_i - w_i)l_{aT}] \\
& + \mu\mathbb{E}[(x_i - w_i)l_{aT'}]\mathbf{Z}^{-1}\mathbb{E}u(w_i,\gamma)(2l_al_{aT} + l_{aaT})
\end{aligned}$$

and

$$
\begin{aligned}
& \mu nACov((\hat{\mu}-\mu)\hat{\mathbb{E}}u(w_i,\hat{\gamma})(l_a^2(x_i,\hat{T})+l_{aa}(x_i,\hat{T})), \hat{\mathbb{E}}u(w_i,\hat{\gamma})(l_a^2(x_i,\hat{T})+l_{aa}(x_i,\hat{T}))-\mathbb{E}u(w_i,\gamma)(l_a^2+l_{aa}) \\
= \;& \mu \lim_{n\to\infty} \mathbb{E}\{(\hat{\mu}-\mu)\hat{\mathbb{E}}u(w_i,\hat{\gamma})(l_a^2(x_i,\hat{T})+l_{aa}(x_i,\hat{T}))\hat{\mathbb{E}}[u(w_i,\hat{\gamma})(l_a^2(x_i,\hat{T})+l_{aa}(x_i,\hat{T}))-\mathbb{E}u(w_i,\gamma)(l_a^2+l_{aa}) \\
& +\mu \lim_{n\to\infty}(n-1)\mathbb{E}[(\hat{\mu}-\mu)u(w_i,\gamma)(l_a^2+l_{aa})(u(w_i,\gamma)(l_a^2+l_{aa})-\mathbb{E}u(w_i,\gamma)(l_a^2+l_{aa})] \\
& +\mu \lim_{n\to\infty}(n-1)\mathbb{E}(\hat{\mu}-\mu)\begin{bmatrix} u_{\gamma'}(l_a^2+l_{aa})(\hat{\gamma}-\gamma)+ \\ u(w_i,\gamma)(2l_a l_{aT'}+l_{aaT'})(\hat{T}-T) \end{bmatrix}(u(w_j,\gamma)(l_a^2+l_{aa})-\mathbb{E}u(l_a^2+l_{aa})) \\
& +\mu \lim_{n\to\infty}(n-1)\mathbb{E}(\hat{\mu}-\mu)u(w_i,\gamma)(l_a^2+l_{aa})[u_{\gamma'}(w_j,\gamma)(l_a^2+l_{aa})(\hat{\gamma}-\gamma)+u(w_i,\gamma)(2l_a l_{aT'}+l_{aaT'})(\hat{T}-T) \\
= \;& \mu\mathbb{E}u(w_i,\gamma)(l_a^2+l_{aa})\mathbb{E}u_{\gamma'}(w_j,\gamma)(l_a^2+l_{aa})\sigma_{\mu\gamma}+\mu\mathbb{E}u(w_i,\gamma)(l_a^2+l_{aa})\mathbb{E}u(w_j,\gamma)(2l_a l_{aT'}+l_{aaT'})[nACov(\hat{T}
\end{aligned}
$$

Putting all pieces together, we have

$$
\begin{aligned}
& nAVar(\widehat{AE}) \\
= \;& \lim_{n\to\infty} n\mathbb{E}(\widehat{AE}-\lim_{n\to\infty}\mathbb{E}\widehat{AE})^2 \\
= \;& nAVar((x_i-w_i)l_a)+\mathbb{E}(x_i-w_i)l_{aT'}\mathbf{Z}^{-1}\mathbb{E}(x_i-w_i)l_{aT}+nAVar(\hat{\mu})[\mathbb{E}u(l_a^2+l_{aa})]^2+\mu^2 nAVar(\hat{K}) \\
& +2nACov(\hat{\mathbb{E}}(x_i-w_i)l_a(x_i,\hat{T}),(\hat{\mu}-\mu)\hat{\mathbb{E}}u(w_i,\hat{\gamma})(l_a^2(x_i,\hat{T})+l_{aa}(x_i,\hat{T}))) \\
& +2\mu nACov(\hat{\mathbb{E}}(x_i-w_i)l_a(x_i,\hat{T}),[\hat{\mathbb{E}}u(w_i,\hat{\gamma})(l_a^2(x_i,\hat{T})+l_{aa}(x_i,\hat{T}))-\mathbb{E}u(w_i,\gamma)(l_a^2+l_{aa})]) \\
& +2\mu nACov((\hat{\mu}-\mu)\hat{\mathbb{E}}u(w_i,\hat{\gamma})(l_a^2(x_i,\hat{T})+l_{aa}(x_i,\hat{T})),\hat{\mathbb{E}}u(w_i,\hat{\gamma})(l_a^2(x_i,\hat{T})+l_{aa}(x_i,\hat{T}))-\mathbb{E}u(w_i,\gamma)(l_a^2+l
\end{aligned}
$$

Q.E.D. ■

Bibliography

[1] C. Ai, Semiparametric Maximum Likelihood Estimator, Econometrica, 65 (1997), pp 933-964.

[2] D. Ackerberg, M. Botticini, Endogenous matching and the empirical determinants of contract form, mimeo, Boston University (1999).

[3] D. Ackerberg, M. Botticini, Endogenous Matching and the Empirical Determinants of Contract Form", Journal of Political Economy, 110 (2002), pp 564-591.

[4] G. Casella, R. Berger, Statistical Inference, 2nd Ed. The Wadsworth Group, (2002).

[5] S. Chen, Rank Estimation of Transformation Models, Econometrica, 70 (2002), pp 1683-1697.

[6] P.A. Chiappori, Econometric models of insurance under asymmetric information, in G.Dionne, ed., Handbook of Insurance, Norht Holland, (2000).

[7] P.A. Chiappori, B. Salanie, Testing contract theory: a survey of some recent work, in Advances in Economics and Econometrics, vol 1, M. Dewatripont, L. Hansen and S. Turnovsky eds, Cambridge University Press. (2003)

[8] H. Dette, N. Neumeyer, and K. F. Pilz, A Simple Nonparametric Estimator of a Strictly Monotone Regression Function, Bernoulli Vol. 12, No. 3 (2006), 469-490.

[9] X. d'Haultfoeuille, P. Fevrier, Identification and Estimation of Incentive Problems: Adverse Selection. CREST-INSEE, Working Paper (2007).

[10] P.S Faynzilberg, P. Kumar, On the generalized principal-agent problem: Decomposition and existence results, Rev. Econ. Design (2000), 5, 23-58.

[11] R. Gibbons, M. Waldman, Executive compensation, in: O. Ashenfelter, D. Card (Eds.), Handbook of Labor Economics, vol. 3, North-Holland, Amsterdam, 1999.

[12] M. R. Haley, The Response of Worker Effort to Piece Rates: Evidence from the Midwest Logging Industry, J. Human Resources XXXVIII (4), (2003), 881-890.

[13] S. Grossman, O. Hart, An analysis of the principal–agent problem, Econometrica 51 (1983), 7–45.

[14] M. Harris, A. Raviv, Optimal incentive contracts with imperfect information, J. Econ. Theory 20 (1979) 231–259.

[15] M.F. Hellwig, K.M. Schmidt, Discrete-time approximation of the Holmstrom–Milgrom Brownian-motion model of intertemporal incentives provision, Econometrica 70 (2002) 2225–2264.

[16] J.L. Horowitz, Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable, Econometrica, 64 (1996) 103-137.

[17] B. Holmstrom, Moral hazard and observability, Bell J. Econ. 10 (1979) 74–91.

[18] B. Holmstrom, P. Milgrom, Aggregation and linearity in the provision of intertemporal incentives, Econometrica 55 (1987) 303–328.

[19] M. Jensen, K.J. Murphy, Performance pay and top-management incentives, J. Polit. Economy 98 (1990) 225–264.

[20] I. Jewitt, Justifying the first order approach to principal-agent problems, Econometrica, 56 (1988), No.5, 1177-1190.

[21] R. Ke, Identifying the Optimality of Contract without Knowing Functions: Moral Hazard and Statistic Inference II (Non-Parametric Identification), 2008, draft.

[22] S.K. Kim, Efficiency of an information system in an agent model, Econometrica, 63 (1995), No. 1, 89-102.

[23] E.P. Lazear, Performance pay and productivity, Amer. Econ. Rev. 90 (2000) 1346–1361.

[24] Q. Li and J. S. Racine, Nonparametric Econometrics: Theory and Practice, Princeton Unversity Press, 2007.

[25] J. Mirrlees, The optimal structure of authority and incentives within an organization, Bell J. Econ. 7 (1976) 105–131.

[26] J. Mirrlees, Information and Incentives: the economics of carrots and sticks, Econ. Journal 107(1997), 1311-29.

[27] J. Mirrlees, Theory of moral hazard and unobservable behavior: Part I, Rev. Econ. Studies, 66 (1999), 3-21.

[28] R.B. Myerson, Optimal coordination mechanisms in generalized principal-agent problems, Journal of Mathematical Economics, 11(1982), 67-81.

[29] W. Newey, Adaptive estimation of regression models via moment restrictions, J. of Econometrics, 38 (1988) pp301-339.

[30] W. Newey, Asymptotic theory for non-linear estimators, draft, MIT (2005).

[31] W. Rogerson, The first order approach to principle-agent problems, Econometrica, 53 (1985), No. 6, 1357-1368.

[32] H. Paarsch, and B. Shearer, The response of worker effort to piece rates: evidence from the British Columbia tree-planting industry, Journal of Human Resource, 34 (1999), 643-667.

[33] Paarsch, H and B. Shearer, Piece rates, fixed wages and incentive effects: statistical evidence from payroll records, International Economic Review., 41 (2000), 59-92.

[34] A. Pagan and A. Ullah, Non-parametric Econometrics, Cambridge University Press, 1999.

[35] C. Prendergast, The provision of incentives in firms, J. Econ. Lit. 37 (1999) 7–63.

[36] C. Prendergast, The tenuous tradeoff between risk and incentives, J. Polit. Economy 110 (2002) 1071–1102.

Chapter 2

A Fixed-Point Approach to Validate the First-Order Approach: Relaxation of Global Concavity

2.1 Introduction

It is well known that the sufficient conditions for the first order approach (FOA) to be valid could be very restrictive. The main concern is the conditions under which one can replace the incentive compatible (IC) constraint by the agent's first order condition (local IC constraint)[1]. One of the sufficient conditions is the global concavity of the agent's utility for every given contract, which is equivalent to the conditions assuring the convexity of the constraint set, in terms of the principal's optimization problem. In this spirit, based on Kuhn-Tucker's theorem, Rogerson's (1985) seminal idea is first to relax the principal' original problem futher, and find that the solution of the double relaxed problem is indeed within the original constraint set.

[1] Mirrlees (1971) provides a famous example that the first order condition for the agent's best response fails to be valid for the principal's program. He also shows an insightful example: if the likelihood ratio function is not bounded from below, a simple two-part tariff contract can arbitrarily closely approach the first best contract. The condition that the likelihood ratio function is bounded from below seems reasonable and will be met in many commonly used density functions, although the most commonly seen normal distribution fails this restriction. In terms of pragmatic application, we also can put a limit liability constraint on the agent's side, which means the principal cannot infinitely punish the agent when performance is bad. Given these two facts, the boundedness likelihood ratio is not the main concern (Kim, 1995).

As a result, he shows that FOA validates with convexity of the distribution function condition (CDFC) together with monotone likelihood ratio property (MLRP). Even though the latter seems to have a very intuitive intepretation in terms of statistics (Milgrom, 1982), the former is really restrictive so that many commonly used density functions fail. In order to remove this difficulty, Jewitt (1988) somewhat relaxes CDFC, keeping MLRP the same, at the cost of introducing restrictions on the agent's utility shape. That is, the absolute risk version measure should not decline too fast[2]. Jewitt's (1988) proof looks different from Rogerson's (1985), but they both require global concavity. Jewitt's (1988) results rely on a concavity-preserving transformation, so that the agent's expected utility is globally concave when the utility is concave in output.

This paper validates FOA based on a new but very intuitive approach. In contrast to Rogerson's original double-relaxed approach, we narrow the contract space first, directly searching for the existence of a solution within the contract class satisfying the Mirrless-Holmstrom characterization ($\mathcal{MH}$ contract). Therefore, we treat the classical moral hazard problem as a game between the principal and agent, and then we transfer the issue of validity to a problem of whether there is a fixed point regarding the agent's best response against a $\mathcal{MH}$ contract.

The main difference between the existing literature and the current approach is that the proof in the current approach does not rely on global concavity of the agent's utility. And global concavity is a special case of our approach. Instead, based on our approach, the FOA is valid even though the agent's utility is convex or even stranger, as long as the best response is continuous in some parameters. Using the current approach, we can unify Jewitt's (1988) and Rogerson's (1985) seminal results. As an application, we restore Jewitt's (1988) results to situations where the log likelihood ratio is not bounded from below, and/or the payment rule is not concave due to the existence of some limited liability constraint. We provide a set of conditions to assure the single-peakness of the agent's utility, though global concavity is not met. According to our proof, the FOA is valid for normal distribution or its monotone

[2]Several other authors also try to relax or generalize the first order approach, using some more technical tools such as rearrangement (Carlier and Dana, 2002, 2005), but the conclusion seems not to make a significant difference to the CDFC. One exception is Araujo and Moreira (2001), who use a generalized Lagrangian approach to deal with the situation where the FOA is not valid. Recently, Mirrlees and Zhou (2007) characterize the principal-agent model by enlarging the agent's choice space. However those papers do not justify the validity of the FOA, and in them, it remains unknown whether we could have a more friendly condition for the FOA to be valid.

tranformations (like log normal) with constant-relative-risk-aversion utility, if the action space is well-specified.

Meanwhile, we present a necessary and sufficient condition for FOA to be valid in a general non-separable utility environment, which also can be applied to a general bi-level non-linear optimization problem. It is well known that when the constraint set is non-convex, there exists a duality gap so that the FOA is not valid. Our approach provides a direct method to judge the validity by checking two conditions: (i) the action chosen by the agent should be a fixed point against the $\mathcal{MH}$ contract; (ii) the optimal action should maximize the Lagrangian under the $\mathcal{MH}$ contract associated with the multipliers being determined by the IR constraint and local IC constraint. It can be seen that a sufficient condition for FOA to be valid is the continuity of the best response correspondence so that the fixed point exists and any effort level can be a fixed point by adjusting the multipliers. Fortunately, the classical moral hazard problem has some good properties so that this continuities is met even in the general non-separable utility case. In addition, even the explicity analytical conclusion is not available, our approach will be very useful for numerical algorithm in practice, once the functions are known.

This paper thus is organized as follows. Section II presents the basic model and several properties of its characterization. Section III provides a fixed approach to validate FOA and prove the existence of solutions and offers two examples where the existing literature does not apply. Section IV deals with some applications to an additive output-generating process and the exponential family with existence of a sufficient statistic. Section V generalizes several theorems to situation where the agent's utility is non-separable and shows a necessary and sufficient condition for the validity of FOA. Section VI presents basic conclusions.

2.2 The Model

2.2.1 Characterizations

In a standard moral hazard problem setting (Holmstrom, 1979), there are a principal and a risk averse agent whose utility functions are $v(.) \in \mathcal{V}$ and $u(.) \in \mathcal{U}$ respectively. The principal could be risk neutral or risk averse[3]. The agent's output x is randomly distributed over region $\mathcal{X} \subset \mathbb{R}$, with an atomless probability density function (p.d.f.) $f(x,a)$, given the agent's effort $a \in \mathbb{A} \subset \mathbb{R}_+$, where $\mathbb{A}$ could be an open interval $[0,\infty)$ or a close interval $[0,\bar{a}]$. Throughout this paper, we assume $\mathbb{A} = [0,\infty)$, unless we point it out explicitly. And the relevance of the boundedness will be discussed later. We assume the support $\mathcal{X}$ does not depend on the effort level a and p.d.f. $f(x,a)$ is continuous and differentiable in a up to some appropriate order. The agent's effort is unobserved by the principal, and the principal makes a take-it-or-leave-it contract $w \in \mathcal{W}$ with $\mathcal{W}$ being a measurable compact functional space. We assume that the disutility of effort $c(a)$ is increasing as effort increases and separable from the monetary utility[4]. The principal solves the following optimization problem:

$$(P1) \quad \max_{\{a,w\}} \int v(x-w) f(x,a) dx$$

subject to the following individual rationality (IR) and incentive compatibility (IC) constraints for the agent,

$$\int [u(w) - c(a)] f(x,a) dx \geq \underline{U} \tag{IR}$$

$$\int [u(w) - c(a)] f(x,a) dx \geq \int [u(w) - c(\tilde{a})] f(x,\tilde{a}) dx, \ \forall a, \tilde{a} \in \mathbb{A}, \tag{IC}$$

where $\underline{U}$ is the outside reservation utility, and the choice set $\mathbb{A}$ implies that $a = 0$ could be a default choice by the agent.

Formally, we write the following assumptions explicitly.

A1: Agent is risk averse and the principal is risk neutral or risk averse;

A2: cost of effort is a strictly increasing function of effort;

A3: expected output is increasing and weakly concave in effort a, namely, (i) $\frac{\partial}{\partial a}\mathbb{E}x > 0$ and $\mathbb{E}x < \infty$ for any $a < \infty$, and (ii) $\frac{\partial^2}{\partial a^2}\mathbb{E}x \leq 0$;

A4: the score is well-defined everywhere, namely $\frac{\partial}{\partial a}\log f(x,a) > -\infty$ for any $x \in \mathcal{X}$;

A4′: or the payment is uniformly bounded from below, namely, $s(x) \geq \underline{s} > -\infty$ for any

[3] It is not necessary to assume that at least one of them is risk averse. What is necessary is $-\frac{v''(.)}{v'(.)} + (-\frac{u''(.)}{u'(.)}) > 0$, which means that the principal can be risk averse or risk taking.

[4] As Rogerson (1985) does, we always are able to reparameterize the cost of effort as $c(a) = a$, without loss of generality. We preserve this opportunity for the moment.

$x \in \mathcal{X}$.

Assumption A3, the concavity of the mean of output, is standard[5]. The first part A3-i is weaker than second order stochastic dominance (SOSD), and therefore much weaker than monotone likelihood ratio property (MLRP)[6]. A3-ii is weaker than the condition (2.10a) in Jewitt (1988), and thus much weaker than the convexity of the distribution function condition (CDFC).

A4 may be the most restrictive assumption, precluding the most familiar normal distribution. Under CDFC, A4 can be dropped if limited liability constraint A4$'$ holds. The economic intuition behind A4 or A4$'$ is that the principal cannot punish the agent by a negative infinite fine, even if the agent's performance is bad. However, without CDFC, replacing A4 by A4$'$ is a non-trivial task, because under A4$'$, Jewitt's (1988) condition does not apply since the payment rule is no longer concave (e.g., normal distribution). We will discuss this issue based on our new approach later.

The above assumptions are fairly general, which may not be sufficient to validate the first order approach. We will first present several important results under general conditions, then make some more specific assumptions to validate the first order approach based on the general findings.

Remark 2.1: *Under the above assumptions A1-A4 or A4$'$, the first best solution exists if at least one of the following conditions holds: (i) cost of effort is strictly convex; (ii) expected output is strictly concave; or (iii) action space is compact.*

2.2.2 The Properties of Optimal Contract

For convenience, use $U(w,a) = \int [u(w) - c(a)] f(x,a)dx$ to denote agent's utility, and use $a^{BR}(w) \in \arg\max_a U(w,a)$ to denote agent's best response to contract w, and denote $\mathbb{A}^{BR}(\mathcal{W})$ as all collections of best responses. Because the agent is always able to choose quit as one of his strategies, we define the implementation as follows.

[5] If the action space is compact, we do not need to specify any assumption about the distribution function. If the action space is not compact, $a \in \mathbb{R}_+$, then it is necessary to make some assumptions to assure the existence of the optimum.

[6] The first part also implies the identifiability of the distribution function, that is, $\mathbb{E}X$ is increasing in a implies there is positive measure x such that $f(x,a) \neq f(x,a')$ *for all* $a \neq a'$. This is very useful in terms of empirical identification.

Definition 2.1: *An effort level a is implemented by a contract w, if: (i) $a \in a^{BR}(w)$, and (ii) $U(w,a) \geq \underline{U}$.*

First of all, we show the first order condition is a necessary constraint for (P1), which means the solution (a, w) to (P1) must satisfy the following first order condition:

$$\int u(w) f_a(x,a)dx - c'(a) = 0. \tag{LIC}$$

We can prove it formally as follows.

Lemma 2.1: *Under assumptions A1-A3, and A4 or A4', suppose $a^* \in a^{BR}(w^*)$, then (i) IR constraint is binding; (ii) and LIC is a necessary condition for (P1).*

Proof. We check (i) first. By contradiction, suppose (w^*, a^*) is an optimal contract for which IR is not binding. Consider the contract $\tilde{w} = u^{-1}(u(w^*) - \varepsilon)$ for any constant $\varepsilon > 0$. We choose ε so that IR is binding. And note that $U(\tilde{w}, a) = U(w^*, a) + \varepsilon$, $\tilde{w}$ implements the same effort level as w^* does. Because $\varepsilon > 0$, the principal is strictly better off with $\tilde{w}$.

(ii) Because $U(w^*, a)$ is a continuous and differentiable function of $a \in [0, \infty)$ for any given w, the optimal happens at a point where the first order condition is satisfied unless the optimizer is unbounded. So it suffices to show that a^* is bounded. Note that for given w^*,

$$\underline{U} \leq U(w^*, a^*) \leq u(\mathbb{E}w^*) - c(a^*).$$

Because $\mathbb{E}v(x - w^*) \leq v(\mathbb{E}(x - w^*))$, as the solution to problem P1, $\mathbb{E}w^*$ is bounded by a linear transformation of $\mathbb{E}x$, therefore

$$u(\alpha\mathbb{E}x + \beta) - c(a^*) \geq \underline{U}$$

where α and β are two constants. Therefore $a^* \in a^{BR}(w^*)$ must be bounded. This implies that $U_a(w^*, a^*) = 0$ must hold. Q.E.D. ■

Remark 2.2: *If the action space is compact, namely, $a \in [0, \bar{a}]$, under assumptions A1-A4 or A4', then (i) LIC must hold if the principal is risk-neutral*[7]*; (ii) when the principal is risk-*

[7]Suppose $(w^*, \bar{a})$ is the optimal contract and $U_a(w^*, \bar{a}) > 0$, then according to the median value theorem, we can construct another contract $\tilde{w} = (1-\alpha)w^* + \alpha\mathbb{E}w^* - \epsilon(\alpha)$ with $\alpha \in (0,1)$ and $\epsilon(\alpha) > 0$ such that for α small enough $U_a(\bar{a}, \tilde{w}) = 0$. By Jensen's inequality, since $u(w)$ is concave, there also exists $\epsilon(\alpha) > 0$ such that

averse and LIC does not hold, then Borch's Rule $0 \le w^{*\prime} \le 1$ must hold for all x. The basic economic intuition from (i) is that for the risk-neutral principal, the optimal solution is always constrained by LIC. But for the risk-averse principal, LIC might not hold. Once the LIC does not hold, the incentive from risk sharing is enough so that Borch's Rule holds.

A straightforward corollary is that for any best response $a^* \in a^{BR}(w)$, the agent's utility with w fixed must be locally concave in a at a^*.

Corollary 2.1: *If conditions for lemma 1 hold, and if (a^*, w^*) is the solution to (P1), $U(w^*, a^*)$ must be locally concave in a at a^*. Formally, we have for all $a^* \in a^{BR}(w^*)$, $U_{aa}(w^*, a^*) \le 0$ and at least for one $a^* \in a^{BR}(w^*)$, $U_{aa}(w^*, a^*) < 0$.*

Proof. For any given w^*, since LIC is a necessary condition, $U_{aa}(w^*, a^*) \le 0$ must be satisfied since a^* is the best response. And since the optimality cannot happen at the boundary points, then there must exist some $a^* \in a^{BR}(w^*)$, then $U_{aa}(w^*, a^*) < 0$ is strict. Q.E.D. ■

When IC is relaxed by LIC, the problem (P1) becomes

$$(P2) \quad \max_{\{a,w\}} \int v(x-w) f(x,a) dx, \text{ s.t. IR and LIC.}$$

Use A to denote the set of (w, a) constrained by IR and IC, and use B to denote the one constrained by IR and LIC. Because the solution to LIC contains all true best responses satisfying LIC, $A \subset B$ because LIC relaxes the constraint so that (P2) adopts some solutions that might be not attainable by the principal under the original IC constraint.

We use $V_i(w, a)$ $(i = A, B)$ to denote the maximized value over constraint set i. To show that the FOA is valid, it is equavalent to show that the solution constrained by LIC is indeed the same as IC, i.e., $V_A(w, a) = V_B(w, a)$. Rogerson (1985) relaxes B even further, namely, $C \supset B$, and shows that the optimization constrained by C actually will happen in A, and therefore, $V_A(w, a) = V_B(w, a)$. The present paper uses a very different approach. We refine the solution set A so that the necessary conditions for the formula of optimal contract can be used to narrow the search for a valid solution.

In order to compare (P1) and (P2) in a more intuitive way, we rewrite the problem (P1)

$U(\tilde{w}, \bar{a}) = U(w^*, \bar{a})$. Therefore, $\bar{a} \in a^{BR}(\tilde{w})$, is still implemented. However, the principal is better off to pay less. The extra profit is $\epsilon(\alpha) > 0$. So $(w^*, \bar{a})$ cannot be optimal contract for any $U_a(w^*, \bar{a}) > 0$. The same logic can be applied to the boundary $a = 0$ where $U_a(w^*, 0) < 0$ cannot be optimal.

based on lemma 1, as follows,

$$(P1') \max_{\{a^{BR}(w),w\}} \int v(x-w)f(x,a)dx$$

s.t.

$$\int [u(w) - c(a^{BR}(w))]f(x,a)dx \geq \underline{U} \quad \text{(IR')}$$

$$\int u(w)f_a(x, a^{BR}(w)))dx - c'(a^{BR}(w)) = 0. \quad \text{(IC')}$$

It can be seen that problem $P1'$ is equavalent to problem $P1$ since $a^{BR}(w)$ is a true best response rather than just a solution to LIC.

For convenience, we define a particular class of contract that comes from the first order condition of (P2).

Definition 2.2: *We call $s(x,a;\lambda,\mu)$ the Mirrlees-Holmstrom contract (MH) if $\frac{v'(x-s(x,a;\lambda,\mu))}{u'(s(x,a;\lambda,\mu))} = \lambda + \mu l_a(x,a)$ whenever $\lambda + \mu l_a(x,a) \geq \frac{v'(x-\underline{s})}{u'(\underline{s})}$, and $s(x,a) = \underline{s}$ whenever $\lambda + \mu l_a(x,a) \leq \frac{v'(x-\underline{s})}{u'(\underline{s})}$, where $l_a(x,a) \equiv \frac{f_a(x,a)}{f(x,a)}$.*

Any contract from the $\mathcal{MH}$ class is indexed by three parameters, λ, μ and a. For convenience, we introduce notation $q(x,a) = \lambda + \mu l_a(x,a)$, and let $r(q)$ solve $\frac{v'(x-r(q))}{u'(r(q))} = q$. First of all, we consider the following mapping. If the contract is chosen from the $\mathcal{MH}$ class, for a given λ and μ, the contract $s(x,a;\lambda,\mu)$ can be indexed by $a \in \mathbb{A}$, say, $s(x,a;\lambda,\mu)$. Therefore, the best reponse to $s(x,a;\lambda,\mu)$ can be regarded as a mapping $a^{BR}_{\lambda,\mu} : \mathbb{A} \rightrightarrows \mathbb{A}$ given λ and μ. For notational compactness, we may suppress the index (λ,μ) in the best response $a^{BR}_{\lambda,\mu}$ and $\mathcal{MH}$ contract $s(x,a;\lambda,\mu)$ with no confusion.

We obtain the following important lemma.

Lemma 2.2: *Under assumptions A1-A3, and A4 or A4', suppose there exist a, λ, μ such that: (i) the effort level a is implemented by the MH contract $s(x,a;\lambda,\mu)$, and (ii) $\lambda \geq 0$ satisfies the complementary condition: $\lambda = 0$ if $\int u(s(x,a;\lambda,\mu))f(x,a)dx - c(a) > \underline{U}$; and $\mu \geq 0$ satisfies the complementary condition $\mu = 0$ if $\int u(s(x,a;\lambda,\mu))f_a(x,a)dx > c'(a)$. Then $s(x,a;\lambda,\mu)$ is Pareto-optimal contract implementing a.*

Proof. (By contradiction). Suppose a is implemented by $s(x,a;\lambda,\mu)$ with condition (ii).

When the IR constraint is binding, and supposing there is another contract w implementing

a and keeping the agent's IR constraint binding, we have:

$$\int u(w)f(x,a)dx = \int u(s(x,a;\lambda,\mu))f(x,a)dx,$$

and

$$0 = \mu[\int u(w)f_a(x,a)dx - c'(a)] = \mu[\int u(s(x,a;\lambda,\mu))f_a(x,a)dx - c'(a)].$$

The profit distance between using $s(x,a;\lambda,\mu)$ and w is:

$$\begin{aligned}
\Delta\Pi &= \int v(x-s(x,a;\lambda,\mu))f(x,a)dx - \int v(x-w)f(x,a)dx \\
&= \int v(x-s(x,a;\lambda,\mu))f(x,a)dx + \mu[\int u(s(x,a;\lambda,\mu))f_a(x,a)dx - c'(a)] \\
&\quad +\lambda[\int u(s(x,a;\lambda,\mu))f(x,a)dx - \int u(w)f(x,a)dx] \\
&\quad -\{\int v(x-w)f(x,a)dx + \mu[\int u(w)f_a(x,a)dx - c'(a)]\} \\
&= \int (v(x-s(x,a;\lambda,\mu)) - qu(s(x,a;\lambda,\mu)) - [v(x-w) - qu(w)])f(x,a)dx \\
&\quad - \int q[u(s(x,a;\lambda,\mu)) - u(w)]f(x,a)dx \\
&= \int \{v(x-s(x,a;\lambda,\mu)) - qu(s(x,a;\lambda,\mu)) - [v(x-w) - qu(w)]\}f(x,a)dx \\
&> 0.
\end{aligned}$$

The last step holds as long as $w \neq s(x,a;\lambda,\mu)$ with positive probability since $s(x,a;\lambda,\mu)$ is the pointwise minimizer of $v(x-w) - qu(w)$ for any $q \geq \frac{v'(x-\underline{s})}{u'(\underline{s})}$; and if $q \leq \frac{v'(x-\underline{s})}{u'(\underline{s})}$, $s(x,a;\lambda,\mu) = \underline{s}$. To see this, note that: (i) when $q \geq \frac{v'(x-\underline{s})}{u'(\underline{s})} \geq 0$, the object $v(x-w) - qu(w)$ is concave and has a unique minizer at $w = r(q)$; (ii) when $\frac{v'(x-\underline{s})}{u'(\underline{s})} > q$, $v(x-w) - qu(w)$ is monotonically decreasing in w.

When the IR constraint is not binding, we have $\lambda = 0$, by the definition of $s(.)$,

$$\int u(w)f(x,a)dx \leq \int u(s(x,a;\lambda,\mu))f(x,a)dx,$$

by the same construction, the above reasoning is true as well. Q.E.D. ■

The above lemma shows the fact that if an action can be a fixed point of best response against a $\mathcal{MH}$ contract, given λ and μ such that the IR cosntraint is binding, then the optimal contract to implement that action must belong to the $\mathcal{MH}$ class. To validate FOA is essentially to check whether an action can be implemented by a $\mathcal{MH}$ contract $s(x, a; \lambda, \mu)$ with IR constraint binding.

It is important to point out the relationship between (λ, μ) and effort a constrained by IR and LIC constraint. Jewitt *et al.* (2008) have found the wonderful relationship independently, though our main result does not require such a strong relationship between (λ, μ) and a. For self-containedness, we provide the following formal proof.

Lemma 2.3: *For every given effort $a > 0$, there exists a unique (λ, μ) with $\mu \geq 0$, $\lambda \geq 0$ solving the following equations*

$$\begin{cases} \lambda \left[\int u(s(x, a; \lambda, \mu) f(x, a) dx - c(a) - \underline{U} \right] = 0 \\ \mu [\int u(s(x, a; \lambda, \mu)) f_a(x, a) dx - c'(a)] = 0 \end{cases} \tag{2.1}$$

where $\int u(s(x, a; \lambda, \mu) f(x, a) dx - c(a) - \underline{U} \geq 0$ and $\int u(s(x, a; \lambda, \mu)) f_a(x, a) dx - c'(a) \geq 0$ should hold.

Proof. Note that for any given λ and a, $\int u(s(x, a; \lambda, \mu)) f_a(x, a) dx$ is strictly monotone in μ. And if $\mu = 0$,

$$\int u(s(x, a; \lambda, 0)) f_a(x, a) dx - c'(a) \leq 0$$

and there exist $\bar{\mu}$ such that every $\mu > \bar{\mu}$

$$\int u(s(x, a; \lambda, \mu)) f_a(x, a) dx - c'(a) > 0$$

Therefore, by the median value theorem, there must exist a unique μ such that

$$\int u(s(x, a; \lambda, \mu)) f_a(x, a) dx - c'(a) = 0$$

holds. If $\int u(s(x, a; \lambda, 0)) f_a(x, a) dx - c'(a) > 0$, we set $\mu = 0$. For the μ satisfying the above conditions, we call it $\mu(\lambda, a)$. Also $\mu(\lambda, a)$ will depend continuously on (λ, a) by the implicit function theorem. Substituting $\mu(\lambda, a)$ into the IR constraint, we have

$$\frac{\partial}{\partial\lambda}\int u(s(x,a;\lambda,\mu(\lambda,a)))f(x,a)dx$$
$$= \int u'(.)r'(.)f(x,a)dx - \frac{\left[\int u'(.)r'(.)l_a f(x,a)dx\right]^2}{\int u'(.)r'(.)l_a^2 f(x,a)dx}$$

If we write $Z_1 = \sqrt{u'(.)r'(.)}$, $Z_2 = \sqrt{u'(.)r'(.)}l_a(x,a)$, by the Cauchy–Schwarz inequality, then we have

$$\mathbb{E}Z_1^2\mathbb{E}Z_2^2 - (\mathbb{E}Z_1Z_2)^2 \geq 0$$

and equality holds only if Z_1 is linear in Z_2[8]. Because $l_a(x,a)$ can not be a constant, Z_1 cannot be linear in Z_2. Therefore, the above inequality must be strict, and

$$\frac{\partial}{\partial\lambda}\int u(s(x,a;\lambda,\mu(\lambda,a)))f(x,a)dx > 0.$$

So if $\int u(s(x,a;0,\mu(0,a)))f(x,a)dx < c(a) + \underline{U}$, then by the median value theorem again, there must exist a unique $\lambda > 0$, making IR constraint binding. If $\int u(s(x,a;0,\mu(0,a)))f(x,a)dx > c(a) + \underline{U}$, then we set $\lambda = 0$, and solve $\mu(0,a)$. Q.E.D. ■

Following from the above proof, we have corollary 2.

Corollary 2.2: *Let $(\lambda^*(a), \mu^*(a))$ be the solution of (1), then the implicit function $\lambda^*(a)$ and $\mu^*(a)$ are continuously differentiable in a.*

Basically, we want to provide a set of conditions under which the following problem P3 will be equivalent to problem P1:

$$(P3): \max_{\{\lambda,\mu,a\}} \int v(x - s(x))f(x,a)dx$$

s.t.

$$\int [u(s(x)) - c(a^{BR}(s))]f(x,a)dx \geq \underline{U} \tag{IR'}$$

$$\int u(s(x))f_a(x,a^{BR}(s)))dx - c'(a^{BR}(s)) = 0. \tag{IC'}$$

where $s(x) \in \mathcal{MH}$.

[8]This fact is also noted by Jewitt, Kadan and Swinkels (2008).

It becomes clear based on the fixed point approach that the validity of FOA requires two conditions, summarized as follows.

Lemma 2.4: *Under assumptions A1-A3, and A4 or A4′, (P3) is equavalent to (P1), if (i) every* $a \in [0, \bar{a}]$ *can be sustained by a certain combination of* (λ, μ)*; (ii) the* (λ, μ) *pair solves system (1).*

Proof. If every $a \in [0, \bar{a}]$ can be covered by (λ, μ), then for any action a^* which is implemented by contract w^*, we can find a pair of (λ^*, μ^*) such that a^* becomes a fixed point against the $\mathcal{MH}$ contract $s(x, a^*, \lambda^*, \mu^*)$ with IR constraint binding. Based on Lemma 1 and Lemma 2, contract $s(x, a^*, \lambda^*, \mu^*)$ will be the least cost contract to implement a^*. Q.E.D. ■

Before proceeding, we provide a tight necessary condition for FOA to be valid.

Proposition 2.1: *FOA is valid only if there is at least a fixed point* $a \in a^{BR}(s(x, a, \lambda^*(a), \mu^*(a)))$, *where* $(\lambda^*(a), \mu^*(a))$ *are functions of* a*, implicitly defined by system (1).*

Proof. Suppose FOA is valid, then the optimal effort a^* is implemented by $s(x, a^*; \lambda, \mu)$, and (λ, μ) is determined by system (1), which means $a^* \in a^{BR}(s(x, a^*, \lambda^*(a^*), \mu^*(a^*)))$, which is equivalent to $a^{BR}(s(x, a, \lambda^*(a), \mu^*(a)))$ at least admits a fixed point. Q.E.D. ■

By adding another condition, actually, we can have a necessary and sufficient condition for the FOA to be valid. We present this stronger result in Section 5, where the agent's utility is more general.

2.3 Validity of FOA and Existence of Solution

If $a^{BR}(s(x, a, \lambda, \mu))$ is continuous in (a, λ, μ), then the necessary condition is met since $a^{BR}(s(x, a, \lambda^*(a), \mu^*(a$ is also continuous in a. We now show that the continuity of $a^{BR}(s(x, a; \lambda, \mu))$ in (a, λ, μ) is also sufficient so that the two conditions in Lemma 4 are met.

2.3.1 Validity of FOA

Theorem 2.1: *Under assumptions A1-A3, and A4 or A4′, (P3) is equivalent to (P1), if either* $a^{BR}(s(x, a, \lambda, \mu))$ *is continuous in* μ *for every given* (a, λ) *or* $a^{BR}(s(x, a, \lambda^*(a, \mu), \mu))$ *is continuous in* μ *for every given* a.

Proof. (By construction) For every effort level $\hat{a}$ to be implemented, we construct a $\mathcal{MH}$

contract and make IR constraint binding as follows.

Step 0. Boundedness of effort level $\hat{a}$.

Since the optimizer a^* must be bounded, let $\bar{a}$ be the upper bound of effort responding to any optimal contract $\bar{a} = \sup a^{BR}(w^*)$.

Step 1. Construction of contract

Given $\hat{a} \leq \bar{a}$, based on Lemma 3, we construct contract $s(x, \hat{a}, \hat{\lambda}, \mu)$ where $\hat{\lambda} = \lambda^*(\hat{a})$ is the solution of system (1), associated with $\hat{\mu}$ (but we will vary μ). Therefore, $a^{BR}(s(x, \hat{a}, \hat{\lambda}, \mu))$ is continuous in μ.

Step 2. Adjustment of $a^{BR}(.)$ by μ

We claim that we can adjust the best response by varying μ based on continuity. Note that when $\mu = 0$, $a^{BR}(.) = 0$, and therefore, if by varying μ, we can achieve $a^{BR}(s(x, \hat{a}, \hat{\lambda}, \mu)) = \bar{a}$, we are done. Suppose that $\hat{a}$ can not be achieved by adjusting μ. Therefore, there must exist a largest fixed point sustained by μ. Let $a^{*\bar{\mu}}_{\max} = \sup_{\mu}\{a \in a^{BR}(s(x, a, \hat{\lambda}, \mu))\}$ be the largest fixed point a which can be sustained by μ. We discuss two cases.

(i) The first case is that if any $a \in [0, a^{*\bar{\mu}}_{\max}]$ is fully covered by adjusting μ, then the new fixed point(s) $a^{*\mu+\epsilon}$ against $s_{\epsilon}(x, a^{*\mu+\epsilon}, \hat{\lambda}, \mu + \epsilon)$ must fall beyond $[0, a^{*\bar{\mu}}_{\max}]$ since the same fixed point can never be sustained by two different μ's due to

$$\frac{\partial}{\partial \mu}[\int u(s(x, a, \hat{\lambda}, \mu)) f_a(x, a) dx] = \int u'(.) r'(.) l_a^2(x, a) f(x, a) dx > 0$$

Therefore, $a^{*\mu+\epsilon} > a^{*\bar{\mu}}_{\max}$, a contradiction to $a^{*\bar{\mu}}_{\max}$ being the largest fixed point.

(ii) The second case is that if $a \in [0, a^{*\bar{\mu}}_{\max}]$ is not fully covered by adjusting μ, then there must exist some gap(s) which cannot be covered by μ anymore. In this case, since $a^{*\mu+\epsilon}$ can not fall into the gaps, nor can the places be covered by μ, $a^{*\mu+\epsilon} > a^{*\bar{\mu}}_{\max}$ will be the case. We obtain the same contradiction.

Based on (i) and (ii), by varying μ, the fixed point is unbounded. Therefore, by continuity, we can adjust μ such that the best response $\hat{a} \in a^{BR}(s(x, \hat{a}, \hat{\lambda}, \mu)$. This means $\int u(s(x, \hat{a}, \hat{\lambda}, \mu)) f_a(x, \hat{a}) dx - c'(\hat{a}) = 0$. By Lemma 3, the solution to system (1) is unique. Therefore, $\mu = \hat{\mu}$ will also make IR constraint binding. Therefore, the two conditions in Lemma 4 are met.

Step 3. Implementation of any action

Since any $\hat{a} \leq \bar{a}$ can be implemented by the above process, FOA is valid.

Similarly, when the contract is $s(x, a, \lambda^*(a, \mu), \mu)$, the proof relies on continuity of $\lambda^*(a, \mu)$ for every given a, and the following monotonicity:

$$\begin{aligned}&\frac{\partial}{\partial \mu}[\int u(s(x, a, \lambda(a, \mu), \mu))f_a(x, a)dx]\\ = \quad & -\frac{\left[\int u'(.)r'(.)l_a(x, a)f(x, a)dx\right]^2}{\int u'(.)r'(.)f(x, a)dx} + \int u'(.)r'(.)l_a^2(x, a)f(x, a)dx > 0\end{aligned}$$

as we show in Lemma 4. Q.E.D. ■

The proof of the above theorem does not require the global concavity of the agent's utility. For instance, in our setting, as long as the best reponse is unique, the agent's utility could be concave, quasi-concave, convex or quasi-convex, or convex first and concave later. Even if the best response is not unique, as long as there exists a path-connected correspondence, the above theorem will be true as well. Therefore, it allows us to validate FOA in some situations where the existing literature does not apply. For example, as we will see in next subsection, when the limited liability constraint is effective, the payment rule is no longer concave even though restrictions are introduced on the monetary utility function.

It is worth pointing out that under some conditions, the fixed point a^* will be a monotone function of μ, and we are able to continuously adjust the fixed point by varying μ.

Proposition 2.2: *If the conditions for Theorem 1 hold, in addition, the $l_{ax}(x) \geq 0$ and $l_{aax} \leq 0$, then for every given μ, the best response correspondence, $a^{BR}(s(x, a, \hat{\lambda}(a, \mu), \mu) : A \rightrightarrows A$ has and has only one unique fixed point. Meanwhile, the fixed point $a^*(\mu)$ is a continuously monotone increasing function of μ.*

Proof. Based on the conditions in Theorem 1, for every μ, the fixed point exists. To show the uniqueness of the fixed point, it sufficies to show the monotonicity. For the fixed-point a^*, let $\lambda^*(a^*, \mu)$ be the λ solving the IR constraint and substitute $\lambda^*(a^*, \mu)$ into the LIC constraint, and take the derivative with respect to μ, we have:

$$\begin{aligned}
& \frac{\partial}{\partial a^*}[\int u(s(x,a^*,\lambda(a^*,\mu),\mu))f_a(x,a^*)dx - c'(a^*)] \\
= & \int u'(.)r'(.)\frac{\partial \lambda^*}{\partial a^*} f_a dx + \mu \int u'(.)r'(.)l_{aa} f_a dx + \int u(r(.)) f_{aa} dx - c''(a^*) \\
= & -\int u'(.)r'(.) f_a dx \frac{\mu \int u'(.)r'(.)l_{aa} f dx}{\int u'(.)r'(.) f dx} + \mu \int u'(.)r'(.)l_{aa} f_a dx + \int u(r(.)) f_{aa} dx - c''(a^*) \\
< & -\mu \int u'(.)r'(.) f_a dx \frac{\int u'(.)r'(.)l_{aa} f dx}{\int u'(.)r'(.) f dx} + \mu \int u'(.)r'(.)l_{aa} f_a dx.
\end{aligned}$$

We want to show $\int u'(.)r'(.)fdx \int u'(.)r'(.)l_{aa}f_a dx - \int u'(.)r'(.)f_a dx \int u'(.)r'(.)l_{aa} f dx \leq 0$ under condition $l_{ax}(x) \geq 0$ and $l_{aax} \leq 0$.

Construct a function,

$$\varphi(\tau) = \int^{\tau} u'(.)r'(.) f dx \int^{\tau} u'(.)r'(.)l_{aa} f_a dx - \int^{\tau} u'(.)r'(.) f_a dx \int^{\tau} u'(.)r'(.)l_{aa} f dx$$

with $\varphi(x_{\min}) = 0$. It suffices to show $\varphi'(\tau) \leq 0$. This is to show

$$\begin{aligned}
\frac{\varphi'(\tau)}{u'(.)r'(.)f} &= \int^{\tau} u'(.)r'(.)l_{aa} f_a dx + l_{aa} l_a \int^{\tau} u'(.)r'(.) f dx \\
& \quad -l_a \int^{\tau} u'(.)r'(.)l_{aa} f dx - l_{aa} \int^{\tau} u'(.)r'(.) f_a dx \\
& \leq 0
\end{aligned}$$

Denote $\tilde{\varphi}(\tau) = \frac{\varphi'(\tau)}{u'(.)r'(.)f}$, we have $\tilde{\varphi}(x_{\min}) = 0$. And to show $\tilde{\varphi}(\tau) \leq 0$, it suffices to show $\tilde{\varphi}'(\tau) \leq 0$. Note that due to $l_{ax} \geq 0$ and $l_{aax} \leq 0$, we have

$$\begin{aligned}
\tilde{\varphi}'(\tau) &= [l_{aa} l_a]' \int^{\tau} u'(.)r'(.) f dx - l_{ax} \int^{\tau} u'(.)r'(.)l_{aa} f dx - l_{aax} \int^{\tau} u'(.)r'(.) f_a dx \\
&= l_{ax} \int^{\tau} u'(.)r'(.)[l_{aa}(\tau, a) - l_{aa}] f dx + l_{aax} \int^{\tau} u'(.)r'(.)[l_a(\tau, a) - l_a] f dx \\
&\leq 0.
\end{aligned}$$

Therefore,

$$\frac{\partial a^*}{\partial \mu} = -\frac{\frac{\partial}{\partial \mu}[\int u(s(x,a^*,\lambda(a^*,\mu),\mu))f_a(x,a^*)dx - c'(a^*)]}{\frac{\partial}{\partial a^*}[\int u(s(x,a^*,\lambda(a^*,\mu),\mu))f_a(x,a^*)dx - c'(a^*)]} > 0$$

where $\frac{\partial}{\partial\mu}[\int u(s(x,a^*,\lambda(a^*,\mu),\mu))f_a(x,a^*)dx - c'(a^*)] > 0$ according to the proof in Lemma 3.

The continuity of $a^*(\mu)$ is because the object $\int u(s(x,a,\lambda(a,\mu),\mu))f_a(x,a)dx - c'(a)$ is continuous and monotone in a, since for each μ, the best response correspondence, $a^{BR}(s(x,a,\hat{\lambda}(a,\mu),\mu) : \mathbb{A} \rightrightarrows \mathbb{A}$ has one unique fixed point. Q.E.D. ■

2.3.2 Existence of Solution

Theorem 1 only specifies the condition for the equavalence between (P1) and (P3). We need to prove the existence of a solution to (P3). Holmstrom's (1979) condition can be applied to assure the existence of a solution, by assuming that the payment $\frac{\partial w}{\partial x} \leq 1$. With CDFC along with MLRP, the existing literature also shows the existence of a solution (Dana, 2005, Jewitt, 2008). Below, we prove the existence of the solution with the results based on Theorem 1.

Theorem 2.2: *If the conditions for Theorem 1 hold, then the solution to (P3) exists. Particularly, when the principal is risk-neutral, LIC is a necessary constraint regardless of the boundedness of the action space.*

Proof. First, by Lemma 3, for each a, $\lambda^*(a)$ and $\mu^*(a)$ are continuously differentiable in a. And given a, $\lambda^*(a)$ and $\mu^*(a)$, $a^{BR}(s(x,a,\lambda^*(a),\mu^*(a)))$ is also continuous in a. Therefore, the objective function

$$\int v(x - s(x,a,\lambda^*(a),\mu^*(a)))f(x,a^{BR}(s(a,\lambda^*(a),\mu^*(a))))dx$$

is continuous in a.

Second, let $\Gamma(a)$ be the set that satisfies two constraints IR' and IC' as listed in P3. $\Gamma(a)$ is non-empty due to the existence of fixed point $a \in a^{BR}(s(a,\lambda^*(a),\mu^*(a)))$ based on continuity. We want to check that the constrained set is closed. Let $a_n \in \Gamma(a_n)$ be a sequence converging to a. By definition, then IC' and IR' hold for a_n. By upper-hemi-continuity of the best response, when $a_n \to a$, then $a^{BR}(s(a_n,\lambda^*(a_n),\mu^*(a_n))) \to a^{BR}(s(a,\lambda^*(a),\mu^*(a)))$, and therefore, IC' and IR' will hold for a based on continuity. Thus, $a \in \Gamma(a)$.

Third, we check the boundedness of a. If $a \in [0,\bar{a}]$ is typically assumed, we are done. If $a \in [0,\infty)$, we want to show that the optimality can not happen at $a \to \infty$. By contradiction, suppose that the optimal solution to (P3) happens at $a^* \to \infty$. Therefore, we have

$$-\infty < M < \int v(x - s(x, a^*, \lambda^*, \mu^*))f(x, a^*)dx$$

for M small enough. By Jesen's inequality, we have

$$\int [x - s(x, a^*, \lambda^*, \mu^*)]f(x, a^*)dx \geq v^{-1}(M).$$

Therefore, $\mathbb{E}s(x, a^*, \lambda^*, \mu^*)$ is bounded by $C_1\mathbb{E}x + C_2$ with two finite constants $C_1 > 0$ and C_2. If $\mathbb{E}s(x, a^*, \lambda^*, \mu^*)$ is bounded by $C_1\mathbb{E}x + C_2$, based on IR constraint,

$$\underline{U} \leq u(C_1\mathbb{E}x + C_2) - c(a^*),$$

therefore, a^* can not be unbounded, a contradition to $a^* \to \infty$.

Putting these three steps together, solution of (P3) exists and it is the same solution for (P1). Q.E.D. ■

Theorem 1 and 2 provide a set of conditions where FOA is valid and the solution to P1 exists. To find the solution is to find the fixed point of the best response correspondence. There is a lot of literature dealing with computation of a fixed point (e.g., Scarf, 1967), and the computation becomes numerically available. Sometime, it is also useful to search for the fixed point based on the following corollary.

Corollary 2.3: *If for every a, we have $a \in a^{BR}(s(x, a; \lambda^*(a), \mu^*(a)))$, then FOA is valid.*

Proof. It directly follows from Theorem 1 since every a could be a fixed point. Q.E.D. ■

It is worth pointing out that $\mathbb{E}v(x-s)$ is continuous in a, but might not be differentiable in a, and therefore the adjoint equation in Holmstrom (1979) might not be applicable in finding the solution. However, if the principal is risk neutral, then $\mathbb{E}(x-s)$ will be continuously differentiable in a, and therefore the adjoint equation will hold too for an inner solution. Formally we provide the following theorem.

Theorem 2.3: *If the conditions in Theorems 1 and 2 hold and in addition, the principal is* *risk neutral, then FOA is valid and the following adjoint equation*

$$\int (x - s(x,a,\lambda,\mu)) f_a(x,a) dx + \mu[\int u(s(x,a,\lambda,\mu)) f_{aa}(x,a) dx - c''(a)] = 0 \qquad (2.2)$$

also holds for every $a \in (0, \bar{a})$*; otherwise,* $a \in \{0, \bar{a}\}$.

Proof. When the principal is risk neutral, we can decompose P3 into a two-stage optimization. The first stage is to choose (λ, μ) to minimize the cost of implementation $\int s(x,a;\lambda,\mu) f(x, a^{BR}(s(x,a;\lambda$ for given any a, and the second stage is to choose the optimal effort a. Let

$$C(a) = \min_{(\lambda,\mu)} \int s(x,a;\lambda,\mu) f(x, a^{BR}(s(x,a;\lambda,\mu)) dx \text{ s.t. IC and IR and } s(.) \geq \underline{s}$$

By Milgrom and Segal (2002), $C(a)$ will be continuously differentiable in $a \in (0, \bar{a})$, even though $a^{BR}(s(x,a;\lambda,\mu)$ is continuous but not differentiable (and at the left (right) boundary point, it is left-hand (right-hand) differentiable). Therefore, the envelope theorems apply and we obtain the adjoint equation in addition to LIC and IR constraints. Q.E.D. ■

In some real situations, there is some exogeneous specification on A4$'$. The cutting point x_0 is exogenously given, rather than an endogeneous solution to $s(x_0, a^*; \lambda^*(a^*), \mu^*(a^*)) = \underline{s}$. Assumption A4$'$ can be replaced by A4$''$.

A4$''$: The limited liability constraint is specified as $w \geq \underline{s}$ *if* $x > x_0$ *and* $w = \underline{s}$ *if* $x \leq x_0$.

Under assumption A4$''$, to find the optimal solution a^*, we need to take constraint $s(x_0, a, \lambda^*(a), \mu^*(a)) \geq \underline{s}$ into consideration. We formally state the following corollary.

Corollary 2.4: *If the conditions in Theorem 3 are met, and A4 or A4$'$ is replaced by A4$''$, then the adjoint equation holds whenever* $s(x_0, a, \lambda^*(a), \mu^*(a)) \geq \underline{s}$*; otherwise*

$$s(x_0, a, \lambda^*(a), \mu^*(a)) = \underline{s}$$

should hold for $a \in (0, \bar{a})$*, where* x_0 *is the cutting point where the limited liability constraint is being activated.*

Proof. When $s(x_0, a, \lambda^*(a), \mu^*(a)) \geq \underline{s}$, based on Theorem 3, we obtain the adjoint equation. Otherwise, based on Kuhn-Tucker's theorem, $s(x_0, a, \lambda^*(a), \mu^*(a)) = \underline{s}$ should hold. Q.E.D. ■

Remark 2.3: *If FOA is valid, under assumption A4, the limited liability constraint must be activited at* x_0 *where* $l_a(x_0, a) < 0$. *Otherwise,* $\frac{v'(x-\underline{s})}{u'(\underline{s})} > \lambda = E\frac{v'(x-s)}{u'(s)} > \frac{v'(x-\underline{s})}{u'(\underline{s})}$, *a contradiction arises.*

We now discuss the connection between the fixed point approach and existing literature. Based on Theorem 1 and Theorem 2, we can prove Rogerson (1985) and Jewitt (1988) in a more intuitive way, and understand the difference between them. We can state their results as follows.

Proposition 2.3: *(Rogerson, 1985) Under assumption A1, A2, and A4 or A4', and assuming (i) MLRP, i.e.,* $l_{ax}(x, a) \geq 0$, *(ii) CDFC, i.e.,* $F_{aa}(x, a) \geq 0$, *then FOA is valid and a solution exists.*

Proof. We can check whether the two conditions in Theorem 1 are met. Under MLRP, we have $\frac{\partial s(x,a;\lambda,\mu)}{\partial x} \geq 0$, while, by CDFC, we have

$$\begin{aligned} U_{aa}(s(x, a; \lambda, \mu), \tilde{a}) &= \int u(s) f_{aa}(x, a) dx \\ &= -\int u'(s) \frac{\partial s}{\partial x} F_{aa}(x, a) dx < 0. \end{aligned}$$

Therefore, $\int u(s(x, a; \lambda, \mu)) f_a(x, \tilde{a}) dx - c'(\tilde{a}) = 0$ has a unique solution $\tilde{a}$ given $\lambda > 0$ and $\mu \geq 0$. And by continuity of $U(a, s)$, when the best response is unique, the best response $a^{BR}(s(x, a; \lambda, \mu)))$ is continuous in (a, λ, μ). Based on Theorem 1, FOA is valid and the solution exists. Q.E.D. ■

Jewitt (1988) relaxes the concavity of $U(w, a)$ to that of $U(s(x, a; \lambda, \mu), a)$.

Proposition 2.4: *(Jewitt, 1988) With assumption (i) MLRP, i.e.,* $l_{ax}(x, a) \geq 0$, *(ii)* $\int^x F_{aa}(\tau, a) d\tau \geq 0$ *for all* x, *(iii)* $l_{axx} \geq 0$ *and* $l_a(x, a)$ *is uniformly bounded from below; and (iv) the principal is risk-neutral and* $u(r(q))$ *concave in* q, *then FOA is valid and a solution exists.*

Proof. Since we utilize the functional form $s(x, a; \lambda, \mu)$, we can integrate by parts one step further as follows,

$$\begin{aligned} U_{aa}(s(x,a;\lambda,\mu),\tilde{a}) &= \int u(s) f_{aa}(x,\tilde{a})dx \\ &= -\int u'(s)s'(x)F_{aa}(x,\tilde{a})dx \\ &= -u'(s(\bar{x},.))\frac{\partial s}{\partial x}(\bar{x},.)\int F_{aa}(x,\tilde{a})dx + \int\left(\int^x F_{aa}(\tau,\tilde{a})d\tau\right)d(u'(s)s'(x)). \end{aligned}$$

When conditions (i), (ii) and (iv) hold, $u'(s)\frac{\partial s}{\partial x}$ is decreasing in x; together with condition (iii), this implies that $U_{aa}(s(x,a;\lambda,\mu),\tilde{a}) < 0$ for any $\tilde{a}$. Therefore, the best response $a^{BR}(s(x,a;\lambda,\mu))$ is continuous in (a,λ,μ). Q.E.D. ■

The connection between Rogerson (1985) and Jewitt (1988) seems very smooth in terms of the fixed point approach. The trade-off between restrictions on utility and on distribution can be seen by integration by parts intuitively. Condition $\int^x F_{aa}(\tau,a)d\tau \geq 0$ can be regarded as a higher order convexity than CDFC.

Remark 2.4: *Rogerson's (1985) result holds if A4 does not hold, but A4′ holds. However, Jewitt's (1988) proof is not robust to the situation where A4 does not hold or there are restrictions on the payment rule. But our approach is able to deal with this issue as we do in example 2.*

2.3.3 Some Properties

With the fixed point approach, we can provide some new conditions to validate FOA. The following proposition is based on Tarsky's fixed point theorem.

Proposition 2.5: *Under assumptions A1-A4 or A4′, if utility and distribution function satisfy condition* $\int u'(s(x,a;\lambda,\mu)s'(x,a;\lambda,\mu)l_{aa}(x,\tilde{a})f_a(x,a)dx \geq 0$ *for any* $\tilde{a}, a \in A$, *then (i)* $a^{BR}(s(x,a,\lambda,\mu))$ *has a fixed point; (ii) if in addition the principal is risk-neutral and the output generating process is additive,*

$$y(x) = a + \varepsilon, \tag{2.3}$$

and FOA therefore is valid.

Proof. (i) We can show a condition under which the best reponse is non-decreasing based on supermodularity. Note that, when

$$\begin{aligned}
\int u(s(x,a_2;\lambda,\mu))f(x,a_2^*)dx - c(a_2^*) &\geq \int u(s(x,a_2;\lambda,\mu))f(x,a_1^*)dx - c(a_1^*) \text{ and} \\
\int u(s(x,a_1;\lambda,\mu))f(x,a_1^*)dx - c(a_1^*) &\geq \int u(s(x,a_1;\lambda,\mu))f(x,a_2^*)dx - c(a_2^*),
\end{aligned}$$

then,

$$\int u(s(x,a_2;\lambda,\mu))[f(x,a_2^*) - f(x,a_1^*)]dx - \int u(s(x,a_1;\lambda,\mu))[f(x,a_2^*) - f(x,a_1^*)]dx] \geq 0 \quad (2.4)$$

Note also that when $\int u(s(x,a;\lambda,\mu))f(x,\tilde{a})dx$ is continuous in $(a,\tilde{a})$, the necessary and sufficient condition for inequality (4) is

$$\int_{x_0} u'(s(x,a;\lambda,\mu)s'(x,a;\lambda,\mu)l_{aa}(x,a)f_a(x,\tilde{a})dx \geq 0$$

where x_0 solves $s(x,a_2;\lambda,\mu) = \underline{s}$. Under this condition, then $a_2^* \geq a_1^*$ must be the case. By Tarsky's fixed point theorem, $a^{BR}(s(x,a;\lambda,\mu)$ has at least one fixed point.

(ii) Based on (i), the fixed point exists for any given (λ,μ). From additivity (3), we have

$$\begin{aligned}
f_a(x,a) &= -g'(\varepsilon)y', \\
l_a(x,a) &= -\frac{g'(\varepsilon)y'}{g(\varepsilon)y'} = -\frac{g'(\varepsilon)}{g(\varepsilon)} \\
f_{aa}(x,a) &= g''(\varepsilon)y' \\
l_{aa}(x,a) &= (\frac{g'}{g})' \\
\text{and } l_{ax}(x,a) &= -(\frac{g'}{g})'y'.
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \frac{\partial}{\partial a}\int u(s(x,a;\lambda,\mu))f(x,\tilde{a})dx \\
= & \ \mu\int u'(.)s'(.)l_{aa}(x,a)f(x,\tilde{a})dx \\
= & \ u(s(\underline{x},a;\lambda,\mu))g(\underline{y}-\tilde{a})-u(s(\bar{x},a;\lambda,\mu))g(\bar{y}-\tilde{a})-\int u(s(x,a;\lambda,\mu))l_a(x,\tilde{a})f(x,\tilde{a})dx \\
= & \ g(\underline{y}-\tilde{a})[u(s(\underline{x},a;\lambda,\mu))-u(s(\bar{x},a;\lambda,\mu))]-\int u(s(x,a;\lambda,\mu))l_a(x,\tilde{a})f(x,\tilde{a})dx \\
= & \ -\frac{\partial}{\partial\tilde{a}}\int u(s(x,a;\lambda,\mu)f(x,\tilde{a})dx+\text{Constant}
\end{aligned}$$

Similarly,

$$\frac{\partial}{\partial a}\int u(s(x,a;\lambda,\mu))f_a(x,\tilde{a})dx=-\frac{\partial}{\partial\tilde{a}}\int u(s(x,a;\lambda,\mu)f_a(x,\tilde{a})dx.$$

Therefore, if $c'(a)$ is non-decreasing, then for given (λ,μ), there is only a unique a solving equation

$$\int u(s(x,a;\lambda,\mu))f_a(x,a)dx-c'(a)=0,$$

which implies that the fixed point a is continuously differentiable in (λ,μ). Therefore, every effort level a can be achieved by adjusting (λ,μ) properly and IR will be binding at the end. Q.E.D. ■

The above proposition depends on some restriction of utility and the output-generating process, but this new set of conditions is not discussed by the existing literature. The advantage of the above proposition is that the cost of effort could take a very general form (e.g., non-smooth). In the future, it will be possible to provide a set of intuitive conditions to validate FOA based on our new approach.

2.3.4 Examples

We provide two examples where the existing literature does not apply. Example 1 violates Jewitt's (1988) specification of utility, while example 2 violates assumption A4. In both examples, the agent's utility is not globally concave in his action.

Example 2.1: *(FOA might be valid when the agent's utility is convex.) The principal is* *risk-neutral. The agent's monetary utility is* $u(w) = \frac{3}{2}w^{\frac{2}{3}}$, *cost of effort* $c(a) = a^2$, *outside reservation utility* $\underline{U} = 0$, *and the distribution function is* $f(x, a) = \frac{1}{a}e^{-\frac{1}{a}x}$ *where* $a \in [0, \bar{a}]$ *is effort and* $x \in [0, \infty)$ *is the output. We assume* $\bar{a}$ *is large enough so that* $U(s(.), 0) \neq U(s(.), \bar{a})$.

In this example, the MH contract is

$$s(x, a, \lambda, \mu) = (\lambda + \mu\frac{x-a}{a^2})^3$$

therefore, the agent's utility under the MH contract is

$$\begin{aligned} U(s(.), \tilde{a}) &= \frac{3}{2}\int(\lambda + \mu\frac{x-a}{a^2})^2 f(x, \tilde{a})dx - \tilde{a}^2 \\ &= \frac{3}{2}[\lambda^2 + \frac{2\lambda\mu}{a^2}(\tilde{a} - a) + \frac{\mu^2}{a^4}(2\tilde{a}^2 - 2\tilde{a}a + a^2)] - \tilde{a}^2 \end{aligned}$$

When $\frac{3\mu^2}{a^4} > 1$, $U(\tilde{a}, s(.))$ *is convex in* $\tilde{a}$, *and when* $\frac{3\mu^2}{a^4} < 1$, $U(\tilde{a}, s(.))$ *is concave. From LIC and IR constraints when* $\tilde{a} = a$, *we have*

$$\frac{3\mu(\lambda a + \mu)}{a^3} = 2a$$

and

$$\frac{3}{2}(\lambda^2 + \frac{\mu^2}{a^2}) - a^2 = 0$$

Therefore, $\mu = \lambda a$, *yielding* $\mu^*(a) = \frac{a^2}{\sqrt{3}}$ *and* $\lambda^*(a) = \frac{a}{\sqrt{3}}$. *To check the validity of FOA, we need to check that for the given MH contract*

$$s(x, a, \lambda^*(a), \mu) = (\frac{a}{\sqrt{3}} + \mu\frac{x-a}{a^2})^3,$$

the best response is continuous in μ *for every given* a. *Therefore, based on Theorem 3, and by the adjoint equation*

$$(1 - 2\sqrt{3}a^2) + \mu(3 - 3) = 0$$

we have

$$a^* = \sqrt{\frac{\sqrt{3}}{6}}, \mu^* = \frac{1}{6}, \lambda^* = \frac{1}{3}\sqrt{\frac{\sqrt{3}}{2}}$$

To check the validity of the solution, we find that $a = \sqrt{\frac{\sqrt{6}}{3}}$ *is indeed a fixed point, which makes the agent's highest expected payoff zero, hitting the IR constraint exactly. The principal's profit is* $\frac{1}{3}\sqrt{\frac{2\sqrt{3}}{3}}$. *(One may assume that if the agent is indifferent in its choice, he will choose the one that the principal instructs.) In this example, we find that when the* MH *contract* $s(x, a, \lambda^*(a), \mu^*(a)) = (\frac{x}{\sqrt{3}})^3$ *is employed, the agent's utility is*

$$U(s(.), \tilde{a}) = \tilde{a}^2 - \tilde{a}^2 = 0$$

which is indifferent against his choice. So any effort the principal favors is weakly implementable.

The observation from the above example is that, when the agent's utility appears to be convex in his action, the principal might be able to choose contracting parameters so that the agent does not have any profitable deviation.

Example 2: *(FOA is valid for normal distribution with limited liability constraint) Suppose the output is normal distribution,* $x \sim N(a, 1)$*, and the utility is* $u(w) = 2\sqrt{w}$*, cost of effort* $c(a) = a^2$*, and the outside reservation utility* $\underline{U} = 0$*. With the limited liability constraint, the* MH *contract is specified as follows:*

$$s(x, a; \lambda, \mu) = \begin{cases} \underline{s} & \text{if } x \leq 0 \\ [\lambda + \mu(x-a)]^2 & \text{if } x \geq 0 \end{cases}$$

where $\underline{s} = 0$.

In this example, the existing literature does not apply due to the existence of the limited liability constraint. We have

$$\begin{aligned} U(s(.), \tilde{a}) &= 2F(0, \tilde{a})\sqrt{\underline{s}} + 2\int_0^\infty (\lambda + \mu(x-a)) f(x, \tilde{a}) dx - \tilde{a}^2 \\ &= \sqrt{\frac{2}{\pi}}\mu e^{-\frac{1}{2}\tilde{a}^2} + [\lambda + \mu(\tilde{a} - a)](1 + \Phi(\frac{\tilde{a}}{\sqrt{2}})) + (\lambda - \mu a)(1 - \Phi(\frac{\tilde{a}}{\sqrt{2}})) - \tilde{a}^2 \end{aligned}$$

and

$$U_a(s(.), \tilde{a}) = \mu(1 + \Phi(\frac{\tilde{a}}{\sqrt{2}}) - 2\tilde{a}$$

where $\Phi(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$ *is the error function. And*

$$U_{aa}(s(.), \tilde{a}) = \sqrt{\frac{2}{\pi}} \mu e^{-\frac{1}{2}\tilde{a}^2} - 2.$$

It can be observed that the agent's utility could be convex, when $\tilde{a}$ *is small enough, but it will be concave when* $\tilde{a}$ *becomes large, and* $U_{aaa}(s(.), \tilde{a}) < 0$ *for any* $\tilde{a} > 0$*. Therefore, the agent's utility is single-peaked. We first solve the LIC and the IR constraint for* $\tilde{a} = a$*. Note that*

$$\begin{aligned} \mu e^{-\frac{1}{2}a^2} \sqrt{\frac{2}{\pi}} + \lambda(1 + \Phi(\frac{a}{\sqrt{2}}) + (\lambda - \mu a)(1 - \Phi(\frac{a}{\sqrt{2}}))] &= a^2 \\ \mu(1 + \Phi(\frac{a}{\sqrt{2}})) &= 2a \end{aligned}$$

Therefore, we have

$$\begin{aligned} \lambda &= \frac{1}{2} \left(a^2 - \frac{2a \left(e^{-\frac{1}{2}a^2} \sqrt{\frac{2}{\pi}} - a(1 - \Phi(\frac{a}{\sqrt{2}})) \right)}{(1 + \Phi(\frac{a}{\sqrt{2}}))} \right) \\ \mu &= \frac{2a}{(1 + \Phi(\frac{a}{\sqrt{2}}))} \end{aligned}$$

Using the constraint $\lambda = \mu a$*, we have equation*

$$\frac{a[a - 2\sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}a^2} - a\Phi(\frac{a}{\sqrt{2}})]}{2(1 + \Phi(\frac{a}{\sqrt{2}}))} = 0$$

Therefore, the only solution is $\lambda = \mu = a = 0.$[9]

2.4 Application to Non-concave Payment Rule with Limited Liability Constraint

In this section, we use the current approach to validate the FOA in situations where the score function is not uniformly bounded from below, CDFC does not hold, and there exists some

[9] By the adjoint equation $1 - \frac{e^{-\frac{1}{2}a^2}}{\sqrt{2\pi}}(\lambda^2 + a^2\mu^2) - 2\mu = 0$, we can find that the profit is maximized at $a = 0.465886$. However, this solution does not satisfy the limited libility constraint.

limited liability constraint. A very useful example is that the output or its monotone tranformation obey a normal distribution. The idea is to show that the best response against a given $\mathcal{MH}$ contract $s(x, \hat{a}, \lambda, \mu)$ is unique. It also turns out that the agent's expected utility is single-peaked.

2.4.1 Additive output-generating process

For the output-generating process (3), if we also assume MLRP holds, then there exists a unique ε^* such that $\varepsilon > \varepsilon^*$ if and only if $l_a(x, a) > 0$. We employ Assumption A4$''$, i.e., the limited liability constraint is exogenously specified at $x = x_0$. Under assumption A4$''$, without loss of generality, we normalize the zero effort as $\underline{a} = y(x_0) - \varepsilon^* = 0$, which implies $a \geq x_0$ (for convenience, we let $\varepsilon^* = 0$). In this case, for a given $\mathcal{MH}$ contract $s(x, \hat{a}, \lambda, \mu)$, it can be seen that the agent's utility might not be globally concave because

$$
\begin{aligned}
& u(\underline{s})F_{aa}(x_0, a) + \int_{x_0} u(s(x, \hat{a}; \lambda, \mu))f_{aa}(x, a)dx \\
= & -\int_{x_0} \frac{\partial}{\partial x} u(s(x, \hat{a}; \lambda, \mu))F_{aa}(x, a)dx \\
> & \ 0
\end{aligned}
$$

when a is close to x_0, even though we put restrictions on the agent's utility like constant relative risk aversion (CRRA). The following proposition provides a set of conditions to make the single-peakedness of the agent's utility.

Proposition 2.6: *If (i) assumption A4$''$ holds, namely, x_0 is exogeneously given; (ii) the boundaries of the effort level $\underline{a}$ and $\bar{a}$ satisfy the normalization $l_a(x_0, \underline{a}) \leq 0$ and $l_a^2(x_0, \bar{a}) + l_{aa}(x_0, \bar{a}) \leq 0$; (iii) the output-generating process is additive and MLRP holds, and $l_{ayy} \leq 0$; (iv) the principal is risk-neutral and the agent's utility satisfies Jewitt's (1988) restriction, namely, $u'''u' - 3u''^2 \leq 0$; and in addition, (v) $c'''(a) \geq 0$ and $c'(0) = 0$; then FOA is valid.*

Proof. Step 1. First of all, we need to normalize the action space. Based on MLRP, let $x^*(a)$ be the unique interior solution to $l_a(x, a) = 0$. For given x_0, we want $l_a(x_0, \underline{a}) \leq 0$ so that for any effort level, the limited liability constraint is activated at some point where the score is negative (for example., if the limited liability constraint is activated at $x_0 = 0$, then we normalize $\underline{a} = 0$. See Remark 3 for some justification). At the same time, we also need to specify the upper bound of the effort. Based on additivity (3) and $l_{ayy} \leq 0$, for $x < x^*$, $l_a^2(x,a) + l_{aa}(x,a)$ crosses the x-axis only once, which is denoted as $x_{\min}$. For $x > x^*$, we claim that $l_a^2(x,a)+l_{aa}(x,a)$ crosses the x-axis odd numbers of times. It can be shown that $\frac{\partial x_{\min}}{\partial a} > 0$. We let $x_0 \geq x_{\min}(\bar{a})$ provided $x_{\min}(\bar{a}) > \underline{a}$.

Step 2. Note that $U_a(s(.),0) > 0$; therefore, to show the single-peakedness of $U(s(.),\tilde{a})$, it suffices to show that $U_{aaa}(s(.),\tilde{a}) \leq 0$. Let $\varpi(x,\hat{a})$ denote $u(s(x,\hat{a};\lambda,\mu))$. Note that $x_{\min}(a) \leq x_0 \leq a$ and $\varpi_{yy}(x,\hat{a}) \leq 0$. Meanwhile, for $x > x_0$, $g''(y-a)$ must at first be less than zero, and greater than zero at the end. So $g''(y-a)$ crosses the x-axis from below and the number of intersection points will be odd (except for the boundary point). We denote those intersections as $(x_1, ..., x_{2k-1})$ and let $x_{2k} = \bar{x}$. Therefore, we have

$$\begin{aligned}
& -\int_{x_0} \varpi_y(x,\hat{a})y'(x)F_{aaa}(x,a)dx \\
= & \int_{x_0} \varpi_y(x,\hat{a})y'(x)g''(y-a)dx \\
= & \sum_{i=0}^{2k-1} \int_{x_i}^{x_{i+1}} \varpi_y(x,\hat{a})y'(x)g''(y-a)dx \\
\leq & \sum_{i=0}^{2k-1} \varpi_y(x_{i+1},\hat{a})[g'(y(x_{i+2})-a) - g'(y(x_i)-a)] \\
\leq & -\varpi_y(x_1,\hat{a})g'(y(x_0)-a) \\
\leq & \ 0
\end{aligned}$$

The condition for $\varpi_{yy}(x,\hat{a}) \leq 0$ is stated by Jewitt (1988), which is $l_{ay} \geq 0$ and $l_{ayy} \geq 0$ and $u'''u' - 3u''^2 \leq 0$. Meanwhile, by $c'''(a) \geq 0$, $U_{aaa}(s(.),\tilde{a}) \leq 0$. Combining these conditions, the agent's utility will be single-peaked. ■

Based on the above proposition, for any normal (or log normal) distribution[10] with $x_0 = 0$, $\sigma \geq \bar{a} - \underline{a}$, if agent's utility is CRRA $u(w) = \frac{1}{\gamma}w^\gamma$ with $\gamma \leq \frac{1}{2}$, then FOA is valid. The restrictions on the boundaries of effort indeed do not hurt because we can tranfer effort level to a compact space by reparameterization since we do not specify the functional form of cost function $c(a)$. But A4$''$ is a key assumption where x_0 is exogenous.

[10]For normal distribution $\mathrm{x}_{\min} = a - \sigma$ and $\mathrm{x}_{\max} = a + \sigma$. This conclusion holds for exponential family distribution with $l_{aa}(x,a) = -\sqrt{\mathbb{E}l_a^2} = -\sigma$.

2.4.2 Exponential family with existence of a sufficient statistic

For exponential family distribution (Brown, 1986), without loss of generality, we can reparameterize the effort and output so that

$$f(x,a)=\frac{e^{ay(x)}\upsilon(x)}{\int e^{ay(x)}\upsilon(x)dx} \text{ where } y'(x)>0 \text{ and } \upsilon(x)>0,\ \int e^{ay(x)}\upsilon(x)dx<\infty \tag{2.5}$$

with normalization $x\in[0,\bar{x}]\subset\mathbb{R}^+$.

The above distribution satisfies MLRP, and therefore A3-i. And it can be shown that $l_a^2(x,a)+l_{aa}(x,a)=0$ has only two solutions, $x_{\min}=y^{-1}(m(a)-\sqrt{\mathbb{E}l_a^2})$ and $x_{\max}=y^{-1}(m(a)+\sqrt{\mathbb{E}l_a^2})$, where $m(a)=\mathbb{E}x$. It can be seen that $\sqrt{\mathbb{E}l_a^2}$ is a constant independent of a. Therefore, we have a similar conclusion.

Proposition 2.7: *For exponential family (5), if (i) assumption A4″ holds, namely, x_0 is exogeneously given; (ii) the boundaries of the effort level $\underline{a}$ and $\bar{a}$ satisfy normalization $y(x_0)\le m(\underline{a})$ and $m(\bar{a})\le y(x_0)+\sqrt{\mathbb{E}l_a^2}$; (iii) $\int_{y^{-1}(m(a))}F_{aaa}(x,a)dy\ge 0$ holds; and (iv) the principal is risk netral and the agent's utility satisfies Jewitt's (1988) restriction, namely, $u'''u'-3u''^2\le 0$; in addtion, (v) $c'''(a)\ge 0$ and $c'(0)=0$; then FOA is valid.*

Proof. Observing that $l_a^2(x,a)+l_{aa}(x,a)=0$ only has two solutions $x_{\min}=y^{-1}(a-\sqrt{\mathbb{E}l_a^2})$ and $x_{\max}=y^{-1}(a-\sqrt{\mathbb{E}l_a^2})$, we see for $x>y^{-1}(a)$, F_{aaa} crosses the x-axis from above only once. Therefore,

$$\begin{aligned}
& -\int_{x_0}\varpi_y(x,\hat{a})y'(x)F_{aaa}(x,a)dx \\
\le\ & -\varpi_y(y^{-1}(a+\sqrt{\mathbb{E}l_a^2}),\hat{a})\int_{x_0}y'(x)F_{aaa}(x,a)dx \\
\le\ & 0
\end{aligned}$$

The last step is because $\int_{x_0}F_{aaa}(x,a)dy\ge 0$ for $x_0\le a$ since $\int_{y^{-1}(m(a))}F_{aaa}(x,a)dy\ge 0$. Q.E.D. ■

The above two propositions should be very useful for empirical research, where the additive output-generating process and exponential families are commonly used. We also explicitly provide the following corollary for Gamma distribution, where CDFC does not hold and the agent's utility is not globally concave.

Corollary 2.5: *For Gamma distribution* $f(x,a) = \frac{e^{-\frac{x}{a}}x^{\alpha-1}}{\Gamma(\alpha)a^{\alpha}}$, *if the effort level is within bound* $A \subset [\frac{x_0}{2+\alpha}, \frac{x_0}{2+\alpha-\sqrt{2+\alpha}}]$ *and utility satisfies* $u'''u' - 3u''^2 \leq 0$, *then FOA is valid.*

Proof. For Gamma distribution, note that $\int_t F_{aaa}(x,a)dy = -\frac{2e^{-\frac{t}{a}}t^{\alpha+1}a(t-(2+\alpha)a)}{\Gamma(a)a^{5+\alpha}} \geq 0$ if $t \leq (2+\alpha)a$, and that the two solutions to $F_{aaa}(x,a) = 0$ are $x_{\min} = a(2+\alpha-\sqrt{2+\alpha})$ and $x_{\max} = a(2+\alpha+\sqrt{2+\alpha})$. Therefore, if $a(2+\alpha-\sqrt{2+\alpha}) \leq x_0 \leq (2+\alpha)a$, we can show the single-peakedness. It turns out that the action space $\mathbb{A} = [\underline{a}, \bar{a}] \subset [\frac{x_0}{2+\alpha}, \frac{x_0}{2+\alpha-\sqrt{2+\alpha}}]$. Q.E.D. ■

2.5 Generalizations and A Necessary and Sufficient Condition

2.5.1 Generalization of Theorems to Non-separable Utility

Theorems 1 and 2 can be generalized to a non-separable environment. The specification of the agent's utility is the same as Alvi (2004). But our result sheds some light on validity of FOA without global concavity. Before proceeding, we extend system (1) to the general utility case as follows:

$$\begin{cases} \lambda\left[\int u(s(x,a;\lambda,\mu),a)f(x,a)dx - \underline{U}\right] = 0 \\ \mu\left[\int [u(s(x,a;\lambda,\mu),a)f_a(x,a) + u_a(s(x,a;\lambda,\mu),a)f(x,a)]dx\right] = 0 \end{cases} \tag{2.6}$$

where $\int u(s(x,a;\lambda,\mu),a)f(x,a)dx - \underline{U} \geq 0$ and $\int [u(s(x,a;\lambda,\mu),a)f_a(x,a) + u_a(s(x,a;\lambda,\mu),a)f(x,a)]dx \geq 0$ should hold.

Theorem 2.4: *Theorems 1 and 2 hold if the agent utility is non-separable,* $u(w,a)$ *with the assumptions (i)* $v_w < 0$, $u_w > 0$, $u_a < 0$ *and (ii)* $u_{ww} < 0$, $u_{aa} \leq 0$, $u_{aw} \leq 0$, $u_{aww} \leq 0$.

Proof. We generalize Theorem 1 first; the proof of generalization of Theorem 2 can be done similarly.

Step 0. Construction of the $\mathcal{MH}$ contract

In the non-separable case, under assumptions (i) and (ii), the $\mathcal{MH}$ contract can be specified as follows:

$$\begin{aligned} -v_x(x-s) + \lambda u_w(s,a) + \mu[u_{aw}(s,a) + u_w(s,a)l_a(x,a)] &= 0 \text{ whenever } q \geq \frac{v_x - \mu u_{aw}}{u_w} \\ s &= \underline{s} \text{ otherwise} \end{aligned}$$

which is because for every given $(x, a; \lambda, \mu)$, there is a unique s that solves the above equation. So the $\mathcal{MH}$ contract $s(x, a; \lambda, \mu)$ is continuous in each of its arguments.

Step 1. Fixed point argument

Similar to the proof of Lemma 2, when there exists (a, λ, μ) such that (i) $a \in a^{BR}(s(x, a; \lambda, \mu))$; (ii) $\lambda \geq 0$ satisfies the complementary condition: $\lambda = 0$ if $U(s(x, a; \lambda, \mu), a) > \underline{U}$; and (iii) $\mu \geq 0$ satisfies the complementary condition: $\mu = 0$ if $U_a(s(x, a; \lambda, \mu), a) > 0$; then $s(x, a; \lambda, \mu)$ is a Pareto-optimal contract implementing the effort a.

Step 3. Continuity of $\lambda(a)$ and $\mu(a)$ for every a.

The proof is similar to Lemma 3. Formally, we state as follows. Note that for every given (μ, a),

$$\frac{\partial s}{\partial \lambda} = -\frac{u_w^2}{(v_{xx}u_w + v_x u_{ww}) + \mu(u_{aww}u_w - u_{aw}u_{ww})} > 0$$

so $\int u(s, a)f(x, a)dx$ is strictly monotone in (μ, a), so either there is a unique $\lambda > 0$ solving

$$\int u(s, a)f(x, a)dx - \underline{U} = 0$$

or $\int u(s, a)f(x, a)dx - \underline{U} > 0$ implying $\lambda = 0$. We call it $\lambda(\mu, a)$, which is continuously differentiable in (μ, a). Substituting $\lambda(\mu, a)$ into the LIC constraint, and observing that

$$\frac{\partial s}{\partial \mu} = -\frac{u_w^2}{(v_{xx}u_w + v_x u_{ww}) + \mu(u_{aww}u_w - u_{aw}u_{ww})}(l_a + \frac{u_{aw}}{u_w})$$

we have,

$$\begin{aligned}
& \frac{\partial}{\partial \mu}[\int u(s(x, a, \lambda(\mu, a), \mu), a)f_a(x, a)dx + \int u_a(s(x, a, \lambda(\mu, a), \mu), a)f(x, a)dx] \\
= & \frac{\partial \lambda(\mu, a)}{\partial \mu} \int [u_w(s(.), a)l_a(x, a) + u_{aw}(s(.), a)]\frac{\partial s}{\partial \lambda} f(x, a)dx \\
& + \int [u_w(s(.), a)l_a(x, a) + u_{aw}(s(.), a)]\frac{\partial s}{\partial \mu} f(x, a)dx \\
= & \mathbb{E}Z_2^2 - \frac{(\mathbb{E}Z_1 Z_2)^2}{\mathbb{E}Z_1^2} \\
\geq & 0
\end{aligned}$$

where $Z_1 = \sqrt{u_w \frac{\partial s}{\partial \lambda}}$, and $Z_2 = \sqrt{u_w \frac{\partial s}{\partial \lambda}}(l_a + \frac{u_{aw}}{u_w})$ and the last step is due to Cauthy-Schwarz inequality. We want to show that there is only a unique μ solving $\mu U_a(s(x,a;\lambda(a,\mu),\mu),a) = 0$ for every given a.

If $l_a + \frac{u_{aw}}{u_w}$ is not a constant, the above Cauchy-Schwarz inequality is strict; therefore, $U_a(s(x,a;\lambda(a,\mu),\mu),a)$ is monotone in μ, as a result, the solution to $\mu U_a(s(x,a;\lambda(a,\mu),\mu),a) = 0$ is unique. If $l_a + \frac{u_{aw}}{u_w}$ is a constant, which requires $u_{aw} \neq 0$ with a strict positive probability. Therefore, we can write $s(x,.)$ in terms of l_a and a constant C, as follows,

$$s(x,a;\lambda(a,\mu),\mu) = w(C - l_a).$$

And $w(.)$ is a decreasing function, implying that $s(x,a;\lambda(a,\mu),\mu)$ is an increasing function of l_a. Meanwhile, note that for any x, $u_w(s(x,.),a)l_a + u_{aw}(s(x.,),a) = Cu_w(s(x,.),a)$ with the constant C. Therefore, $u(s(x,.),a)l_a(x,a) + u_a(s(x,.),a) = C(u(s(x,.),a) - u(\underline{w},a))$, and we have

$$\mu U_a = \mu C \int (u(s(x,a;\lambda(a,\mu),\mu),a) - u(\underline{w},a))f(x,a)dx = 0.$$

As a result, either $C = 0$ or $\mu = 0$ since $s \neq \underline{w}$, which implies that $\frac{v_x(x-s)}{u_w(s,a)} = \lambda$ is a constant. Note that

$$C = \frac{\mathbb{E}[u_w(s,a)l_a + u_{aw}(s,a)]}{\mathbb{E}u_w(s,a)} < 0$$

due to $\mathbb{E}u_w(s(x.,),a)l_a = Cov(u_w(s(x,.),a),l_a) < 0$[11] and $\mathbb{E}u_{aw}(s,a) \leq 0$. Therefore $\mu = 0$ is the unique solution to $\mu U_a(s(x,a;\lambda(a,\mu),\mu),a) = 0$. We call this solution $\mu^*(a)$. As a result, $\mu^*(a)$ and $\lambda^*(\mu^*(a),a)$ both are continuously differentiable in a.

Step 4. Adjustment of μ to implement every effort $a \in [0,\bar{a}]$.

The remainder of proof is essentially the same as Theorem 1. Fixed (λ,a), if the best response is continuous in μ, then we are able to adjust μ such that the complementary condition for IR constraint holds. So far, we generalize Theorem 1. The rest for existence is similar to Theorem 2. Q.E.D. ■

Remark 2.5: *If the principal is risk-neutral, then we can show that $\mu > 0$ once LIC holds under a MH contract, since $Cov(u, \frac{1}{u_w} - \mu\frac{u_{aw}}{u_w}) > 0$.*

[11] Because $s(x,a;\lambda(a,\mu),\mu) = w(C-l_a)$ is an increasing function of l_a and $u_{ww} < 0$, so $Cov(u_w(s(x,.),a),l_a) < 0$. This conclusion also holds when there is limited liability constraint A4′ since the limited liability constraint is activited at x_0 where $l_a(x_0,a) < 0$.

2.5.2 A Necessary and Sufficient Condition for Validity of FOA

Furthermore, we extend proposition 1 to the following theorem, which is a necessary and sufficient condition for FOA to be valid.

Theorem 2.5: *Under the assumptions (i) $v_w < 0$, $u_w > 0$, $u_a < 0$ and (ii) $u_{ww} < 0$, $u_{aa} \leq 0$, $u_{aw} \leq 0$, $u_{aww} \leq 0$, the FOA is valid if and only if there at least exist one $a^* \in [0, \overline{a}]$ so that (i) $a^* \in a^{BR}(s(x, a^*; \lambda^*(a^*), \mu^*(a^*)))$ and (ii)*

$$a^* \in \arg\max_a \int v(x - s(x, a; \lambda^*(a), \mu^*(a)))f(x, a)dx$$

where $(\lambda^*(a), \mu^*(a)$ is defined by system (6).

Proof. Necessary part. If the FOA is valid, the solution must be one of the fixed points. So (i) is necessary. At the same time, if the FOA is valid, the principal's original problem is equivalent to the following problem:

$$\max_{\{a,\lambda,\mu\}} \mathbb{E}v(x - s(x, a; \lambda, \mu)). \text{ s.t. } \mathbb{E}u(s(x, a; \lambda, \mu), \tilde{a}) \geq \underline{U} \text{ and } a \in \arg\max_{\tilde{a}} \mathbb{E}u(s(x, a; \lambda, \mu), \tilde{a})$$

We construct the Lagrangian

$$\begin{aligned} & L(s(x, a; \lambda^*(a), \mu^*(a)), a) \\ = \ & V(x - s(x, a; \lambda^*(a), \mu^*(a))) + \lambda^*(a)[U(s(x, a; \lambda^*(a), \mu^*(a)), a) - \underline{U}] + \mu^*(a)U_a(s(x, a; \lambda^*(a), \mu^*(a)), a) \end{aligned}$$

Note that for every given a, there is a unique $(\lambda^*(a), \mu^*(a))$ solving system (6); therefore, the above problem is equivalent to the following problem:

$$\max_{a \in a^{BR}(s(x,a;\lambda^*(a),\mu^*(a)))} \int L(s(x, a; \lambda^*(a), \mu^*(a))), a)f(x, a)dx$$

Therefore the solution a based on the FOA has the following property:

$$\begin{aligned} & \int \{v(x - s(x, a; .)) + \lambda^*(a)[u(s(x, a; .), a) - \underline{U}] + \mu^*(a)[u_a(s(x, a; .), a) + u(s(x, a; .), a)l_a(x, a)]\}f(x, a) \\ \geq \ & \int \{v(x - s(x, \hat{a}; .)) + \lambda^*(\hat{a})[u(s(x, \hat{a}; .), \hat{a}) - \underline{U}] + \mu^*(\hat{a})[u_a(s(x, \hat{a}; .), \hat{a}) + u(s(x, \hat{a}; .), \hat{a})l_a(x, \hat{a})]\}f(x, \hat{a}) \end{aligned}$$

for any $\hat{a} \neq a$.

As a result, both conditions (i) and (ii) are satisfied when the FOA is valid.

Sufficient part. If the optimal action a^* can be implemented by a $\mathcal{MH}$ contract $s(x, a; \lambda^*(a), \mu^*(a))$, we are done. Suppose there is an optimal action a^* being not implemented by any $\mathcal{MH}$ contract, but some contract w. We discuss two subcases: either LIC is binding under (w, a^*) or not binding (IR is always binding for the optimal contract).

(i) LIC is binding under (w, a^*).

By the optimality of w, IR constraint is binding as well, therefore,

$$\int u(w, a^*) f(x, a^*) dx - \underline{U} = 0$$

and

$$\int [u_a(w, a^*) + u(w, a^*) l_a(x, a^*)] f(x, a^*) dx = 0$$

Let a be the action satisfies conditions (i) and (ii). Therefore, the profit distance between (s, a) and (w, a^*) is:

$$\begin{aligned}
& \Delta\Pi \\
= & \int v(x - s(x, a; .)) f(x, a) dx - \int v(x - w) f(x, a^*) dx \\
= & \int \{v(x - s(x, a; .)) + \lambda^*(a)[u(s(x, a; .), a) - \underline{U}] + \mu^*(a)[u_a(s(x, a; .), a) + u(s(x, a; .), a) l_a(x, a)]\} f(x, a) \\
& - \int \{v(x - w) + \lambda^*(a^*)[u(w, a^*) - \underline{U}] + \mu^*(a^*)[u_a(w, a^*) + u(w, a^*) l_a(x, a^*)]\} f(x, a^*) dx \\
> & \int \{v(x - s(x, a; .)) + \lambda^*(a)[u(s(x, a; .), a) - \underline{U}] + \mu^*(a)[u_a(s(x, a; .), a) + u(s(x, a; .), a) l_a(x, a)]\} f(x, a) \\
& - \int \begin{pmatrix} v(x - s(x, a^*; \lambda^*(a^*), \mu^*(a^*))) + \lambda^*(a^*)[u(s(x, a^*; \lambda^*(a^*), \mu^*(a^*), a^*) - \underline{U}] \\ +\mu^*(a^*)[u_a(s(x, a^*; \lambda^*(a^*), \mu^*(a^*)), a^*) + u(s(x, a^*; \lambda^*(a^*), \mu^*(a^*), a^*) l_a(x, a^*)] \end{pmatrix} f(x, a^*) dx \\
\geq & \ 0
\end{aligned}$$

The second last step is because $s(x, a^*; \lambda^*(a^*), \mu^*(a^*))$ is the pointwise minimizer of the object

$$v(x - w) + \lambda^*(a^*)[u(w, a^*) - \underline{U}] + \mu^*(a^*)[u_a(w, a^*) + u(w, a^*) l_a(x, a^*)$$

Since $\Delta\Pi > 0$ results, we obtain a contradiction to the optimality of w.

(ii) LIC is not binding under (w, a^*)[12]

When LIC is not binding under w, we claim that

$$\begin{aligned} & \int [u_a(s(x,a^*;\lambda^*,\mu^*),a^*) + u(s(x,a^*;\lambda^*,\mu^*),a^*)l_a(x,a^*)]f(x,a^*)dx \\ \geq & \int [u_a(w,a^*) + u(w,a^*)l_a(x,a^*)]f(x,a^*)dx \\ > & \ 0. \end{aligned}$$

To show the above inequality, note that

$$l_a(x,a^*) = \frac{v_x(x-s) - \lambda u_w(s,a^*)}{\mu u_w(s,a^*)} - \frac{u_{aw}(s,a^*)}{u_w(s,a^*)}$$

therefore, if we let λ^* solve IR constraint, and for notational convenience, compress $s(x,a^*;\lambda^*(a^*,\mu),\mu) = s$, we have

$$\begin{aligned} & \int [u_a(s,a^*) + u(s,a^*)l_a(x,a^*)]f(x,a^*)dx - \int [u_a(w,a^*) + u(w,a^*)l_a(x,a^*)]f(x,a^*)dx \\ = & \int \{[u_a(s,a^*) - u_a(w,a^*)] + [u(s,a^*) - u(w,a^*)](\frac{v_x(x-s) - \lambda^* u_w(s,a^*)}{\mu u_w(s,a^*)} - \frac{u_{aw}(s,a^*)}{u_w(s,a^*)})\}f(x,a^*)dx \\ \geq & \int \{u_{aw}(s,a^*)\frac{u(s,a^*) - u(w,a^*)}{u_w(s,a^*)} + [u(s,a^*) - u(w,a^*)](\frac{v_x(x-s) - \lambda^* u_w(s,a^*)}{\mu u_w(s,a^*)} - \frac{u_{aw}(s,a^*)}{u_w(s,a^*)})\}f(x,a^*) \\ = & \int [u(s,a^*) - u(w,a^*)]\frac{v_x(x-s) - \lambda^* u_w(s,a^*)}{\mu u_w(s,a^*)} f(x,a^*)dx \\ = & \frac{1}{\mu}\int [u(s,a^*) - u(w,a^*)]\frac{v_x}{u_w} f(x,a^*)dx \end{aligned}$$

Note also that $\mu \geq 0$, so it suffices to show

$$\int [u(s,a^*) - u(w,a^*)]\frac{v_x(x-s)}{u_w(s,a^*)} f(x,a^*)dx \geq 0$$

for $\mu = \mu^*(a^*)$.

[12]Here we can rule out the case $U_a(w,a^*) < 0$ by assuming $U_a(\underline{w},0) = u_a(\underline{w},0) = 0$ for some fixed payment $\underline{w}$ solving $u(\underline{w},0) = \underline{U}$. This seems very reasonable. Even for linear cost fuction, we can let $c(a) = \mathbf{1}(a > 0)a$.

By the definition of w, we have

$$\begin{aligned}
& \int v(x-w)f(x,a^*)dx \\
> & \int L(s(x,a;\lambda^*(a),\mu^*(a))),a)f(x,a)dx \\
\geq & \int L(s(x,a^*;\lambda^*(a^*),\mu^*(a^*))),a^*),a^*)f(x,a^*)dx \\
= & \int v(x-s(x,a^*;\lambda^*(a^*),\mu^*(a^*)))f(x,a^*)dx
\end{aligned}$$

By the concavity, we have,

$$\begin{aligned}
& \int [u(s(x,a^*;\lambda^*(a^*),\mu^*(a^*)),a^*)-u(w,a^*)]\frac{v_x(x-s)}{u_w(s,a^*)}f(x,a^*)dx \\
\geq & \int [s(x,a^*;\lambda^*(a^*),\mu^*(a^*))-w]v_x(x-s(x,a^*;\lambda^*(a^*),\mu^*(a^*)))f(x,a^*)dx \\
\geq & \int [v(x-w)-v(x-s(x,a^*;\lambda^*(a^*),\mu^*(a^*)))]f(x,a^*)dx \\
\geq & \ 0
\end{aligned}$$

As a result, for effort level a^*, if w is the optimal contract implementing a^* and $\int[u_a(w,a^*)+u(w,a^*)l_a(x,a^*)]f(x,a^*)dx>0$, then

$$\int [u_a(s(x,a^*;\lambda^*(a^*),\mu^*(a^*)),a^*)+u(s(x,a^*;\lambda^*(a^*),\mu^*(a^*)),a^*)l_a(x,a^*)]f(x,a^*)dx>0$$

as well. Hence, the solution to system (6) must result $\mu^*(a^*)=0$. Based on $\mu^*(a^*)=0$, we can construct the same Lagrangian as subcase (i). Finally, we obtain $\Delta\Pi>0$ as a contradiction. This implies that $(s(x,a^*;\lambda^*(a^*),\mu^*(a^*)),a^*)$ should be the solution to the principal's original problem. Q.E.D. ■

The above theorem provides an important method to justify the validity of FOA and computation of the solution. Because the value function $\mathbb{E}v(x-s(x,a;\lambda^*(a),\mu^*(a)))$ of relaxed problem (with LIC constraint) or the Lagrangian $L(s(x,a;\lambda^*(a),\mu^*(a)),a)$ is a continuously deferentiable function of a, to find the maximizer(s) of them, we only need to focus on the steady points where the adjoint equations hold, or the boundary points. Therefore, the two-step procedure to check the validity of FOA or solve the principal-agent problem can be stated as follows:

(i) Find the maximizer of the expectation of Lagrangian, $\mathbb{E}L(s(x, a; \lambda^*(a), \mu^*(a)), a)$ by using the adjoint equation, together with LIC and IR constraints.

(ii) Check whether the maximizer a^* is a fixed point of the best response correspondence $a^{BR}(s(x, a; \lambda^*(a), \mu^*(a)))$.

If the above two steps are passed, then FOA is valid and we find the solution, if not, FOA is not valid. This theorem and algorithm can be applied to a very general environment.

2.5.3 Implications for A General Non-linear Optimization

It is valuable to point out the applicability of the fixed point approach to a general non-linear optimization problem. Consider a constrained optimalization problem like

$$(P5) \quad \max_{\{a,w\}} V(w, a)$$

s.t.

$$\begin{aligned} a &\in \arg\max_a U(w, a) \quad (\text{IC"}) \\ \underline{U} &\leq U(w, a) \qquad\quad (\text{IR"}) \end{aligned}$$

where we assume both $V(w, a)$ and $U(w, a)$ are continuously differentiable over a compact space $\mathbb{A} \times \mathcal{W}$.

For problem (P5), it is well-known that there might exist some duality gap when the constrained set is not convex or $U(w, a)$ is not globally concave in a. Our fixed-point approach provides a necessary and sufficient condition to judge whether (P5) can be solved by only using the local first order condition of IC".

Theorem 2.6: *If $V(w, a)$ and $U(w, a)$ are continuously differentiable over a compact space $A \times W$, then (P5) can be solved based on FOA if and only if there exist some $s(a, \lambda, \mu) \in \arg\max_w V(w, a) + \lambda[U(w, a) - \underline{U}] + \mu U_a(w, a)$ and $a^* \in A$ such that (i) $a^* \in a^{BR}(s(a^*, \lambda(a^*), \mu(a^*)))$ and (ii) $a^* \in \arg\max_a V(s(a, \lambda(a), \mu(a)), a) + \lambda(a)[U(s(a, \lambda(a), \mu(a)), a) - \underline{U}] + \mu(a) U_a(s(a, \lambda(a), \mu(a)), a)$,* *where* $\lambda(a) \geq 0$ *and* $\mu(a) \geq 0$ *solves the system*

$$\begin{cases} \lambda[U(s(a,\lambda,\mu),a) - \underline{U}] = 0 \\ \mu U_a(s(a,\lambda,\mu),a) = 0 \end{cases} \tag{2.7}$$

with the complementary conditions $\lambda = 0$ *if* $U(s(a,\lambda,\mu),a) > \underline{U}$ *and* $\mu = 0$ *if* $U_a(s(a,\lambda,\mu),a) > 0$.

Proof. (The proof is quite similar to the proof of Theorem 5.)

Necessary part. If (P5) can be solved by FOA, then the solution must have properties $s(a,\lambda,\mu)$ such that $a \in a^{BR}(a, s(a,\lambda,\mu))$ and (λ,μ) solving the system of (7) with the complememtary conditions. Therefore, to search the optimal (a,λ,μ) is equivalent to choose a to maximize the Lagraingian,

$$L(s(a,\lambda(a),\mu(a)),a) = V(s(a,\lambda(a),\mu(a)),a)+\lambda(a)[U(s(a,\lambda(a),\mu(a)),a)-\underline{U}]+\mu(a)U_a(s(a,\lambda(a),\mu(a)).$$

Sufficient part. First of all, for problem (P5), at least of IC" or IR" must hold for the optimal solution (suppose not, we can find a profitable deviation $\tilde{w}$ such that $u(\tilde{w},a) + \varepsilon = u(w,a)$ where a constant ε is chosen to make at least one of IC" or IR" constraints binding). We show that there is no profitable deviation from (i) and (ii). By contradiction, suppose there is an optimal $w \neq s(a,\lambda,\mu)$ implementing the optimal $a \notin a^{BR}(s(a,\lambda(a),\mu(a)))$, we discuss several subcases as follows.

Subcase (i): both IC" and IR" binding

In this subcase, we have the profit distance,

$$\begin{aligned} \Delta\Pi &= V(w,a) - V(s(a^*,\lambda(a^*),\mu(a^*)),a^*) \\ &= V(w,a) + \lambda(a)[U(w,a) - \underline{U}] + \mu(a)U_a(w,a) - L(s(a^*,\lambda(a^*),\mu(a^*)),a^*) \\ &< L(s(a,\lambda(a),\mu(a)),a) - L(s(a^*,\lambda(a^*),\mu(a^*)),a^*) \\ &< 0, \end{aligned}$$

a contradiction to the optimality of (w,a).

Subcase (ii): only IC" binding, IR" not binding.

In this subcase, choose $s(a, \lambda, \mu) = w$, so we have $U(w, a) = U(s(a, \lambda, \mu), a)$ and

$$U_a(s(a, \lambda, \mu), a) = U_a(w, a) > 0$$

Therefore the solutions to system (7) will be $\lambda(a) > 0$ and $\mu(a) = 0$. We can do the same proof by construction as subcase (i).

Subcase (iii): only IR" binding, IC" not binding.

It is similar to subcase (ii), except now $\lambda(a) = 0$ and $\mu(a) > 0$.

Putting all pieces together, we show that hte two conditions in Theorem 6 are sufficient to solve problem (P5). Q.E.D. ■

The above theorem holds under a very general environment. It does not require $s(a, \lambda, \mu)$ to be unique, nor the continuities of $\lambda(a)$ or $\mu(a)$ in a. However, because the second condition in the theorem seems less tractable, we can provide some sufficient condition instead, where the global concavity can be relaxed by the following three conditions: (i) $a \in a^{BR}(s(a, \lambda, \mu))$, i.e., the best response admits a fixed point, where $s(a, \lambda, \mu) = \arg\min_w V(a, w) + \lambda[U(a, w) - \underline{U}] + \mu U_a(a, w)$ is the unique minimizer for given (a, λ, μ); (ii) there exists (λ, μ) such that two complementary conditions are met as follows:

$$\lambda \geq 0 \text{ and } \lambda = 0 \text{ if } U(a, w) > \underline{U}$$

and

$$\mu \geq 0 \text{ and } \mu = 0 \text{ if } U_a(a, w) > 0$$

and (iii) every a could be a fixed point by adjusting (a, λ, μ) properly.

Remark 2.6: *There are several differences between (P5) and the principal-agent problem: (i) $s(a, \lambda, \mu)$ is a number for given a, so it is hard to have both IC and IR constaints binding*[13]*; (ii) in Theorem 5, we make conditions to assure the uniqueness of $s(x, a; \lambda, \mu)$, and $\lambda(a)$ and $\mu(a)$ have nice properties under those conditions, but in problem (P5), those properties are not available. For example, when the agent's utility is separable $U(a, w) = u(w)c_1(a) - c_2(a)$, then*

[13] *For example, suppose $V(a, w) = a - w$, $U(a, w) = 2\sqrt{w}a - a^2$, and $\underline{U} = 0$. In this case, IR will not binding at optimum. The solution is $\mu^* = a^* = \frac{1}{2}$, $\lambda^* = 0$ and $w^* = \frac{1}{4}$.*

μ cannot adjust effort a when IR is binding.

2.6 Conclusions and Discussion

The global concavity of the agent's utility is a too strong condition for validity of FOA. This paper provides a new approach to justify FOA based on the existence of a fixed point of best reponse correspondence, which is weaker than the global concavity condition since the latter implies the existence of a fixed point. Based on this approach, we can check the validity of FOA directly by checking the existence of a fixed point against a $\mathcal{MH}$ contract, which makes it more transparent to judge the validity in practice. And using our approach, we restore the validity of FOA when the payment schedule is not concave due to the existence of some limited liability constraint. For a concrete example, FOA can be applied to a useful case such as a normal distribution combined with a CRRA utility, under some specifications of effort space. We also find some nice properties still hold in the situation where the agent's utility is non-separable. Particularly, we provide a necessary and sufficient condition for FOA to be valid, which also sheds light on a general non-linear optimization without a convex constrained set. In the future, we expect that the relaxation of global concavity would bring some additonal chances to find a set of more intuitive conditions to validate FOA. At least, the numerical algorithm can be built on this ground even if an explicit solution is not available.

Bibliography

[1] A. Araujo, and H. Moreira, A general Lagrangian approach for non-concave moral hazard problems, Journal of Mathematical Economics, 35, (2001), 17-39.

[2] L. D. Brown, Fundamentals of Statistical Exponential Families: With Applications in Statistical Decision Theory, volume 9 of Institute of Mathematical Statistics Lecture Notes Monograph Series. Hayward, California 1986, .

[3] G. Carlier, and R.-A. Dana, Existence and monotonicity of solutions to moral hazard problems Journal of Mathematical Economics, 41 (2005), 826-843

[4] E. Alvi, First-order approach to principal-agent problems: a generation, The General Paper on Risk and Insurance Theory, 22. (1997), 59-65

[5] P.S Faynzilberg, P. Kumar, On the generalized principal-agent problem: Decomposition and existence results, Rev. Econ. Design 5 (2000), 23-58.

[6] S. Grossman, O. Hart, An analysis of the principal–agent problem, Econometrica 51 (1983) 7–45.

[7] B. Holmstrom, Moral hazard and observability, Bell J. Econ. 10 (1979) 74–91.

[8] B. Holmstrom, P. Milgrom, 1987, Aggregation and linearity in the provision of intertemporal incentives, Econometrica 55 (1987) 303–328.

[9] I. Jewitt, Justifying the first order approach to principal-agent problems, Econometrica, 56 (1988), 1177-1190.

[10] I. Jewitt, O. Kadan, J. Swinkels, Moral Hazard with Bounded Payments, Journal of Economic Theory, (2008) forthcoming.

[11] R. Ke, 2008, Identifying Contract Optimuality Non-parametrically with Moral Hazard: First Order Approach and Statistical Inference, 2008, Job Market Paper, Department of Economics, MIT.

[12] S.K. Kim, Efficiency of an information system in an agent model, Econometrica, 63 (1995), 89-102.

[13] P. R. Milgrom and I. R. Segal, Envelope Theorems for Arbitrary Choice Sets, Econometrica, Vol. 70 (2002). 583-601,

[14] J. Mirrlees, The optimal structure of authority and incentives within an organization, Bell J. Econ. 7 (1976) 105–131.

[15] J. Mirrlees, 1997, Information and Incentives: the economics of carrots and sticks, Econ. Journal 107(1997), 1311-29.

[16] J. Mirrlees, 1999, Theory of moral hazard and unobservable behavior: Part I, Rev. Econ. Studies, 66 (1999), 3-21.

[17] J. Mirrlees and Y. Zhou, Principal-Agent Problems Revisted Part I: Static Models, Working Paper, Chinese University of HongKong, 2006.

[18] R.B. Myerson, 1982 Optimal coordination mechanisms in generalized principal-agent problems, Journal of Mathematical Economics, 11(1982), 67-81.

[19] W. Rogerson, 1985, The first order approach to principle-agent problems, Econometrica, 53 (1985), 1357-1368.

[20] H. Scarf., The approximation of fixed points of a continuous mapping. SIAM J. Appl. Math., 15 (1967), 1328–1343.

Chapter 3

Auctioning Social Surplus: First Best Bayesian Mechanism with Ex Post Invidual Rationality

3.1 Introduction

Suppose a social planner wants to allocate resources efficiently ex post, and he also subjects to budget balance ex post. Basically, we have the following seven combinations of interest (See Table 1).

Table. 1 Combinations of Properties of Ex Post Efficient Mechanisms.

	Individual Rationality	Incentive Compatibilty	Budget Balance
Ex ante	(1)	(4)	N.A.
Interim	(2)	(5)*	N.A.
Ex Post	(3)*	(6)	(7)*

The well-known classic Vickery-Clark-Groves (VCG hereafter) mechanism can be either ex post individually rational or ex post budget balance, but cannot be both. Holmstrom (1977) provides a necessary and sufficient condition for ex post budget balance. Another well-known mechanism by Arrow (1979) and d'Aspremont Varet-Gared (1979) is ex post budget balance but might not be interim individually rational. The celebrated result by Myerson-Satterthewaite

(1983) shows the impossibility of having combination (2), (5) and (7) in the above table, when the agent's preferences are linear and the seller owns the object initially. Several other authors provide some additional insights with slight change of assumptions in endowment or divisibility of the good, where conterveiling incentive matters (Crampton, Gibbons and Klemperer, 1987; McAfee, 1992, *among others*). In their story, combinations (2), (5) and (7) might be possible.

Recently, Ledyard and Palfrey (2006) has fully characterized the solution to an interim socially efficient mechnism, i.e., combination (2), (5) and (7). And Krishna and Perry (1998) has generalized the classic Vickery-Clark-Groves mechanism to discuss the sufficient and necessary condition for existence of interm socially efficient mechanism. Unfortunately, Ely and Cheung (2003) shows combination of (3), (6) and (7) is impossible.

This paper is considering the possibility of the combination (3), (5) and (7), star marked in table 1. Our interest on ex post individual rationality is motivated by the fact that, the participant may opt out of ex post once the ex post allocation can not give him higher utility than the outside reservation utility. The consequence of this option is subtle: because the participant expects that somebody will opt out, his strategy at the interim stage will change given that the designer is not a budget breaker. Therefore, the interim incentive compatible constraint will be different from the mechanism without an option to quit. We first show that the condition for a mechanism being robust against ex post option to quit needs to satisfy an ex post individual rationality constraint. Then we check the condition under which an ex post efficient allocation exists, satisfying incentive compatible (interim), budget balance (ex post) and individual rational (ex post). The importance and significance of the ex post IR constraint have been discussed by Dudek, Kim and Ledyard (1995). Recently, there has been increasing literature on exploring robustness of mechanism design, which requires ex post IR, even ex post incentive compatibility (Morries, 2003; Bergemann, S. Morris, 2005). Our strategy of proof is to construct a concrete mechanism or auction which has these good properties.

Our mechanism or auction has a very intuitive intepretation, and seems simple in terms of pragmatic use, which can be regarded as an auction to sell social surplus (if any). The game can be decomposed into two stages: (i) the first stage, all bidders compete for the right to own entitlement (license). The owner of entitlement can charge full consumer surplus to the other bidders, but in order to win the entitlement, he needs to pay a lump sum transfer to the others;

(ii) the second stage is a trivial game, where the bidders are charged according to their first stage reports. By constructure, this mechanism is always ex post budget balance. We show it will be also ex post individual rational under some conditions, depending on the context. In private good cases, it will be ex post individual rational whenever VCG runs expected social surplus. However, in public good cases, it sensitively depends on the flexibility of total supply and the number of participants.

We also use this mechanism to explore the possibility of no trade (Myerson-Satterthewaite theorem) in a generalized environment, where divisibility of trade object, distribution of initial endownment and concavity of preference have receieved full consideration. We show trade could happen when either endownment is extreme or relatively symmetric when preference is concave. But if we impose the ex post IR constraint instead of its interim counterpart, even though the endownment is symmetric, no trade happens, in constrast to the existence of efficient partner dissolving mechanism (Clampton, Gibbons, Klemperer 1987). We provide a set of conditions for non-existence of ex post socially efficient trade mechnanisms.

The closest paper to ours is Dudek, Kim and Ledyard (1995) (DKL, hereafter), though our discovery is independent of theirs and the motivation is quite different. They propose an ex post individually rational mechanism to allocate a single unit private good among agents whose reservation utility is type-independent[1]. However, our paper proposes an explicit auction format for privitization[2], both for private good and public good, either for endogeneous or exogeneous quantity and derives the necessary and sufficient condition for the existence of ex post individually rational mechanisms. In addition, we also discuss how flexibity of endowment and type-dependent reservation affect the existence of the ex post individual rationality.

The connections of our paper to standard auctions also can be seen in the following senses. In private good cases (whether or not the quantity of supply is fixed), compared with the standard auction, the present auction format generates a risk-free revenue (same as any efficient allocation) to the seller (Eso, 1999), and meets all bidders ex post individual rationality without any side-payment. We can explicitly characterize the bidding strategy under asymmetric

[1]Their paper is also motivated by the importance of a transfer among agents.

[2]The bidding function in our auction will be consistent with DKL's mechanism when either in the symmetric independent private case or the two-bidder asymmetric independent private case. We also discuss the relevance of choice of an auction format, especially in public good cases.

distribution; as a contrast, finding the solutions to the standard asymmetric first price auction is very complicated. In public good cases, our auction connects with multi-unit auction or divisible good auction (Wilson 1979, Ausubel, 2004, Wang and Zender, 2001; *among others*), but there are several differences worth mentioning. First, the bidding strategy in our auction is much more tractable, so that we can work out the solution explicitly, for general quansi-concave preferences. As a consequence, we can check the revenue easily, while in a standard multi-unit auction, it is hard to invert the demand function and work out the formula of expected revenue in general quasi-concave utility form. Second, the seller can earn a risk-free revenue, which is the highest revenue among all efficient allocations, and the bidders' ex post individual rationalities are met when the number of bidders is two. Third, under symmetric situation, the allocation in our auction is always efficient, while the standard auction may not be (Ausubel, 2004).

The paper then is organized as follows. Section 2 describes the basic setting of the environment and associated solution concepts. Section 3 deals with private good and Section 4 deal with public good. In section 5, we propose a specific auction to implement the mechanisms we proposed in section 3 or 4. We finally summarize the findings in Section 5. Technical proofs are in the Appendices.

3.2 Preliminaries

3.2.1 Classical environment

Environment: There are n players. Each i's utility is $u_i = v_i(x_i, \theta_i) - m_i$, where $v_i(x_i, \theta_i)$ is utility from consumption of good x_i, and m_i is money payment[3], where θ_i is only known by player i, which is a random variable drawn from some set $\Theta_i \subset \mathbb{R}$, with cumulation distribution function (c.d.f) $F_i(.)$ and probability distribution function (p.d.f) $f_i(.)$, but the distribution is common knowledge. And we assume $v_i(.,.)$ is an increasing function of x_i and θ_i, and satisfies

[3] This preference can be regarded as a general form of common value. $v_i(x, \theta_i) = \mathbb{E}[u(x, S)/S = \theta_i]$, where S is some random variable common to all players, and θ_i is a private signal. Many public good consumptions have such a feature, such as hospital space, energy, public transportation and so on. This specification of utility function is general enough to cover the private value and public good situation, associated with a certain form of cost function. In a pure public good case, all $x_i = x$. One may generalize the utility function to $v_i(h_i(\mathbf{x}), \theta_i)$ with $h_i : \mathbb{R}^n \to \mathbb{R}$, as Mookherjee and Reichelstein (1989).

supermoduarity, i.e., $v_i(x_i, \theta_i) + v_i(x_i', \theta_i') \geq v_i(x_i', \theta_i) + v_i(x_i', \theta_i)$ for any x_i', x_i, θ_i', θ_i (this means, the preference appears an increasing difference). We use $\mathbf{v} = (v_1, v_2, ..., v_n)$ to denote the collection of all variables and use subcript $-i$ to indicate all individuals except i ($\boldsymbol{\theta}$ and $\mathbf{x}$ have the same treatment).

Suppose the cost of building $\mathbf{x}$ takes a general form $c(\mathbf{x})$, where $\mathbf{x} = (x_1, x_2, ..., x_n)$ is the bundle of the goods. We assume $c(\mathbf{x})$ is non-decreasing with any arguments, and at least for some i, it is strictly increasing. To capture the possible externality, we assume $c(x_1, x_2, ..., x_n)$ appears piecewise submodular, namely, $c(x_i, x_j', x_{-ij}) + c(x_i', x_j, x_{-ij}) \geq c(x_i, x_j, x_{-ij}) + c(x_i, x_j, x_{-ij})$. This cost function can be a pure public good case if $c(x_1, x_2, ..., x_n) = c(\max x_i)$ such as national security, or partially/fully excludible public good $c(x_1, x_2, ..., x_n) = c(\sum x_i)$. Particularly, we allow $c(\mathbf{x})$ to have an infinite marginal cost at some exogenous point, which corresponds to situations where the total endownment is given (resambling a pure exchange economy in text books).

Allocation Rule: Let $\mathbf{X} \subset \mathbb{R}^n$ be an arbitrary set of allocations (feasible), and let $\mathbf{x} : \Theta \rightarrow \mathbf{X}$ be the social choice rule. Throughout this paper, we are interested in the following allocation rule, which is called optimal if

$$\mathbf{x}^*(\boldsymbol{\theta}) \in \begin{cases} \arg\max_{\mathbf{x}} \sum v_i(x_i, \theta_i) - c(\mathbf{x}) \text{ given state } \boldsymbol{\theta}\text{, if quantity is endogenous} \\ \arg\max_{\mathbf{x}} \sum v_i(x_i, \theta_i) \text{ s.t. } c(\mathbf{x}) \leq \bar{c} \text{ given state } \boldsymbol{\theta}\text{, if quantity is exogenous} \end{cases}$$

Unless pointed out explicitly, $S(\boldsymbol{\theta})$ is used to denote the social surplus of either case, i.e. $S(\boldsymbol{\theta}) = \max_{\mathbf{x}} \sum v_i(x_i, \theta_i) - c(\mathbf{x})$ or $S(\boldsymbol{\theta}) = \max_{\mathbf{x}} \sum v_i(x_i, \theta_i)$ $s.t. c(\mathbf{x}) \leq \bar{c}$. The following proposition is a standard result based on supermodularity.

Proposition 3.1: *Assuming $v_i(x_i, \theta_i)$ appears an increasing difference, and $c(x)$ appears a decreasing difference, then (i) in the endogenous quantity case, the $S(\theta)$ is supmodular and $x_i^*(\theta_i, \theta_j, \theta_{-ij})$ is non-decreasing in θ_i and θ_j. (ii) In the exogeneous quantity case, $x_i^*(\theta_i, \theta_j, \theta_{-ij})$ is non-decreasing in θ_i; in addition if $\sum x_i = \bar{x}$ and $v(x_i, \theta_i)$ differentiable in x_i, then $S(\theta)$ is submodular and $x_i(\theta_i, \theta_j, \theta_{-ij})$ is non-increasing with θ_j.* (**Proof see Appendix A1**)

Incentive compatibility: There are two concepts of interest, interim incentive compatible (IIC) or ex post incentive compatible (EPIC). They are defined as follows:

Definition 3.1: *A direct mechanism* $< x, M >$ *with* $x : \Theta \to X$ *being allocation rule and* $M : \Theta \to R^n$ *being payment rule, is interim incentive compatible if,*

$$\begin{aligned} \text{IC(Bayesian)} \quad &: \quad \mathbb{E}_{\theta_{-i}} v_i(x_i(\theta_i, \theta_{-i}), \theta_i) - \mathbb{E}_{\theta_{-i}} M_i(\theta_i, \theta_{-i}) \\ &\geq \quad \mathbb{E}_{\theta_{-i}} v_i(x_i(\tilde{\theta}_i, \theta_{-i}), \theta_i) - \mathbb{E}_{\theta_{-i}} M_i(\tilde{\theta}_i, \theta_{-i}) \ \forall \tilde{\theta}_i, \theta_i \in \Theta_i \end{aligned}$$

The above concept means truth-telling is a Bayesian Nash Equilibrium (B.N.E.) strategy at the interim stage. If we use ex post equilibrium as solution concept, the incentive compatibility will be stronger, as follows.

Definition 3.2: *A direct mechanism* $< x, M >$ *with* $x : \Theta \to X$ *being allocation rule and* $M : \Theta \to R^n$ *being payment rule, is ex post incentive compatible if,*

$$\begin{aligned} \text{IC (Ex post)} \quad &: \quad v_i(x_i(\theta_i, \theta_{-i}), \theta_i) - M_i(\theta_i, \theta_{-i}) \\ &\geq \quad v_i(x_i(\tilde{\theta}_i, \theta_{-i}), \theta_i) - M_i(\tilde{\theta}_i, \theta_{-i}) \ \forall \tilde{\theta}_i, \theta_i \in \Theta_i \ \theta_{-i} \in \Theta_{-i} \end{aligned}$$

Individual Rationality: Paralleling IC constraint, the participation constraint also can be defined as interim *individual rationality,*

$$\text{IR (Interim): } \mathbb{E}_{\theta_{-i}}[v_i(x_i(\theta_i, \theta_{-i}), \theta_i) - M_i(\theta_i, \theta_{-i})] \geq \underline{u}_i(\theta_i), \ \forall \theta \in \Theta$$

or *ex post individual rationality,*

$$\text{IR (Ex post): } \quad v_i(x_i(\theta_i, \theta_{-i}), \theta_i) - M_i(\theta_i, \theta_{-i}) \geq \underline{u}_i(\theta_i), \ \forall \theta \in \Theta$$

where $\underline{u}_i(\theta_i)$ is reservation utility, usually normalized to be zero if it is type independent. We will discuss this later.

Budget Balance: Ex post budget balance is defined as,

$$\text{BB (Ex post): } \sum M_i(\theta_i, \theta_{-i}) = \begin{cases} c(\mathbf{x}) \text{ if endowment is endogenous} \\ 0 \text{ if endowment is exogenous} \end{cases}$$

It is acceptable that under some situations, budget balance is not a problem, like union negatiation (government may subsidize one of the parties), while in many other situations, budget balance ex post is a constraint hard to break through[4].

It is well known (Groves (1971), Clark (1973), Vickery (1961), Laffont and Green (1977) Holmstrom, 1979 *among others*) that VCG mechanism with payment rule:

$$M_i^V(\hat{\theta}_i, \hat{\theta}_{-i}) = S(\underline{\theta}_i, \hat{\theta}_{-i}) - \sum_{j \neq i} v_j(x_j^*(\hat{\theta}_i, \hat{\theta}_{-i}), \hat{\theta}_j) + c(\mathbf{x}^*(\hat{\theta}_i, \hat{\theta}_{-i})) \tag{3.1}$$

will implememt the optimal allocation rule in terms of both ex post IC and ex post IR. However, it is also well known that VCG defined as (1) can not be budget balance ex post. Another seminal mechanism propsed by Arrow (1979) and d'Aspremont-Garad-Varet (1979), (AGV hereafter) is budget balance ex post by constructure, but it may not be individual rational, either interim or ex post. The following two lemmas state the positve and negative sides of the existence of desirable mechanisms.

Lemma 3.1: *(Krishna and Perry, 1998; Krishna, 2002) There exists an efficient, incentive compatible (interim) and individual rational (interim) mechanism that balances the budget if and only if the VCG mechanism results in an expected surplus.*

$$\mathbb{E}\Delta \equiv \sum \mathbb{E}_{\theta_{-i}} S(\underline{\theta}_i, \theta_{-i}) - (n-1)\mathbb{E}_{\boldsymbol{\theta}} S(\theta_i, \theta_{-i}) \geq \sum \underline{u}_i(\underline{\theta}_i)$$

where $\underline{\theta}_i = \arg\min_{\theta_i} E_{\theta_{-i}} S(\theta_i, \theta_{-i}) - \underline{u}_i(\theta_i)$.

Krishna and Perry (1998) generalizes the conclusion that, among all individual rational (interim) and incentive compatible (interim) mechanisms, VCG maximizes the payment. If VCG runs deficit, no other mechanism can be better off.

However the negative result appears if instead using a stronger solucton concept as EPIC and EPIR. The following lemma states the impossibility (Chung and Ely, 2003).

Lemma 3.2: *(Chung and Ely, 2003) If the utility is increasing difference, there does not exist an optimal allocation that is ex post IR, ex post IC and ex post budget balance.*

The existing literature tells that if ex post BB is imposed, one has to compromise, yielding

[4]To taking an extreme example, regarding all humen beings as pariticipants in a global mechanism, then no third part can subsidize people living on earth. Of course ex post budget constraint may be too strong, sometimes it can be relaxed to a feasible constraint (Palfery and Ledyard,1999), which is non-defict constraint, $\sum M_i(\theta_i, \theta_{-i}) \geq c(\mathbf{x})$, where money left over is burned.

IR and IC to interim sense. Interim IC is understandable since it is the best play given the information set the agent achieves, but interim IR exproiates the participant's right to quit even if participation is indeed not profitable ex post. This paper attents to discover what happens if this right is entitled to the players, and then they can "vote by foot" at the last minute when all types are revealed. In order to incorparate this feature, we extend the classical Bayesian game to a sequential game.

3.2.2 Effectiveness of option to quit ex post

The game is modified as the following:

Game: (i) Designer anounces the allocation rule of game $<\mathbf{x},\mathbf{M}>$ with $\mathbf{x}$ being the allocation rule and $\mathbf{M}:\Theta\rightarrow\mathbb{R}^n$ with $M=(M_1,M_2,...,M_n)$ as the rule of monetary transfer when the report is $\hat{\boldsymbol{\theta}}$. (ii) Knowing his own type θ_i only, each decides to be in or out of the mechanism; if he chooses "in", he needs to report his type (simulatenously with all opponents); (iii) all reports $\hat{\boldsymbol{\theta}}\in\Theta$ are published and $<\mathbf{x},\mathbf{M}>$ is proposed; (iv) given the mechanism, for individual i, if he accepts $(x_i(\hat{\theta}_i,\hat{\theta}_{-i}),M_i(\hat{\theta}_i,\hat{\theta}_{-i}))$ then he is implemented by what he accepts, or he can leave permanently (without any penality), obtaining an outside option $\underline{u}_i(\theta_i)$; (iv) finally, an up-to-date mechanism $<\mathbf{x},\mathbf{M}>^*$ is enforced among all individuals who have not left. The allocation and payment in $<\mathbf{x},\mathbf{M}>^*$ might be different from the original mechanism $<\mathbf{x},\mathbf{M}>$ because some payment and allocations might no longer be plausible due to some participants' quiting. We denote this game by $\Gamma(n,\mathbf{v},\mathbf{x},\mathbf{M},\boldsymbol{\Theta})$.

Equilibrium: The equilibrium here not only requires each player to tell the truth as a Bayesian Nash Equilibrium, but also requires that the expectation operation at the interim stage should be based on ex post participation set $\Theta^*\subset\Theta$, where $\Theta^*=\times_i\{\text{all } \theta_i\in\Theta_i\text{: } v_i(x_i(\theta_i,\theta_{-i}),\theta_i)-M_i(\theta_i,\theta_{-i})\geq\underline{u}_i(\theta_i)\}$. The complication here is that Θ^* is endogenized by rule of game $<\mathbf{x},\mathbf{M}>$, and affects the enforceability of $<\mathbf{x},\mathbf{M}>$ in turn. For example, if a proposed mechanism results in $\Theta^*=\Phi$, then $<\mathbf{x},\mathbf{M}>$ loses any power to be effective.

Implementability: A direct mechanism $<\mathbf{x},\mathbf{M};\boldsymbol{\Theta}^*>$ such that $\mathbf{x}:\Theta^*\rightarrow\mathbb{R}^n$ and $\mathbf{M}:\Theta^*\rightarrow\mathbb{R}^n$ with $\Theta^*\subset\boldsymbol{\Theta}$ is said to be implementable if

(i) for any $i \in \mathcal{N}^* \equiv |\boldsymbol{\Theta}^*|$, truth-telling is a BNE, i.e.

$$\begin{aligned} \text{IC(Bayesian)} \quad : \quad & \mathbb{E}_{\theta^*_{-i}} v_i(x_i(\theta_i, \theta_{-i}), \theta_i) - \mathbb{E}_{\theta^*_{-i}} M_i(\theta_i, \theta_{-i}) \\ \geq \quad & \mathbb{E}_{\theta^*_{-i}} v_i(x_i(\tilde{\theta}_i, \theta_{-i}), \theta_i) - \mathbb{E}_{\theta^*_{-i}} M_i(\tilde{\theta}_i, \theta_{-i}) \ \forall \tilde{\theta}_i, \theta_i \in \Theta^*_i \end{aligned}$$

(ii) Participation set is consistent, i.e.,

$$v_i(x_i(\theta_i, \theta_{-i}), \theta_i) - M_i(\theta_i, \theta_{-i}) \geq \underline{u}_i(\theta_i) \text{ for } \forall (\theta_i, \theta_{-i}) \in \Theta^*_i$$

Under this setting, a lot of mechanisms are no longer implementable even though they are initially implementable without requirement (ii). For example, AGV is one of the mechanism that fails (ii), as example 1 shows.

Regarding the complication of the above equilibrium, we particularly are interested in a typical implementable mechanism that all participants will not leave ex post, i.e., $\boldsymbol{\Theta}^* = \Theta$. This is a *full participation* mechanism. Therefore, the linkage between the current setting and classical mechaism design literature is obvious, throung the following theorem.

Theorem 3.1: *A direct mechanism $< x, M; \Theta^* >$ is full participation B.N.E. implementable if and only if (i) $< x, M; \Theta >$ is intereim incentive compatible, i.e.,*

$$\begin{aligned} \text{IC (Bayesian)} \quad : \quad & \mathbb{E}_{\theta_{-i}} v_i(x_i(\theta_i, \theta_{-i}), \theta_i) - \mathbb{E}_{\theta_{-i}} M_i(\theta_i, \theta_{-i}) \\ \geq \quad & \mathbb{E}_{\theta_{-i}} v_i(x_i(\tilde{\theta}_i, \theta_{-i}), \theta_i) - \mathbb{E}_{\theta_{-i}} M_i(\tilde{\theta}_i, \theta_{-i}) \ \forall \tilde{\theta}_i, \theta_i \in \Theta_i \end{aligned}$$

and (ii) each individual's ex post IR constraint is met, i.e.,

$$\text{IR (Ex post):} \qquad v_i(x_i(\theta_i, \theta_{-i}), \theta_i) - M_i(\theta_i, \theta_{-i}) \geq \underline{u}_i(\theta_i), \ \forall \theta \in \Theta$$

Proof. If part: use backward induction, if ex post IR is met, based on information set $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, then not leaving is a best response to other players, given other individuals' not leaving. Given this future best response, back to the interim stage, IC constraint is consistent with support condition over $\boldsymbol{\Theta}^* = \Theta$; therefore, truth-telling is a B.N.E based on the information set θ_i. Therefore, $< \mathbf{x}, \mathrm{M}; \boldsymbol{\Theta} >$ is implementable.

Only if: If ex post IR condition is not met, at least some players are leaving, then it is not full participation implementable. Q.E.D. ■

Based on the above theorem, giving the optional right to the participant ex post can be thought of as putting the addtional ex post IR constraint on the program. For conceptual convenience, we define the ex post socially efficient mechanism below:

Definition 3.3: *A direct mechanism (x, M) is called ex post soically efficient if it maximizes "social surplus" ($x \in x^*$) and at the same time satisfies IC (Bayesian), IR (Ex post) and BB (Ex post).*

The natural question arises here is, how significant is the difference that this extra constraint brings in? Intuitively, participants under ex post individually rational mechanisms, seem to have higher expected utility than under interim IR since they can always leave for higher payoff by voting by foot. Therefore, in order to "bribe" the players not to leave in equilibrium, the designer seemingly needs to have more surplus. In other words, is the expect surplus of VCG enough for such a kind of mechanism design?

3.2.3 Characterization of IC and Budget Balance Mechanism

Before proceeding, we characterize the incentive compatible condition first, which is a standard result in the existing literature. Let

$$m_i(z_i) = \int_{\theta_{-i}} M_i(z_i, \theta_{-i}) f_{-i}(\boldsymbol{\theta}_{-i}) d\boldsymbol{\theta}_{-i}$$

be the expected payment when individual i reports z_i, therefore, in the social allocation game, for individual θ_i, his expected utility is

$$U_i(z_i, \theta_i) = \int_{\theta_{-i}} v_i(x_i^*(z_i, \theta_{-i}), \theta_i) f_{-i}(\boldsymbol{\theta}_{-i}) d\boldsymbol{\theta}_{-i} - m_i(z_i) \tag{3.2}$$

It is well known that (Myerson, 1979; Laffort and Green, 1978; *among others*), a direct mechanism is incentive compatible, if and only if,

$$m'(\theta_i) = \left[\frac{\partial}{\partial z_i} \int_{\theta_{-i}} v_i(x_i^*(z_i, \theta_{-i}), \theta_i) f_{-i}(\boldsymbol{\theta}_{-i}) d\boldsymbol{\theta}_{-i} \right]_{z_i = \theta_i} \geq 0 \tag{3.3}$$

Remark 3.1: *$x_i^*(z_i, \theta_{-i})$ needs not to be differentiable, but we assume that $\int_{\theta_{-i}} v_i(x_i^*(z_i, \theta_{-i}), \theta_i) f_{-i}(\theta_{-i})d$ is differentiable in z_i. Of course, assuming $v_i(x_i^*(z_i, \theta_{-i}), \theta_i)$ is continuously differentiable function in (x, θ) is conventional.*

For the ex post budget balance mechnisms, the following lemma states the link between lemma 1 and the lowest type's payoff in the class of budget balance mechanism.

Lemma 3.3: *For any incentive compatible and ex post budget balance mechanism, the following inequality must be true:*

$$\sum U_i(\underline{\theta}_i) \leq \mathbb{E}\Delta$$

where $U_i(\underline{\theta}_i)$ is the lowest type's expected payoff under that mechanism. **(Proof see Appendix A2)**

The above lemma builds an interesting linkage between budget balance mechanism and individually rational mechanism. Particularly, if $\underline{u}_i(\underline{\theta}_i) = 0$, it had better hold $\sum U_i(\underline{\theta}_i) = \mathbb{E}\Delta$, since individual rationality condition $\sum U_i(\underline{\theta}_i) \geq 0$ holds if and only $\mathbb{E}\Delta \geq 0$. For the typical approach to construct a budget balance mechanism, the crutial issue is to construct $m_i(\theta_i)$. Note that under any budget balance mechanism, there must be somebody who pays and others who are paid. So the allocation of payee/payer-ship is a key instrument, like conterveiling incentive. Our constructure of the mechanism is based on this intrinsic property of budget balance mechanism.

3.2.4 Related literature

1. Mechanism design and public good provision (Vickery, 1961; Groves, 1973; Clark, 1971), Green and Laffont (1977, 1979), Holmstrom (1977), Holmstrom and Myerson (1983). There is a lot of literature in this field.

2. Auction of shares (Wilson, 1979 QJE; Ausubel, 2004, AER among other): auction degign for divisible good quantity. Conclusion: no efficient allocation in general.

3. Partnership dissolving (Crampton, Gibbons and Klemperer, 1987, EMA, MacAfee, 1992, JET, Modouano *with others*, 2002;). There is efficient resolution in symmetric independent environment if the initial endownment is very symmetric. But not in general.

4. Ex post implementation and ex post mechanim design. Bergermann and Morries (2003), Chung and Ely (2003). In general, it is impossible to have an efficient mechanism satisfying ex

post IC, IR and BB.

3.3 Private Good Senario

3.3.1 Mechanism under SIPV environment

In the case without externality, neither in utility nor in production, for individual i, his utility from consumption is $v_i(x_i, \theta_i)$ over x_i, and no one else benefits from x_i at all (this mean the only gets utility from his own consumption). We assume that θ_i is independent crossing individuals, and we will discuss both identical distribution or asymmetric distribution later. The cost function is $c(\max_i x_i)$. In this case,

$$S_i(\theta^{n:n}) = \max_{x_i} v_i(\theta^{n:n}, x_i) - c_i(x_i)$$

where we use $\theta^{i:j}$ to denote the i-th smallest order statisics among j random variables. It is reasonable to assume $S_i(\underline{\theta}_i) \geq 0$, implying that production is always socially efficient, otherwise, the socially efficient decision might be no production since outside option is higher[5].

In this case, VCG payment is the following:

$$M_i^V(\hat{\boldsymbol{\theta}}) = \begin{cases} S(\hat{\theta}_i) + c(x^*(\hat{\theta}_i)) & \text{if } \max_{j \neq i} \hat{\theta}_j < \hat{\theta}_i \\ \frac{1}{\#\mathcal{N}_j}[S(\theta^{n-1:n}) + c(x^*(\theta_i))] \text{ for any } & \hat{\theta}_i \in \mathcal{N}_j \\ 0 & \text{if } \max_{j \neq i} \hat{\theta}_j > \hat{\theta}_i \end{cases}$$

VCG runs expected surplus if and only if $\mathbb{E}S(\theta)^{n-1:n} \geq 0$, which is always the case since $S_i(\underline{\theta}_i) \geq 0$. The question here is: can we allocate this expected surplus properly so that a budget balance mechanism is also ex post IR?

We propose the following mechanism to answer the above question.

M1: *(i) The highest type agent makes production decision by himself (therefore it is opti-*

[5] If $c(x)$ is continuous, the first best solution x^{1st} solves $\frac{\partial}{\partial x} v(x, \theta^{n:n}) = c'(x)$, If x is not continous, like a binary variable, $S(\theta^{n:n})$ is still well-defined.

mal); (ii) based on the report $\hat{\boldsymbol{\theta}}$*, the payment rule associated with this allocation is:*

$$M_i^F(\hat{\boldsymbol{\theta}}) = \begin{cases} \frac{n-1}{n}\mathbb{E}[S(\tau^{n:n})/\tau^{n:n} \leq \hat{\theta}_i] & \text{if } \max_{j\neq i}\hat{\theta}_j < \hat{\theta}_i \\ \frac{1}{\#\mathcal{N}_j}\frac{n-1}{n}\mathbb{E}[S(\tau^{n:n})/\tau^{n:n} \leq \hat{\theta}_i] & \text{for any } \hat{\theta}_i \in \mathcal{N}_j \\ -\frac{1}{n}\mathbb{E}[S(\tau^{n:n})/\tau^{n:n} \leq \hat{\theta}_j^{n:n}] & \text{if } \max_{j\neq i}\hat{\theta}_j > \hat{\theta}_i \end{cases} \tag{3.4}$$

Under this decentralized mechanism, individual i's payoff is

$$u_i(\hat{\theta}_i, \theta_{-i}; \theta_i) = \mathbf{1}(\max_{j\neq i}\hat{\theta}_j < \hat{\theta}_i)S(\theta_i) - M_i^F(\hat{\boldsymbol{\theta}})$$

where $\mathbf{1}(Z)$ is indication function for event Z.

Remark 3.2: *This mechanism decentralizes the production decision, comparing with VCG, where the production decision is centralized. There is a centralized counterpart of M1, where the designer determines* $x^*(\theta)$ *and sells to the highest type reporter. The agents' behavior is equavalent (in term of strategy) under both versions. This conclusion holds even in the environment of externality.*

We claim this mechanism **M1** or $< \mathbf{x}^*(\boldsymbol{\theta}), \mathbf{M}^F(\boldsymbol{\theta}) >$ is an ex post socially effecient mechanism if and only if VCG runs expected social surplus, under SIPV environment.

Theorem 3.2: *Under SIPV, the proposed mechanism M1:* $< x^*(\theta), M^F(\theta) >$ *is ex post soically efficient if and only if VCG mechanim runs expected surplus.*

Proof. (i) To check the incentive compatibility, note that,

$$\begin{aligned} m(\theta) &= G(\theta)\left[\frac{n-1}{n}\frac{\int_{\underline{\theta}}^{\theta} S(\tau)dF(\tau))^n}{F(\theta)^n}\right] - \frac{1}{n-1}\int_{\theta}^{\bar{\theta}}\frac{n-1}{n}\left(\frac{\int_{\underline{\theta}}^{z} S(\tau)dF(\tau))^n}{F(z)^n}\right)dG(z) \\ &= (n-1)\int_{\underline{\theta}}^{\bar{\theta}} S(\tau)F(\tau))^{n-1}dF(\tau) - (n-1)\int_{\theta}^{\bar{\theta}} S(\tau)F(\tau))^{n-2}dF(\tau) \end{aligned}$$

therefore,

$$m'(\theta) = g(\theta)S(\theta)$$

which accords with the necessary and sufficient condition of interim IC.

(ii) Ex post budget balance is met by construction.

(iii) We check interim IR first, since if interim IR fails, the game can not proceed. It is easy to chek that

$$U(\theta) = G(\theta)S(\theta) - m(\theta)$$

therefore,

$$U(\underline{\theta}) = -m(\underline{\theta}) = (n-1)\int_{\underline{\theta}}^{\bar{\theta}} S(\tau)(1-F(\tau))F(\tau))^{n-2}dF(\tau) = \frac{1}{n}\mathbb{E}\Delta$$

So if and only if $\mathbb{E}\Delta \geq 0$, at the interim stage, nobody will quit. And at the ex post stage, if $\max_{j\neq i}\theta_j > \theta_i$, his ex post payoff $u^l(\theta) = \frac{1}{n}\mathbb{E}[S(\tau^{n:n})/\tau^{n:n} \leq \theta_j^{n:n}] \geq 0$; if $\max_{j\neq i}\theta_j < \theta_i$,

$$u^w(\theta) = S(\theta) - \frac{(n-1)}{n}\frac{\int_{\underline{\theta}}^{\theta} S(\tau)dF(\tau))^n}{F(\theta)^n} \geq u^l(\theta) > 0$$

This complete the proof of argument. Q.E.D. ■

It is also seen that M1 satisfies ex post monotonicity, defined as follows.

Definition 3.4: *A mechanism satisfies interim (ex post) payoff monotonicity if the interim (ex post) payoff is non-decreasing given any realization of other bidders' type.*

It is clear that IC interim requires interim payoff monotonicity, but not ex post monotonicity. And ex post monotonicity at least implies that in the ex post stage, the higher type can at least never be worse than the lower type. The existing mechanisms such as AGV do not satisfy ex post payoff monotonicity (as we will see an example later). Ex post payoff monotonicity is closely related to ex post incentive compatibility, because ex post payoff monotonicity is necessary for IC ex post but not sufficient. Meanwhile, ex post payoff monotonicity may imply that with some subset of realizations of state of the world, ex post IC is met.

Proposition 3.2: *Under SIPV, M1 satisfies ex post payoff monotonicity.*

Proof. Given any realization of other players' type, if individual i's type is lower than the winner's, say $\theta_i < \theta^{n:n}$, then his ex post payoff is independent of his type; if he himself is the highest type, $\theta_i = \theta^{n:n}$, then his ex post pay off is increasing of his type since $S(\theta) - \frac{(n-1)}{n}\frac{\int_{\underline{\theta}}^{\theta} S(\tau)dF(\tau))^n}{F(\theta)^n}$ is increasing with θ. The only trick here is to show that there is no drop when his type passes the pivotal point $\theta^{n-1:n}$. Note that $u^w(\theta) > u^l(\theta)$, so when passing the pivotal point, his payoff jumps rather than drops. And at the point $\theta_i = \theta^{n-1:n}$, the tie-setting makes his payoff in between $u^w(\theta)$ and $u^l(\theta)$, therefore, his expost payoff is non-decreasing given any realization of other bidders' type. Q.E.D. ■

Although M1 is not ex post incentive compatible, the ex post payoff monotonicity still generates several notable robust properties, in terms of ex post implementation. The agents will not regret the allocation under M1: the winner definitely does not want to be a loser by lowering his report; and the loser probably does not want to increase his report to become a winner under some subset of realizations as well. It seems acceptable that the winner can not renege on his payment, unless he wants to give up his winner-ship. Formally, we define winner's no veto power as below.

Definition 3.5: *Winner has no veto power in payment (WNVPP), if he cannot underpay without giving up his winner-ship.*

The above assumption rules out the winner's default such as simply paying less when he obtains the object. But we still allow the winner to change his report if he wants to switch the allocation. We have the following proposition.

Proposition 3.3: *Under SIPV, M1 with WNVPP is ex post incentive compatible with probability,*

$$\Pr(EPIC) = n\int_{\underline{S}}^{\bar{S}}\left[F(S^{-1}(\frac{\int_{\underline{\theta}}^{\theta} S(\tau)dF(\tau))^n}{F(\theta)^n}))\right]^{n-1} dF(S^{-1}(\tau)).$$

especially, when $F(S^{-1}(\tau)) = S^{\gamma}$, a power function, then

$$\Pr(\text{EPIC}) = (\frac{n\gamma}{n\gamma+1})^{(n-1)\gamma} \to \frac{1}{e} \cong 0.37$$

Proof. Given his oppenent's truth-telling ex post, the winner will not report $\tilde{\theta} > \theta$ ex post. And with WNVPP, the winner can not report $\tilde{\theta} < \theta$, but still holds the object. Moreover, the winner does not want to report $\tilde{\theta} \leq \theta_{n-1:n}$ too, since he earns less by changing allocation. Given the winner's behavior, the losers do not benefit by reporting $\tilde{\theta} < \theta$. The loser may only mispresent his type $\tilde{\theta} > \theta$ when the second highest type agent's realization is close enough to the winner's,

$$S_{n-1:n} - \frac{n-1}{n}\frac{\int_{\underline{\theta}}^{\theta_{n:n}} S(\tau)dF(\tau))^n}{F(\theta_{n:n})^n} \geq \frac{1}{n-1}\frac{\int_{\underline{\theta}}^{\theta_{n:n}} S(\tau)dF(\tau))^n}{F(\theta_{n:n})^n}$$

This event happens with probability

$$\Pr(S_{n-1:n} \geq \frac{\int_{\underline{\theta}}^{\theta_{n:n}} S(\tau)dF(\tau))^n}{F(\theta_{n:n})^n}) = 1 - n\int_{\underline{S}}^{\bar{S}} \left[F(S^{-1}(\frac{\int_{\underline{\theta}}^{\theta} S(\tau)dF(\tau))^n}{F(\theta)^n}))\right]^{n-1} dF(S^{-1}(\tau))$$

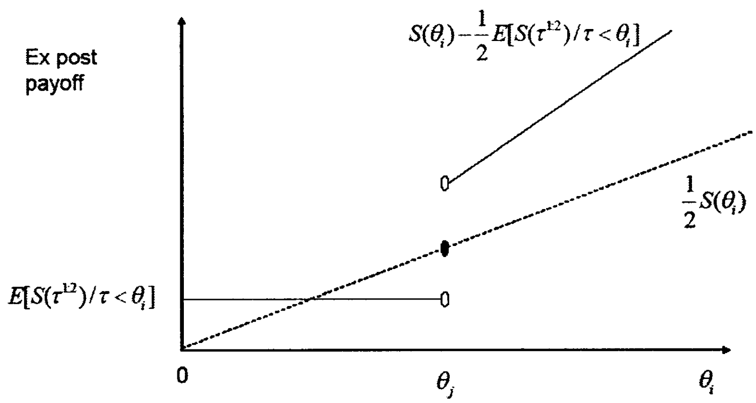
Thus, we have $\Pr(\text{EPIC}) = 1 - \Pr(S_{n-1:n} \geq \frac{\int_{\underline{\theta}}^{\theta_{n:n}} S(\tau)dF(\tau))^n}{F(\theta_{n:n})^n})$. It is easy to check when $F(S^{-1}(\tau)) = S^{\gamma}$, $\frac{\int_{\underline{\theta}}^{\theta} S(\tau)dF(\tau))^n}{F(\theta)^n} = \frac{n\gamma}{n\gamma+1}$, and $\Pr(\text{EPIC}) = (\frac{n\gamma}{n\gamma+1})^{(n-1)\gamma}$ follows. Q.E.D. ■

The above proposition indicates that if an interim incentive compatible mechanism wants to maximize the probability of ex post incentive compatibility, then ex post payoff monotonicity is required[6]. For example, the probability for AGV to be EPIC is zero even with WNVPP. We are wondering if other mechanisms also have nice properties as M1. The answer is no. M1 is the generically unique mechanism that satisfies ex post monotonicity and ex post individual rationality among all budget balance Bayesian mechanisms.

Theorem 3.3: *Under SIPV, M1 is the (generically) unique symmetric mechanism that is ex post socially efficient and satisfies ex post monotonicity (maximize probability of ex post IC). (Proof see A3)*

We compare M1 and AGV below. Take $n = 2$ as an example, the ex post pay-off structure of **M1** is indicated by Figure 1, which is a non-decreasing function. In a concrete example below, probability of winner being worse than loser in AGV is slightly greater than 60%, and probability not ex post IR is about 28%. The probability for **M1** to be EPIC is $\frac{\sqrt{3}}{3}$.

[6] The implication of the above proposition can also be understood as follows. If the designer only accepts the appeal that is pivotal (dismisses any ex post change of report if it is not pivotal), then with some probability, M1 is ex post incentive compatible.



(Insert Figure 1 here)

Example 3.1: Comparison between AGV and M1.

Assume $n = 2$, θ *uniformly distributed over* $[0, 1]$, $c(x) = \frac{1}{\rho}\theta^{\rho}$, $v = x\theta$. *It is easy to see the first best social welfare is* $S(\theta) = \frac{1}{\gamma}\theta^{\gamma}$ *where* $\gamma \equiv \frac{\rho}{\rho-1}$.

Payoff Under AGV.

If i wins, his expost payoff is:

$$S(\theta_i) - \frac{\int_{\theta_j}^{\bar{\theta}} S(\theta)dF(\theta)}{1 - F(\theta_j)} = \frac{1}{\gamma(\gamma+1)}[\frac{\theta_i^{\gamma}(\gamma+1)(1-\theta_j) - (1-\theta_j^{\gamma+1})}{(1-\theta_j)}$$

Because when θ_j *is close to* θ_i *enough, since* $\gamma > 1$, *we have*

$$\theta_i^{\gamma}(\gamma+1)(1-\theta_i) - (1-\theta_i^{\gamma+1}) = \gamma\theta_i^{\gamma}(1-\theta_i) - (1-\theta_i^{\gamma}) < 0$$

So if the two agent's types are close enough, the winner may suffer even though he wins the object. For example, if $\gamma = 2$, *the cost is quadratic, then,*

$$\theta_j^2 < \theta_i^2 < \frac{1+\theta_j+\theta_j^2}{3}$$

is possible. The possibility that the winner's ex post IR fails is about a third.

$$\text{Pr(winner's EPIR fails in } AGV) = \int_0^1 [\frac{1+\theta+\theta^2}{3} - \theta^2]d\theta = \frac{5}{18}$$

And the probability of winner being worse than loser is:

$$\begin{aligned}&\Pr(\text{winner being worse than loser in } AGV)\\ =& \int_0^{\frac{1}{2}\sqrt{3}-\frac{1}{2}}[\frac{2(1+\theta+\theta^2)}{3}-\theta^2]d\theta+\int_{\frac{1}{2}\sqrt{3}-\frac{1}{2}}^{1}(1-\theta^2)d\theta = 0.60\end{aligned}$$

If we look at ex post IC, we find for any realization of θ, there is at least one agent who wants to change the allocation and report.

Payoff Under M1.

If i winns, his ex post payoff is

$$u^w(\theta)=\frac{1}{2}\theta^2-\frac{1}{8}\theta^2=\frac{3}{8}\theta^2$$

if i loses, his ex post pay off is $u^w(\theta)=\frac{1}{8}\theta^2$. In both case, the participant's ex post payoff is positive. Meanwhile under $M1$, nobody would like to make change if an ex post realization falls into region $\theta_j \geq \sqrt{3}\theta_i$, which happens with probability $\frac{\sqrt{3}}{3}=58\%$.

3.3.2 Asymmetric Distribution Case

We now consider environments where the distribution is not identical, but still independent. As we will see, the payment rule can be characterized by a set of non-homogeneous linear ordinary differential equations. Moreover, we can obtain some important observations without requirement to solve the system analytically. As a special case, when $n=2$, we can completely solve the payment function without putting restriction on distributions.

With a little loss of generality, we assume that all θ_i's are in the same support $[\underline{\theta},\overline{\theta}]$, being drawn according to c.d.f. $F_i(\theta_i)$ association with continuous differentiable p.d.f. $f_i(\theta)$.

Let $G_j(z_j)=\Pi_{k\neq j}F_k(z_j)$ be the c.d.f. of the random variable $z_j=\max_{k\neq j}z_k$, association with p.d.f. $g_j(z_j)=G_j'(z_j)$, and let $\Pi(z)=\Pi_{k=1}^n F_k(z)$. The distribution that j wins given i is not the winner is,

$$\begin{aligned}
F_{n-1:n-1}(z_j + \Delta z \backslash z_j &> \theta_i) - F_{1:n-1}(z_j \backslash z_j > \theta_i) \\
&= \Pr(Z_j \in [z_j, z_j + \Delta z], z_j > \max_{k \neq i,j} z_k \backslash z_j > \theta_i) \\
&= \frac{\Pr(Z_j \in [z_j, z_j + \Delta z], z_j > \max_{k \neq i,j} z_k, z_j > \theta_i)}{\Pr(\max_{k \neq i} z_k > \theta_i)} \\
&= \frac{(F_j(z_j + \Delta z) - F_j(z_j)) \left[\Pi_{k \neq j,i} F_k(z_j)\right]}{1 - G_i(\theta_i)}
\end{aligned}$$

Therefore,

$$f_{n-1:n-1}(z_j \backslash z_j > \theta_i) = \frac{f_j(z_j) \left[\Pi_{k \neq j,i} F_k(z_j)\right]}{1 - G_i(\theta_i)}$$

Let $M_i(\theta_i)$ be the winner i's payment when he reports θ_i, then the expected payment for i is:

$$m_i(\theta_i) = G_i(\theta_i) M_i(\theta_i) - \frac{1}{n-1} \sum_{j \neq i} \int_{\theta_i}^{\bar{\theta}} M_j(z) \left[\Pi_{k \neq j,i} F_k(z_j)\right] dF_j(z) \tag{3.5}$$

Note that $m_i(\theta_i)$ is incentive compatible if and only if

$$m_i'(\theta) = g_i(\theta) S_i(\theta)$$

therefore, we can characterize the payment rule $M_i(\theta_i)$ through the following ODEs. Take derivative w.r.t.θ_i on both sides of equation (5), obtain an ODE for any i,

$$S_i(\theta_i) \sum_{k \neq i} \frac{G_k f_k}{F_i} = M_i \sum_{k \neq i} \frac{G_k f_k}{F_i} + G_i M_i' + \frac{1}{n-1} \sum_{k \neq i} \frac{G_k f_k}{F_i} M_k$$

This is

$$S_i(\theta_i) \sum_{k \neq i} \frac{f_k}{F_k} = M_i \sum_{k \neq i} \frac{f_k}{F_k} + M_i' + \frac{1}{n-1} \sum_{k \neq i} \frac{f_k}{F_k} M_k$$

where we compress the augument θ for convenience.

Write $\frac{f_k}{F_k} = q_k$, $q_{-i} = \sum_{k \neq i} q_k$, denote $\mathbf{M}(\theta) = \begin{pmatrix} M_1(\theta) \\ M_2(\theta) \\ \cdots \\ M_n(\theta) \end{pmatrix}$,

$$\mathbf{A}(\theta) = -\begin{pmatrix} q_{-1}(\theta) & \frac{1}{n-1}q_2(\theta) & \cdots & \frac{1}{n-1}q_n(\theta) \\ \frac{1}{n-1}q_1(\theta) & q_{-2}(\theta) & \frac{1}{n-1}q_3(\theta) & \frac{1}{n-1}q_n(\theta) \\ \cdots & \cdots & \cdots & \cdots \\ \frac{1}{n-1}q_1(\theta) & \frac{1}{n-1}q_2(\theta) & \frac{1}{n-1}q_3(\theta) & q_{-n}(\theta) \end{pmatrix}$$

$\mathbf{B}(\theta) = \begin{pmatrix} q_{-1}(\theta) \\ q_{-2}(\theta) \\ \cdots \\ q_{-n}(\theta) \end{pmatrix} S(\theta)$. Finally, the ODE system is,

$$\mathbf{M}'(\theta) = \mathbf{A}(\theta)\mathbf{M}(\theta) + \mathbf{B}(\theta) \quad \text{with} \quad \mathbf{M}(\underline{\theta}) = \frac{1}{n}\mathbf{S}(\underline{\theta}) \tag{3.6}$$

The solution of the above non-homogenous linear ODE system with initial condition $M_i(\underline{\theta}) = \frac{1}{n}S(\underline{\theta})$ exists and is unique.

It seems hard to solve the above system in general if n is large. But for us, the information from the above characterization is enough for us to have the following important result.

Proposition 3.4: *Under asymmetric but independent environment, (i) the payment rule $M(\theta)$ characterized by (6) is ex post socially efficient association with allocation x^*; (ii) the lowest type's expected payoff coincides with expected surplus of VCG $\sum U_i(\underline{\theta}) = E\Delta$.* (**Proof see Appendix A4**).

The above proposition generalizes the result of SIPV. And the second property shows that under this payment rule the lowest type agents' social benefit is exactly the expected social surlus in VCG. (AGV does not have property (ii)).

Remark 3.2: *This observation can be generalized to the situation that utility function is also heterogenous. In that situation, we can re-parameterize the type and the similar reasoning still applies (with a little loss of generality in the support restriction).*

For intuition, we solve the explicit solution in a two player case, for any distribution. We have the following proposition.

Proposition 3.5: *Under environment of asymmetric independent private value (AIPV), the payment rule of M1 is,*

$$M_i(\theta_i) = \frac{1}{2}S(\underline{\theta}) + \int_{\underline{\theta}}^{\theta_i} \frac{f_j(\theta)\int_{\underline{\theta}}^{\theta} F_i(z)F_j(z)S'(z)dz}{F_i(\theta)F_j(\theta)^2} d\theta \tag{3.7}$$

and satisfies ex post monotonicity. **(Proof see Appendix A5)**

Therefore, we find even under asymmetric distribution, the ex ante side-payment is not required to meet both interim IR and ex post IR. For intuition, we provide the following example.

Example 3.2: Asymmetric payment function when type distribution is a power function.

If the distribution is $F_i = \theta^{a_i}$, it can be calculated that,

$$U_i(\underline{\theta}) = \frac{1}{\gamma}\left(\frac{a_j}{\gamma + a_j} - \frac{a_j}{a_j + a_i + \gamma}\right) > 0$$

So definitely, the above regime is ex post socially efficient although AGV is not interim soically efficient. We find that $M_i(\theta) \geq M_j(\theta)$ iff $a_j \geq a_i$.

$$M_i(\theta_i) = \int_{\underline{\theta}}^{\theta_i} \frac{a_j\theta^{a_j-1}\int_{\underline{\theta}}^{\theta} z^{a_i+a_j}z^{\gamma-1}dz}{\theta^{a_i+2a_j}} d\theta = \frac{a_j}{a_i + a_j + \gamma}\theta_i^{\gamma}$$

The stronger type agent's payment scheme will be flatter than the weaker one's, as we find in standard first price auction (Maskin and Riley, 2001).

As an important result, we now improve lemma 1 to a stronger version, by incorporating ex post IR.

Theorem 3.4: *If type is independent, under private value case, (i) there exist ex post socially efficient mechanisms if and only if VCG runs expected surplus, $E\Delta \geq 0$; (ii) in addition, M1 is the unique payment rule generates ex post monotonicity.*

Proof. (i) The necessary part is commonly known in the existing literature, same as the proof of lemma 1. We only show the sufficiency. First consider the payment structure $\mathbf{M}^F$, defined in **M1**. By revenue equivalence theorem, any two incentive compatible mechanims differ in their payoff up to a constant. So there exist constants c_i^F such that,

$$U_i^F(\theta_i) = \mathbb{E}_{\theta_{-i}} S(\theta_i, \boldsymbol{\theta}_{-i}) - c_i^F$$

It is also the case that constants c_i^V exist, such that,

$$U_i^V(\theta_i) = \mathbb{E}_{\theta_{-i}} S(\theta_i, \boldsymbol{\theta}_{-i}) - c_i^V$$

If VCG runs an expected surplus, $\mathbb{E}\Delta \geq 0$ means,

$$\mathbb{E} \sum c_i^V \geq \mathbb{E} \sum c_i^F$$

for all $i \geq 1$, define $d_i = c_i^F - c_i^V$, and let $d_1 = -\sum_{i=2}^n d_i$. Then we can construct a mechanism $\mathbf{M}^{\#}$ by

$$M_i^{\#}(\theta) = M_i^F(\theta) + d_i$$

and this means $\mathbf{M}^{\#}$ is also incentive compatible. We only need to check $\mathbf{M}^{\#}$ is ex post individual rational. Importantly, note that $M_i^F(\theta)$ is ex post individual rational, therefore, the difference of ex post payoff between $M_i^{\#}(\theta)$ and $M_i^F(\theta)$ is also up to a constant[7]. So it is sufficient to check the ex post using similar construction.

For $i \neq 1$,

$$U_i^{\#}(\theta_i) = U_i^F(\theta_i) + d_i = U_i^F(\theta_i) + c_i^F - c_i^V = U_i^V(\theta_i) \geq 0$$

and

$$\begin{aligned} U_1^{\#}(\theta_1) &= U_1^F(\theta_1) + d_1 \\ &\geq U_1^F(\theta_1) + d_1 \\ &= U_1^V(\theta_1) \\ &\geq 0 \end{aligned}$$

[7]This is the key part of the proof. In the proof of lemma 1 (Krishna and Perry, 1998; Krishna, 2002), their proof is based on AGV, because AGV *pe se* may not be IR ex post, $M_i^A(\theta) + d_i$ may not be IR ex post as well, although $M_i^A(\theta) + d_i$ may be IR interim.

Since both VCG and $< \mathbf{x}^*(\theta), M^F(\theta) >$ are ex post individual rational, then $\mathbf{M}^{\#}$ is also ex post individually rational.

(ii) The proof is similar to theorem 3. The key observation is that the payment must depend on the winner's type only. Q.E.D. ■

In fact, **M1** does not require ex ante side-payment, even under the asymmetric environment, which is an advantage in terms of pragmatic implimentation.

3.4 General Preference with Presence of Externality

Now we consider environment with presence of externality, either due to utility interaction like public good, or cost complementarity/substitution like spill-over. For tractability, here we discover a two-agent case, and will discuss an n-agent generalization later. It will be shown that the mechanism works in a 2-agent case might not work in the n-agent case. Interestingly, there is dramatically different implication between endogenous endowment sitution and its exogeneous countpart[8]. So we deal with them separately.

3.4.1 Endogenuous Quantity

Basics

For notational convenience, let

$$f^{(n-1)}(\theta) = (n-1)! f(\theta_{n-1:n-1})...f(\theta_{1:n-1})$$

be the joint pdf of n-1 order statatistics. And we deonte the expected consumption to i, when i's type is ranked as j-th order statistic as:

$$\bar{v}_i^{(j)}(x_i^*(.), \theta) = \underbrace{\int_{\theta}^{\bar{\theta}} ... \int_{\theta}^{\theta_{j-2}}}_{j-1} \underbrace{\int_{\underline{\theta}}^{\theta} \{ \int_{\underline{\theta}}^{\theta_{j+1}} \int_{\underline{\theta}}^{\theta_{j+2}} ... \int_{\underline{\theta}}^{\theta_{n-1}}}_{n-j} v_i(x_i^*(.., \theta_{j-1}, \theta, \theta_{j+1}, ..), \theta) f^{(n-1)}(\theta_{-i}) d\theta_{-i}.$$

[8] When endowment is endogenous, $S(\boldsymbol{\theta})$ will be supermodular, while endowment is exogeneous, $S(\boldsymbol{\theta})$ will be submodular, see proposition 1.

Similarly, his expected social surplus is,

$$\bar{S}^{(j)}(\theta) = \underbrace{\int_{\theta}^{\bar{\theta}} \dots \int_{\theta}^{\theta_{j-2}}}_{j-1} \underbrace{\int_{\underline{\theta}}^{\theta} \int_{\underline{\theta}}^{\theta_{j+1}} \int_{\underline{\theta}}^{\theta_{j+2}} \dots \int_{\underline{\theta}}^{\theta_{n-1}}}_{n-j} S(.., \theta_{j-1}, \theta, \theta_{j+1}, ..) f^{(n-1)}(\theta_{-i}) d\theta_{-i}$$

We propose the mechanism as follows:

M2: *(i) The designer chooses optimal alloction rule* $\mathbf{x}^*(\hat{\boldsymbol{\theta}})$ *according to report* $\hat{\boldsymbol{\theta}}$*;*

(ii) each player receives consumption $x_i^*(\hat{\theta}_i, \hat{\theta}_{-i})$*, and the payment rule is the following:*

$$M_i^k(\hat{\theta}_i, \hat{\theta}_{-i}) = \begin{cases} -S_{-i}(\hat{\theta}_i, \hat{\theta}_{-i}) + [(1-k)\beta^k(\hat{\theta}_i) + k\beta^k(\hat{\theta}^{n-2:n})] & \text{if } \hat{\theta}_i > \max_{j \neq i} \hat{\theta}_j \\ \frac{1}{\# \mathcal{N}_j} \sum [S(\hat{\theta}_i, \hat{\theta}_{-i}) - \beta^k(\hat{\theta}_i)] & \text{if } \hat{\theta}_i = \hat{\theta}_j \text{ for } i, j \in \mathcal{N}_j \\ v_i(x_i(\hat{\theta}_i, \hat{\theta}_{-i}), \hat{\theta}_i) - \frac{[(1-k)\beta^k(\hat{\theta}^{n:n}) + k\beta^k(\hat{\theta}^{n-1:n})]}{n-1} & \text{if } \hat{\theta}_i < \max_{j \neq i} \hat{\theta}_j \end{cases}$$

where

$$\beta^k(\theta) = \frac{\int_{F^{-1}(k)}^{\theta} \left(\frac{d}{d\tau} \bar{S}^{(1)}(\tau) - \sum_{j=1}^{n} \frac{d}{d\tau} \bar{v}_i^{(j)}(x_i^*(., \tau, .), \tau) + m'(\tau) \right) \frac{(F(\tau)-k)^{n-1}}{F(\tau)^{n-2}} d\tau}{(F(\theta) - k)^n} \tag{3.8}$$

with $m'(\tau) = \left[\frac{\partial}{\partial z} \int_{\theta_{-i}} \frac{\partial}{\partial x} v(x(z, \theta_{-i}), z) d\mathbf{F}_{-i}(\boldsymbol{\theta}_{-i}) \right]_{z=\tau}$ and with $k \in [0,1]$ as a constant.

Under this mechanism, the individual with type θ_i pretends to report $\tilde{\theta}_i$ will have the following ex post payoff:

$$u_i(\theta_i, \tilde{\theta}_i) = \begin{cases} v_i(x_i(\tilde{\theta}_i, \theta_{-i}), \theta_i) + S_{-i}(\tilde{\theta}_i, \theta_{-i}) - [(1-k)\beta^k(\theta_i) + k\beta^k(\theta^{n-1:n})] & \text{if } \theta_i > \max_{j \neq i} \theta_j \\ \frac{1}{\# \mathcal{N}_j} \sum [S(\theta_i, \theta_{-i}) - \beta_i] & \text{if } \theta_i = \theta_j \text{ for } i, j \in \mathcal{N}_j \\ v_i(x_i(\tilde{\theta}_i, \theta_{-i}), \theta_i) - v_i(x_i(\tilde{\theta}_i, \theta_{-i}), \tilde{\theta}_i) + \frac{[(1-k)\beta^k(\theta^{n:n}) + k\beta^k(\theta^{n-1:n})]}{n-1} & \text{if } \theta_i < \max_{j \neq i} \theta_j \end{cases}$$

The above mechanism can be understood as follows. The highest type agent gets the entitlement to charge all social surplus at the cost of paying lump sum payment to the remaining losers; while the remaining losers need to pay the consumptions (thus earn zero consumption surplus) according to their reports but are paid by lump sum transfer from the winner. We claim that (i) the mechanism **M2** is interim socially efficient for any $k \in [0,1]$ if and only if $\mathbb{E}\Delta \geq 0$, and (ii) there exist some k such that **M2** is ex post socially efficient.

The following theorem states the results regarding (i).

Theorem 3.5: *If the distribition is i.i.d., and utility is symmetric, mechanism M2:* $< x^*(\theta), M^k(\theta) >$ *is: (i) budget balance and incentive compatible, and (ii) interim socially efficient if and only if* $\mathbb{E}\Delta \geq 0$ *for any* $k \in [0,1]$. **(Proof see Appendix A5)**

Under **M2**, the lowest type agent's payoff is exactly a$\frac{1}{n}$ share of total VCG expected social surplus (deficit). Obviously, **M2** is one kind of budget balance mechanism other than AGV. Compared with AGV, **M2** is interim budget balance if and only if VCG runs expected social surplus[9].

Remark 3.4: $\beta^k(\theta)$ *may be negative or even non-monotonic. If* $\beta^k(\theta)$ *is negative, this means the remaining people need to subsidize the winner in equilibrium.*

Two-player case

If $n = 2$, we claim that $k = 1$ is expost socially efficient. It is easy to know that from formula (8),

$$\beta^1(\theta) = \frac{\int_\theta^{\bar{\theta}} (1 - F(\tau))(S(\tau, \tau) - h(\tau))dF(\tau)}{[1 - F(\theta)]^2}$$

where

$$h(\tau) \equiv \frac{\int_\tau^{\bar{\theta}} \frac{\partial}{\partial \tau} v(x^*(\tau, z), \tau)dF(z)}{f(\tau)}$$

The following theorem states the result regarding **M2** when $n = 2$.

Theorem 3.6: *Given that endowment is endogenous, if distribution is i.i.d. and preference is symmetric, when* $n = 2$*, M2 with* $k = 1$ *has the following properties: (i) lump sum payment function* $\beta^1(\theta)$ *is monotonic (without restriction on distribution); (ii) M2 is ex post socially efficient if and only if* $\mathbb{E}\Delta \geq 0$*. and (iii) M2 also satisfies ex post monotonicity*

Proof. (i) If the endownment is endogenous, then $\frac{\partial}{\partial \tau} v(x^*(\tau, z), \tau)$ is an increasing function

[9] To see this, note in symmetric situation, under AGV, the lowest type's payoff is,

$$\begin{aligned} U^A(\underline{\theta}_i) &= \mathbb{E}_{\theta_{-i}} S(\underline{\theta}_i, \theta_{-i}) - \frac{1}{n-1} \mathbb{E}_{\theta_{-i}} \sum_{j \neq i} \mathbb{E}_{\theta_{-j}} [S_{-j}(\theta_j, \theta_{-j})] \\ &= \mathbb{E}_{\theta_{-i}} S(\underline{\theta}_i, \theta_{-i}) - \frac{(n-1)}{n} [\mathbb{E}_{\boldsymbol{\theta}} S(\theta_i, \theta_{-i}) + \mathbb{E}_{\theta_{-i}} \mathbb{E}_{\theta_{-j}} c(\mathbf{x}(\theta_j, \theta_{-j}))] \\ &\leq \mathbb{E}_{\theta_{-i}} S(\underline{\theta}_i, \theta_{-i}) - \frac{(n-1)}{n} \mathbb{E}_{\boldsymbol{\theta}} S(\theta_i, \theta_{-i}) \end{aligned}$$

of z, since $x^*(\tau, z)$ is an increasing function of z. Therefore,

$$\begin{aligned}
\beta^1(\theta) &= \frac{\int_\theta^{\bar\theta}(1-F(\tau))(S(\tau,\tau)-h(\tau))dF(\tau)}{[1-F(\theta)]^2} \\
&\leq \frac{\int_\theta^{\bar\theta}(1-F(\tau))(S(\tau,\tau)-\frac{\partial}{\partial\tau}v(x^*(\tau,\tau),\tau)\frac{\int_\tau^{\bar\theta}dF(z)}{f(\tau)})dF(\tau)}{[1-F(\theta)]^2} \\
&= \frac{\int_\theta^{\bar\theta}(1-F(\tau))S(\tau,\tau)dF(\tau)}{[1-F(\theta)]^2} - \frac{\frac{1}{2}\int_\theta^{\bar\theta}\frac{dS(\tau,\tau)}{d\tau}(1-F(\tau))^2d\tau}{[1-F(\theta)]^2} \\
&= \frac{1}{2}S(\theta,\theta)
\end{aligned}$$

The last step comes from integration by parts. Note that when $\theta \longrightarrow \bar\theta$,

$$\beta^1(\bar\theta) = \frac{1}{2}S(\bar\theta,\bar\theta)$$

This means $\beta^1(\theta)$ is uniformly bounded by $\frac{1}{2}S(\theta,\theta)$, but achieves $\frac{1}{2}S(\theta,\theta)$ at the boundary, which implies that

$$\beta'(\theta) \geq \frac{1}{2}\frac{d}{d\theta}S(\theta,\theta) > 0$$

Suppose this is not true, then for some $\theta < \bar\theta$, there always exists ϵ small enough such that for $\theta \in [\bar\theta - \epsilon, \bar\theta]$, $2\beta(\theta) > S(\theta,\theta)$, a contradiction.

(ii) We only need to check ex post IR constraint. To meet the loser's ex post IR, the bid should be non-negative. To see this,

$$\begin{aligned}
&\beta^1(\underline{\theta}) \\
=& \int_{\underline{\theta}}^{\bar\theta}[S(\theta,\theta)-h(\theta)](1-F(\theta))dF(\theta) \\
=& \int_{\underline{\theta}}^{\bar\theta}\int_{\underline{\theta}}^{\theta}S(\theta,\tau)dF(\tau)dF(\theta) - \int_{\underline{\theta}}^{\bar\theta}\frac{1-F(\theta)}{f(\theta)}\int_{\underline{\theta}}^{\theta}\frac{\partial}{\partial\theta}v(x^*(\theta,\tau),\theta)dF(\tau)dF(\theta) \\
&+ \int_{\underline{\theta}}^{\bar\theta}\frac{1-F(\theta)}{f(\theta)}\int_{\theta}^{\bar\theta}\frac{\partial}{\partial\theta}v(x^*(\theta,\tau),\theta)dF(\tau)dF(\theta) \\
=& \int_{\underline{\theta}}^{\bar\theta}\int_{\underline{\theta}}^{\theta}S(\theta,\tau)dF(\tau)dF(\theta) - \int_{\underline{\theta}}^{\bar\theta}\int_{\underline{\theta}}^{\bar\theta}\frac{1}{\lambda(\theta)}\frac{\partial}{\partial\theta}v(x^*(\theta,\tau),\theta)dF(\tau)dF(\theta) \\
=& \frac{1}{2}\mathbb{E}\Delta = U(\underline{\theta})
\end{aligned}$$

As long as $\mathbb{E}\Delta \geq 0$, then the bid will be non-negative overall, and the lowest type agent's payoff will be non-negative overall (not only interim but also ex post). For the winner, it is easy to see for any $\theta_i > \theta_j$,

$$S(\theta_i, \theta_j) - \beta(\theta_j) > 0$$

since $\beta(\theta_j) \leq \frac{1}{2}S(\theta_j, \theta_j)$.

(iii) It is ready to see, from the monotonicity of the bid, that the loser's ex post pay-off is monotonic over his own type, and the winner's pay-off is monotonic over his own type too since payment is independent of his own type. The only point we need to check is the pivotal point. As $\theta_j \longrightarrow \theta_i$, the winner still be better off since $\beta(\theta_j) \leq \frac{1}{2}S(\theta_j, \theta_j)$. The equality only holds at $\theta = \overline{\theta}$. Q.E.D. ■

The above theorem says that for a two-person case, mechanism **M2** solves the allocation problem well. And we verify that the necessary and sufficient condition for existence of an ex post socially efficient mechanism is the same as that of interim socially efficient one. It is worth pointing out that $\beta^0(\theta)$ might not be always ex post individually rational, due to the fact

$$S(\theta, \underline{\theta}) - \beta(\theta) \geq 0$$

might fail when $\beta(\underline{\theta}) = \mathbb{E}\Delta > 0$ but $S(\underline{\theta}, \underline{\theta}) = 0$.

Remark 3.5: *If the outside reservation utility $\underline{u}_i(\theta_i)$ is symmetric but type-dependent, then we need to modify the social welfare $S(\theta)$ as net social welfare $S(\theta) - \sum \underline{u}_i(\theta_i)$, and each agent's gain from the project will be net gain $v_i(x_i^*, \theta_i) - \underline{u}_i(\theta_i)$. The conclusions from the above theorem are still true as we will see in the next section.*

Do other values of $k \in [0, 1)$ have such kind of properties? For example, $k = 0$, The following corollary states the result.

Corollary 3.2: *For M2 when $n = 2$, under the same condition as theorem 4, there exists a cut-off $k^* \in [0, 1)$ such that for $k \geq k^*$, M2 with $k < 1$ possesses the same properties as $k = 1$ if $\mathbb{E}\Delta > 0$.*

Proof. The proof crucially depends on the property of $k = 1$. Note that $\beta^{k\prime}(\theta)$ is a continous function of k and $\beta^{k\prime}(\theta) > 0$ for $k = 1$, therfore, for k close enough to 1, $\beta^{k\prime}(\theta) \geq 0$ will still hold. At the same time, observing that

$$h'(\theta) = \frac{-\frac{\partial v}{\partial \theta}v(x^*(\theta,\theta),\theta)f(\theta) + \int_\theta^{\bar{\theta}} \frac{d}{d\theta}\left(\frac{\partial v}{\partial \theta}v(x^*(\theta,\tau),\theta)\right) dF(\tau)}{f(\theta)}$$
$$-\frac{f'(\theta)}{f(\theta)}\frac{\int_\theta^{\bar{\theta}} \frac{\partial v}{\partial \theta}v(x^*(\theta,\tau),\theta)dF(\tau)}{f(\theta)}$$

will be negative as θ close enough to $\overline{\theta}$, which means for k close enough to 1, for $\theta \geq F^{-1}(k)$, we have

$$\begin{aligned}
\beta^{k\prime}(\theta) &= \frac{f(\theta)}{F(\theta)-k}[S(\theta,\theta) - h(\theta) - 2\frac{\int_{F^{-1}(k)}^{\theta}(F(\tau)-k)(S(\tau,\tau)-h(\tau))dF(\tau)}{[F(\theta)-k]^2}] \\
&= \frac{f(\theta)}{F(\theta)-k}\frac{\int_{F^{-1}(k)}^{\theta}(F(\tau)-k)^2(\frac{d}{d\tau}S(\tau,\tau)-h'(\tau))d\tau}{[F(\theta)-k]^2} \geq 0
\end{aligned}$$

Moreover, if $\mathbb{E}\Delta > 0$, therefore, it is possible to have $\beta(\underline{\theta}) \geq 0$ for k to be close enough to 1. In sum, take $k^* = \sup_k\{k : \beta'(\theta) \geq 0 \text{ and } \beta(\underline{\theta}) \geq 0\}$, therefore, $\beta^k(\theta)$ will be monotonic and non-negative. It is also easy to see that $\beta^k(\theta) \leq \frac{1}{2}S(\theta,\theta)$ by the similar derivation:

$$\begin{aligned}
\beta^k(\theta) &= \frac{\int_{F^{-1}(k)}^{\theta}(F(\tau)-k)(S(\tau,\tau)-h(\tau))dF(\tau)}{[F(\theta)-k]^2} \\
&\leq \frac{\int_{F^{-1}(k)}^{\theta}(F(\tau)-k)(S(\tau,\tau)-\frac{\partial}{\partial\tau}v(x^*(\tau,\tau),\tau)\frac{\int_\tau^{\bar{\theta}} dF(z)}{f(\tau)})dF(\tau)}{[F(\theta)-k]^2} \\
&= \frac{\int_{F^{-1}(k)}^{\theta}(F(\tau)-k)\left[S(\tau,\tau)-\frac{\partial}{\partial\tau}v(x^*(\tau,\tau),\tau)\frac{1-F(\tau)}{f(\tau)}\right]dF(\tau)}{[F(\theta)-k]^2} \\
&= \frac{1}{2}S(\theta,\theta) - \frac{\frac{1}{2}(1-k)\int_{F^{-1}(k)}^{\theta}\frac{dS(\tau,\tau)}{d\tau}(F(\tau)-k)d\tau}{[F(\theta)-k]^2} \\
&\leq \frac{1}{2}S(\theta,\theta)
\end{aligned}$$

Q.E.D. ■

For intuition, we provide a concrete example.

Example 3.3: *Two residents are living in a small town. They can build a public good (like internet) together and share with each other or build the good on their own. Suppose that the utility function for each individual i over the size of public good x is $v(x,\theta_i) = \theta_i(2x - x^2)$,* *and the cost of x is $c(x) = cx^2$. We assume type θ_i drawn from $[a, 1]$ with c.d.f. $F(\theta) = \frac{\theta-a}{1-a}$. If they build the good autarkily, they choose $x_i(\theta_i) \in \arg\max_x v(x, \theta_i)$, particularly, $\underline{u}(\underline{\theta}) = \max_x v(x, \underline{\theta})$*[10]*. Does there exist an ex post socially efficient mechanism for them to cooperate? Note that*

$$x^*(\boldsymbol{\theta}) = \frac{\sum \theta_i}{c + \sum \theta_i}$$

and

$$S(\boldsymbol{\theta}) = \frac{(\sum \theta_i)^2}{c + \sum \theta_i}$$

In this example, use $c = 1$, $a = \frac{1}{2}$, and it can be computed $\mathbb{E}\Delta = 4(6\ln 5 - 9\ln 2 - 3\ln 3) = 0.490 > 2u(\underline{\theta}) = \frac{1}{3}$.

For M2 with k=1, we have

$$h(\tau) \equiv \frac{\int_\tau^1 \frac{\partial}{\partial \tau} v(x^*(\tau, z), \tau) dF(z)}{f(\tau)} = 1 - \tau - \frac{1}{1+2\tau} + \frac{1}{2+\tau}$$

and

$$\begin{aligned}
\beta^1(\theta) &= \frac{\int_\theta^1 (1 - F(\tau))(S(\tau,\tau) - h(\tau)) dF(\tau)}{[1 - F(\theta)]^2} \\
&= \frac{\int_\theta^1 (1-\tau)(\frac{4\tau^2}{1+2\tau} - (1 - \tau - \frac{1}{1+2\tau} + \frac{1}{2+\tau})) d\tau}{(1-\theta)^2} \\
&= -\frac{1}{2(1-\theta)^2}[1 + [\theta(5 - 2\theta) - 4]\theta + \ln 27 - 6\ln(\theta + 2) + 3\ln(1 + 2\theta)]
\end{aligned}$$

It can be shown that $\beta^1(\theta)$ is monotone and non-negative (In the following Fig. The red dot line is $\frac{1}{2}S(\theta, \theta)$, the green dots line is $\underline{u}(\theta)$ and the black solid line is $\beta^1(\theta)$).

[10] We can verify that $\arg\min_{\theta_i} \mathbb{E}S(\theta_i, \theta_j) - \underline{u}(\theta_i) = \underline{\theta} = \frac{1}{2}$ due to monotonicity of $\mathbb{E}S(\theta_i, \theta_j) - \underline{u}(\theta_i)$ in θ_i.

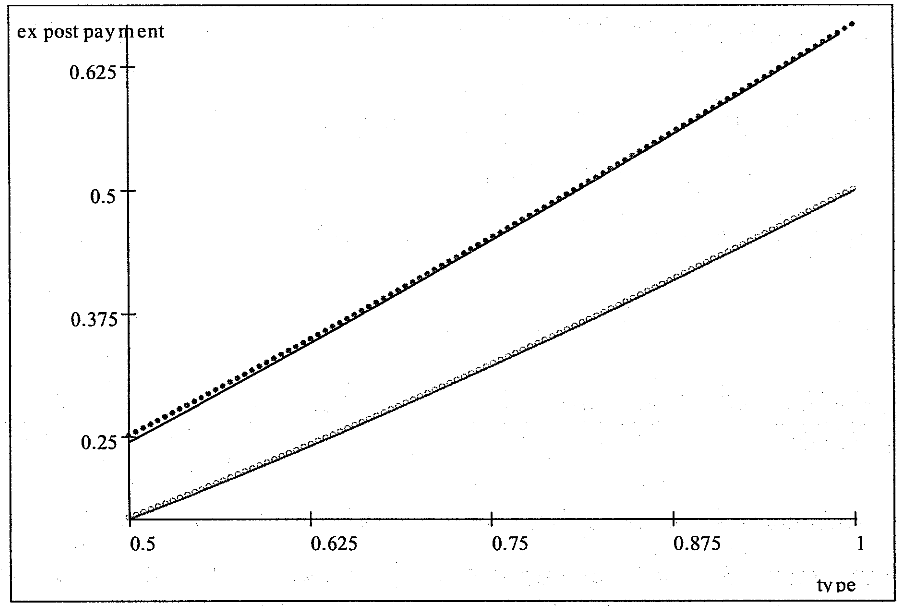


Fig. 2. Ex post payoff functions

To verify ex post IR, we find $\beta^1(\theta) - \underline{u}(\theta) > 0$, *and* $s(\theta_i, \theta_j) - \beta^1(\theta_j) > \underline{u}(\theta_i)$ *for any* $\theta_i > \theta_j$ *by noting that*

$$\frac{(\sum \theta_i)^2}{1 + \sum \theta_i} - \frac{\theta_i^2}{1+\theta_i} > \frac{2\theta_j^2}{1+2\theta_j}$$

since the LHS of the above inequality is an increasing function of θ_i.

We also take a look at M2 with $k = 0$. *The payment function is,*

$$\begin{aligned} \beta^0(\theta) &= \frac{\int_{\frac{1}{2}}^{\theta} F(\tau)(S(\tau,\tau) - h(\tau))dF(\tau)}{F(\theta)^2} \\ &= \frac{2(-\frac{3}{8} + 2\theta - \frac{7}{2}\theta^2 + 2\theta^3 - \ln\frac{3125}{128} + 5\ln(2+\theta) - 2\ln(1+2\theta))}{(1-2\theta)^2} \end{aligned}$$

Therefore, ex post IR

$$\underline{u}(\theta) \leq \beta^0(\theta) \leq S(\theta, \underline{\theta}) - \underline{u}(\theta)$$

fails.

Meanwhile, M3 with $k = 0$ *proposed by the next subsection is also not ex post individual rational i.e., the payment or bid does not satisfy*

$$\underline{u}(\theta) \leq b^0(\theta) \leq S(\theta, \theta) - \underline{u}(\theta)$$

Although when $\underline{u}(\theta) = 0$, it will be true, $0 \le b^0(\theta) \le S(\theta, \theta)$.

Remark 3.6: *The mechanism considered by CGK is not ex post individually rational, even proceeded with a side-payment. One of the advantages of M2 is without need of side-payments.*

3.4.2 Fixed Endowment

Basics

When the endowment is fixed, then VCG always runs expected social surplus. This observation comes from the following:

$$\begin{aligned}
\mathbb{E}\Delta &= \int_{\boldsymbol{\theta}} S(\boldsymbol{\theta}) d\mathbf{F}(\boldsymbol{\theta}) - \sum \int_{\theta_{-i}} \left(\int_{\underline{\theta}_i}^{\bar{\theta}_i} (1 - F_i(\theta_i)) \frac{\partial S(\theta_i, \theta_{-i})}{\partial \theta_i} d\theta_i \right) d\mathbf{F}_{-i}(\boldsymbol{\theta}_{-i}) \\
&= \sum \left(\int_{\boldsymbol{\theta}} v_i(x_i^*(\theta_i, \theta_{-i}), \theta_i) d\mathbf{F}(\boldsymbol{\theta}) - \int_{\theta_{-i}} \left(\int_{\underline{\theta}_i}^{\bar{\theta}_i} (1 - F_i(\theta_i)) \frac{\partial v_i(x_i^*(\theta_i, \theta_{-i}), \theta_i)}{\partial \theta_i} d\theta_i \right) d\mathbf{F}_{-i}(\boldsymbol{\theta}_{-i}) \right) \\
&\ge \sum \left(\int_{\boldsymbol{\theta}} v_i(x_i^*(\theta_i, \theta_{-i}), \theta_i) d\mathbf{F}(\boldsymbol{\theta}) - \int_{\theta_{-i}} \left(\int_{\underline{\theta}_i}^{\bar{\theta}_i} (1 - F_i(\theta_i)) dv_i(x_i^*(\theta_i, \theta_{-i}), \theta_i) \right) d\mathbf{F}_{-i}(\boldsymbol{\theta}_{-i}) \right) \\
&= \sum \int_{\theta_{-i}} v_i(x_i^*(\underline{\theta}_i, \theta_{-i}), \underline{\theta}_i) d\mathbf{F}_{-i}(\boldsymbol{\theta}_{-i}) \\
&\ge 0
\end{aligned}$$

The reason that $\frac{\partial S(\theta_i,\theta_{-i})}{\partial \theta_i} = \frac{\partial v_i(x_i^*(\theta_i,\theta_{-i}),\theta_i)}{\partial \theta_i}$ is similar to lemma 3; the third equality is due to $\frac{\partial v_i(x_i^*(\theta_i,\theta_{-i}),\theta_i)}{\partial \theta_i} \le \left[\frac{\partial v_i(x_i^*(z_i,\theta_{-i}),\theta_i)}{\partial z_i} \right]_{\theta_i = z_i} + \frac{\partial v_i(x_i^*(\theta_i,\theta_{-i}),\theta_i)}{\partial \theta_i}$.

However, **M2** is no longer ex post individually rational. To see this, take $n = 2$ as an example, it is noted that now $\frac{\partial}{\partial z} v(x^*(\tau, z), \tau)$ is a decreasing function of z. It is no longer the case that theorem 5 holds in general[11]. We propose the following allocation rule (Reverse Order Allocation).

M3: *(i) The designer chooses the optimal allocation rule $\mathbf{x}^*(\hat{\boldsymbol{\theta}})$ according to report $\hat{\boldsymbol{\theta}}$;*

---

[11]For example, suppose $v(x_i, \theta_i) = \theta_i \sqrt{x_i}$, $\theta_i \tilde{} U[0,1]$, $\sum x_i = 1$, then $S(\theta_i, \theta_j) = \sqrt{\sum \theta_i^2}$, in this case, it is impossible to have either $\beta^1(\theta) < S(\theta, \theta)$ or $\beta^0(\theta) > 0$ overall.

(ii) each player receives consumption $x_i^*(\hat{\theta}_i, \hat{\theta}_{-i})$, and the payment rule is the following:

$$M_i^T(\hat{\theta}_i, \hat{\theta}_{-i}) = \begin{cases} S(\hat{\theta}_i, \hat{\theta}_{-i}) - [(1-k)r^k(\theta_i) + kr^k(\hat{\theta}^{2:n})] & \text{if } \hat{\theta}_i < \min_{j \neq i} \hat{\theta}_j \\ \frac{1}{\#\mathcal{N}_j} \sum \{S(\hat{\theta}_i, \hat{\theta}_{-i}) - [(1-k)r^k(\theta_i) + kr^k(\hat{\theta}^{2:n})]\} & \text{if } \hat{\theta}_i = \min_{j \neq i} \hat{\theta}_j \text{ for } i, j \in \mathcal{N}_j \\ -v_i(x_i^*(\hat{\theta}_i, \hat{\theta}_{-i}), \hat{\theta}_i) + [(1-k)r^k(\hat{\theta}^{1:n}) + kr^k(\hat{\theta}^{2:n})] & \text{if } \hat{\theta}_i > \min_{j \neq i} \hat{\theta}_j \end{cases}$$

where

$$r^k(\theta) = \frac{\int_{F^{-1}(1-k)}^{\theta} \frac{[k-1+F(\tau)]^{n-1}}{(1-F(\tau))^{n-2}} (\sum_{j=1}^{n} \frac{d}{d\tau} \bar{v}_i^{(j)}(x_i^*(\tau, .), \tau) - m'(\tau) - \frac{d}{d\tau} \bar{S}^{(n)}(\tau)) d\tau}{[k-1+F(\theta)]^n} \tag{3.9}$$

Similarly, it can be justified that the above mechanism is (i) budget balanced and incentive compatible, and (ii) interim socially efficient if and only if $\mathbb{E}\Delta \geq 0$ for any $k \in [0, 1]$. **(See Appendix A6 for detail)**.

Two-player case

We claim M3 with $k = 1$ has nice properties in dealing with an exogeneous endowment case.

When $n = 2$,

$$r(\theta) = \frac{\int_{F^{-1}(1-k)}^{\theta} [k-1+F(\tau)](S(\tau, \tau) + \mu(\tau)) dF(\tau)}{[k-1+F(\theta)]^2}$$

where

$$\mu(\tau) = \frac{\int_{\underline{\theta}}^{\tau} \frac{\partial}{\partial \tau} v(x^*(\tau, z), \tau)) dF(z))}{f(\tau)}$$

And truth telling is a globally optimal strategy.

Theorem 3.7: *When $n = 2$, M3 is ex post socially efficient and satisfies ex post monotonicity.*

Proof. It is obvious that $r^1(\theta)$ is non-negative. And note that

$$\begin{aligned}
r^1(\theta) &= \frac{\int_{\underline{\theta}}^{\theta} F(\tau)(S(\tau,\tau)+\mu(\tau))dF(\tau)}{F(\theta)^2} \\
&\le \frac{\int_{\underline{\theta}}^{\theta} F(\tau)S(\tau,\tau)dF(\tau) + \int_{\underline{\theta}}^{\theta} F(\tau)^2 \frac{d}{d\tau} S(\tau,\underline{\theta})d\tau}{F(\theta)^2} \\
&= \frac{\int_{\underline{\theta}}^{\theta} F(\tau)S(\tau,\tau)dF(\tau) + F(\theta)^2 S(\theta,\underline{\theta}) - 2\int_{\underline{\theta}}^{\theta} F(\tau)S(\tau,\underline{\theta})dF(\tau)}{F(\theta)^2} \\
&\le S(\theta,\underline{\theta})
\end{aligned}$$

The last step is due to submodularity $S(\underline{\theta}_i,\theta_j) \ge \frac{1}{2}[S(\theta_j,\theta_j)+S(\underline{\theta}_i,\underline{\theta}_i)] \ge \frac{1}{2}S(\theta_j,\theta_j)$[12]. Thus $S(\theta_i,\theta_j) - r^1(\theta_j) \ge 0$ for any $\theta_i \le \theta_j$.

To verify the ex post monotonicity, first of all, note that

$$\begin{aligned}
r^1(\theta) &= \frac{\int_{\underline{\theta}}^{\theta} F(\tau)(S(\tau,\tau)+\mu(\tau))dF(\tau)}{F(\theta)^2} \\
&= \frac{1}{2}S(\theta,\theta) + \frac{2\int_{\underline{\theta}}^{\theta} F(\tau)\mu(\tau)dF(\tau) - \int_{\underline{\theta}}^{\theta} F(\tau)^2 \frac{d}{d\tau}S(\tau,\tau)d\tau}{2F(\theta)^2} \\
&\ge \frac{1}{2}S(\theta,\theta) + \frac{2\int_{\underline{\theta}}^{\theta} F(\tau)^2 \frac{\partial}{\partial\tau} v(x^*(\tau,\tau),\tau))d\tau - \int_{\underline{\theta}}^{\theta} F(\tau)^2 \frac{d}{d\tau}S(\tau,\tau)d\tau}{2F(\theta)^2} \\
&= \frac{1}{2}S(\theta,\theta)
\end{aligned}$$

which implies $r^{1\prime}(\theta) \ge \frac{1}{2}\frac{d}{d\theta}S(\theta,\theta) \ge 0$ since $r^1(\underline{\theta}) = \frac{1}{2}S(\underline{\theta},\underline{\theta})$. Q.E.D. ■

However, $r^0(\theta)$ might not be always ex post individually rational. Note that the ex post individual rationality requires,

$$S(\theta_i,\theta_j) - r^0(\theta_i) \ge 0 \text{ for any } \theta_j \ge \theta_i.$$

When $\theta_j \to \theta_i \to \underline{\theta}$, it is possible that $S(\underline{\theta}_i,\underline{\theta}_j) - r^0(\underline{\theta}) = S(\underline{\theta}_i,\underline{\theta}_j) - \mathbb{E}\Delta < 0$ when $\mathbb{E}\Delta > 0$ but $S(\underline{\theta}_i,\underline{\theta}_j) = 0$.

[12] This is due to the fact that $x_i^*(\theta_i,\theta_j),\theta_i)$ must decrease with θ_j since $\sum x_i = \bar{x}$. Therefore, based on supmodularity of $v(x_i^*(\theta_i,\theta_i),\theta_i)$, $S(\theta_i,\theta_j) + S(\theta_j,\theta_i) \ge S(\theta_i,\theta_i) + S(\theta_j,\theta_j)$.

Example 3.4: *The utility is the same as in example 3, but* $\Sigma x_i = 1$. *We have*

$$S(\theta) = \sum \theta_i - \frac{\theta_i \theta_j}{\theta_i + \theta_j}$$

$$x_i(\theta_i) = \frac{\theta_i}{\theta_i + \theta_j}$$

$$\mathbb{E}\Delta = 2\mathbb{E}S(0,\theta) - \int_0^1 \int_0^1 \left(\theta_1 + \theta_2 - \frac{\theta_1 \theta_2}{\theta_1 + \theta_2}\right) d\theta_1 d\theta_2 = 2 - 0.795 > 0$$

And $\mu(\tau) = (\ln 4 - \frac{1}{2})\tau$ *and* $r^0(\theta) = \frac{\ln 4 + 1}{6}(2\theta + 1)$. *So* $r^0(\theta)$ *is not ex post IR. But* $\frac{1}{2}S(\theta, \theta) = \frac{3}{4}\theta < r^1(\theta) = \frac{\ln 4+1}{3}\theta \leq S(\underline{\theta}, \theta) = \theta$, *therefore* $r^1(\theta)$ *is ex post IR and satisfies ex post monotonicity. In fact, as we will show later, there is a window of distribution of endowment satisfying ex post IR even if the outside reservation is type-dependent.*

3.4.3 Discussion for case of n>2

The above allocation rules do not work well when $n > 2$. The issue is that $\beta^k(\theta)$ or $r^k(\theta)$ is no longer always bounded (either from below or above). Take $\beta^1(\theta)$ as an example, when $\theta \to \underline{\theta}$, $\beta^1(\underline{\theta})$ will be not bounded in general, due to the fact that item $\frac{\int_{\theta_{-i}} \frac{\partial}{\partial \tau} S(\theta, \theta_{-i}) d\mathbf{F}_{-i}(\theta_{-i})}{F(\theta)^{n-2}}$ will dominate all other items when $n > 2$ as $\theta \to \underline{\theta}$. The economic intuition is that, due to the externality, the subsidy to the boundary type will be too high, once the number of player is more than 2. This situation happens not only in $\beta^k(\theta)$ or $r^k(\theta)$, but also under any allocation rule such as giving the entitlement to the j-th order highest type. Whether $\beta^k(\theta)$ or $r^k(\theta)$ is still valid depends on the functional form of utility $v(x, \theta)$. This fact demonstrates the sensitivity of the choice of a mechanism.

3.5 Auction-like Implementation and Bilateral Trade

3.5.1 Implement mechanism by auction

The mechanisms proposed in previous sections, M1, M2, or M3, can be implemented by a realisitic form of auction, especially under symmetric independent environment. For example, the first price auction with post auction redistribution will exactly implement the payment rule in M1 under SIPV. To see this, let b be the bidding function, then under the first price rule, the payoff is:

$$u_i(\theta_i) = \begin{cases} S(\theta_i) - b_i & \text{if } b_i > \max_{j \neq i} b_j \\ \frac{1}{\# \mathcal{N}_j} \sum [S(\theta_i) - b_i] & \text{if } b_i = b_j \ \forall i, j \in \mathcal{N}_j \\ \frac{1}{n-1} \max_{j \neq i} b_j & \text{if } b_i < \max_{j \neq i} b_j \end{cases} \tag{3.10}$$

In equilibrium, the optimal bidding strategy will be consistent with $\frac{n-1}{n}\mathbb{E}[S(\tau^{n:n})/\tau^{n:n} \leq \theta]$. However, the ex post individual rationality will be senstive to the payment rule. For example, the second price auction in our context will not be ex post individually rational, though it is still interim individually rational. In general, we can state our auction rule as follows:

(1) Government/designer runs a sealed price auction to pick a winner, and deterimines the supply of products according to the collected messages.

(2) The winner pays the cost, but gets the entitlement to charge the remaining bidders according to their consumption given their announced types;

(3) In order to obtain this entitlement, the winner(s) pays a lump sum transfer to the remaining losers. The rule of the game is known before bidding starts (in M2 who bids the highest wins; but in M3 who bid lowest wins).

(4) The revenue (deficit) collected from the winner's payment is redistributed among all bidders (perhaps including the winner himself), the redistribution policy is known before bidders submit their bids.

Our advice for chioce of auction rule can be summarized as the following.

Table 2: Chioce of auction rule in different context

	Private good	Public good	Public good
Endowment		Endogeneous	Exogeneous
Title Allocation	The highest type	The highest type	The lowest type
Price Rule	1st price auction	Second price auction	Second price auction
# of Bidder	$n \geq 2$	$n = 2$	$n = 2$

In SIPV environment, based on our first price auction, the auctioneer can earn a risk-free revenue that is the same as any efficient auction in terms of expectation. So if the auctioneer is risk averse, he can induce the bidders to compete for the entitlement, and only charge the participants an entry fee. The participants then will behave exactly as what we have described in section 3. And this auction will be ex post individually rational, so that nobody will quit due to an outside option. Another important implication is that this auction form is also collusion-proof, from the auctioneer's prespective.

Moreover, in public good case, the present auction form possesses several important advantages, compared with auction of share when the quantity is continuous. In continuous quantity or multi-unit auction case, the bidder needs to submit their demand function, and in general, the allocation is inefficient (Wilson, 1979, Ausubel, 2004). However, in our formulation, the allocation is efficient, and the auctioneer's revenue is maximized among all efficient allocations. For intuition, we provide the following comparisons.

Suppose the auctioneer has one unit good for sale, say, $\sum x_i = 1$. The bidder i's utility from x_i is $2\theta_i\sqrt{x_i}$. If the auctioneer runs the uniform price auction, it can be shown that the price rule is

$$p^*(\theta_i, \theta_j) = \sqrt{\theta_i^2 + \theta_j^2} \text{ with demand function } x_i^d = \frac{\theta_i^2}{p^2}.$$

Under this pricing rule, the agent's expected utility at interim stage will be

$$U_i(\theta_i) = \mathbb{E}(2\theta_i\sqrt{x_i^d} - p^* x_i^d) = \theta_i^2 \mathbb{E}\frac{1}{\sqrt{\theta_i^2 + \theta_j^2}} \geq 0$$

Therefore, the total expected revenue will be

$$R^p = \mathbb{E}p^*(\theta_i, \theta_j)$$

For simplicity, suppose θ_i is uniformly drawn from $[0, 1]$, then

$$R^p = \frac{1}{3}(\sqrt{2} + \arg\sinh 1) \cong 0.765.$$

If we run the auction proposed by M3, the auctioneer can charge a risk free revenue

$$\mathbb{E}\Delta = 4\mathbb{E}\theta_j - 2\mathbb{E}\sqrt{\theta_i^2 + \theta_j^2} = 2 - 2 * 0.765. < R^p$$

But when $\underline{\theta}$ is large enough, $\mathbb{E}\Delta > R^p$ will be possible.

3.5.2 Type-dependent Reservation Utility and Bilateral Trade

Endogeneous quantity supply

If the outside reservation utility is type-dependent, *a la* Myerson-Satterthwaite (1983), do theorem 5 and theorem 6 still hold? It depends on whether endowment is flexible or not. Theorem 5 is still true, but theorem 6 does not hold. Suppose each individual i's reservation utility is $\underline{u}_i(\theta)$, and let $S^{\#}(\theta_i, \theta_j) = S(\theta_i, \theta_j) - \sum \underline{u}_i(\theta_i)$ be the net social improvement, and $v_i^{\#}(x_i, \theta_i) = v_i(x_i, \theta_i) - \frac{1}{2}\sum \underline{u}_i(\theta_i)$ be the net utility improvement of individual i. Therefore, the payment rule will be based on $v_i^{\#}(x_i, \theta_i)$. Under mechanism M2, the payment rule can be modified as follows:

$$M_i^{1\#}(\hat{\theta}_i, \hat{\theta}_j) = \begin{cases} -S_{-i}^{\#}(\hat{\theta}_i, \hat{\theta}_j) + \beta^{1\#}(\hat{\theta}_j) + \frac{1}{2}(\underline{u}_i(\underline{\theta}_i) - \underline{u}_j(\underline{\theta}_j)) & \text{if } \hat{\theta}_i > \hat{\theta}_j \\ \frac{1}{2}\sum[S(\hat{\theta}_i, \hat{\theta}_{-i}) - \beta^{1\#}(\hat{\theta}_i)] & \text{if } \hat{\theta}_i = \hat{\theta}_j \\ v_i^{\#}(x_i(\hat{\theta}_i, \hat{\theta}_j), \hat{\theta}_i) - \beta^{1\#}(\hat{\theta}_i) - \frac{1}{2}(\underline{u}_i(\underline{\theta}_i) - \underline{u}_j(\underline{\theta}_j)) & \text{if } \hat{\theta}_i < \hat{\theta}_j \end{cases}$$

where

$$\begin{aligned} \beta^{1\#}(\theta) &= \frac{\int_{\theta}^{\bar{\theta}}(1 - F(\tau))(S^{\#}(\tau, \tau) - h^{\#}(\tau))dF(\tau)}{[1 - F(\theta)]^2} + \frac{1}{2}(\underline{u}_i(\underline{\theta}_i) - \underline{u}_j(\underline{\theta}_j)) \\ &= \beta^1(\theta) - \frac{1}{2}\sum \underline{u}_i(\theta) + \frac{1}{2}(\underline{u}_i(\underline{\theta}_i) - \underline{u}_j(\underline{\theta}_j)). \end{aligned}$$

The following proposition justifies that M2 is still ex post individually rational.

Proposition 3.6: *If the reservation utility is type-dependent $\underline{u}_i(\theta_i)$, when n=2, M2 with k=1 is still ex post socially efficient if and only if $\mathbb{E}\Delta \geq \sum \underline{u}_i(\theta_i^{\#})$, where $\theta_i^{\#} = \arg\min_{\theta_i} E_{\theta_j} S(\theta_i, \theta_j) - \underline{u}_i(\theta_i)$.*

Proof. It can be shown that

$$\begin{aligned} \beta^{1\#}(\theta) &= \frac{\int_{\theta}^{\bar{\theta}}(1 - F(\tau))(S^{\#}(\tau, \tau) - h^{\#}(\tau))dF(\tau)}{[1 - F(\theta)]^2} + \frac{1}{2}(\underline{u}_i(\underline{\theta}_i) - \underline{u}_j(\underline{\theta}_j)) \\ &= \frac{\int_{\theta}^{\bar{\theta}}(1 - F(\tau))(S(\tau, \tau) - h(\tau))dF(\tau)}{[1 - F(\theta)]^2} - \underline{u}(\theta) + \frac{1}{2}(\underline{u}_i(\underline{\theta}_i) - \underline{u}_j(\underline{\theta}_j)) \\ &\leq \frac{1}{2}S(\theta, \theta) - \frac{1}{2}\sum \underline{u}_i(\theta) + \frac{1}{2}(\underline{u}_i(\underline{\theta}_i) - \underline{u}_j(\underline{\theta}_j)). \end{aligned}$$

Note that $\frac{d}{d\theta}\left(\frac{1}{2}S(\theta,\theta) - \frac{1}{2}\sum \underline{u}_i(\theta)\right) \geq 0$, and $\beta^{1\#}(\bar{\theta}) = \frac{1}{2}S(\bar{\theta},\bar{\theta}) - \frac{1}{2}\sum \underline{u}_i(\bar{\theta})$, therefore, $\beta^{1\#}(\theta)$ will be monotonically increasing in θ. We want to show the following two inequalities:

$$\begin{aligned} S(\theta_i,\theta_j) - \frac{1}{2}(\underline{u}_i(\theta_i) + \underline{u}_j(\theta_i)) - \beta^{1\#}(\theta_i) &\geq \underline{u}_j(\theta_j) \text{ for } \theta_i \leq \theta_j \\ \frac{1}{2}(\underline{u}_i(\theta_i) + \underline{u}_j(\theta_i)) + \beta^{1\#}(\theta_i) &\geq \underline{u}_i(\theta_i) \text{ for } \theta_i \leq \theta_j \end{aligned}$$

For the first inequality, by the supermodularity of $S(\theta_i,\theta_j)$, we have

$$\begin{aligned} & S(\theta_i,\theta_j) - \frac{1}{2}(\underline{u}_i(\theta_i) + \underline{u}_j(\theta_i)) - \beta^{1\#}(\theta_i) - \underline{u}_j(\theta_j) - \frac{1}{2}(\underline{u}_i(\underline{\theta}_i) - \underline{u}_j(\underline{\theta}_j)) \\ \geq\ & S(\theta_i,\theta_j) - \underline{u}_j(\theta_j) - \frac{1}{2}S(\theta_i,\theta_i) - \frac{1}{2}(\underline{u}_i(\underline{\theta}_i) - \underline{u}_j(\underline{\theta}_j)) \\ \geq\ & \frac{1}{2}S(\theta_i,\theta_i) - \underline{u}_j(\theta_i) - \frac{1}{2}(\underline{u}_i(\underline{\theta}_i) - \underline{u}_j(\underline{\theta}_j)) \\ \geq\ & \frac{1}{2}S(\underline{\theta}_j,\underline{\theta}_j) - \underline{u}_j(\underline{\theta}_j) - \frac{1}{2}(\underline{u}_i(\underline{\theta}_i) - \underline{u}_j(\underline{\theta}_j)) \\ \geq\ & 0 \end{aligned}$$

For the second inequality, note that by the proof of theorem 5,

$$\frac{1}{2}\sum \underline{u}_i(\theta_i) + \beta^{1\#}(\theta_i) - \underline{u}_i(\theta_i) = \beta^1(\theta_i) - \underline{u}_i(\theta_i) + \frac{1}{2}(\underline{u}_i(\underline{\theta}_i) - \underline{u}_j(\underline{\theta}_j))$$

will increase with θ_i due to $\frac{1}{2}\frac{dS(\theta_i,\theta_i)}{d\theta_i} \geq \underline{u}_i'(\theta_i)$, therefore it sufficies to show,

$$\beta^{1\#}(\underline{\theta}) = \frac{1}{2}\mathbb{E}\Delta - \underline{u}_i(\underline{\theta}) + \frac{1}{2}(\underline{u}_i(\underline{\theta}_i) - \underline{u}_j(\underline{\theta}_j)) = \frac{1}{2}[\mathbb{E}\Delta - \sum \underline{u}_i(\underline{\theta})] \geq 0.$$

Note that $\underline{\theta} = \arg\min_{\theta_i}\{\mathbb{E}_{\theta_j}S(\theta_i,\theta_j) - \underline{u}_i(\theta_i)\}$, therefore, $\beta^{1\#}(\underline{\theta}) \geq 0$ if and only if $\mathbb{E}\Delta - \sum \underline{u}_i(\underline{\theta}) \geq 0$. Q.E.D. ■

The above justification applies to the trade problem where the quantity is endogeneous. Unfortunately, this conclusion may not hold for an endowment economy, as we will see below.

Fixed quantity supply

Suppose the quantity of supply is fixed at $\sum \bar{x}_i = 1$, and let $\underline{u}_i(\theta) = v(\bar{x}_i,\theta)$ be the alternative option value of trade. According to Krishna and Perry (1998) or Palfrey and Ledyard

(2007), there exist interim socially efficient trade mechanisms if and only if $\mathbb{E}\Delta - \sum \underline{u}_i(\underline{\theta}_i^{\#}) \geq 0$, where $\underline{\theta}_i^{\#} = \arg\min_{\theta_i} \mathbb{E}_{\theta_j} S(\theta_i, \theta_j) - \underline{u}_i(\theta_i)$. So far we know that the classic Myerson-Satterwaite scenario fails this condition when preference is linear, initial endowment is extreme, and the quantity is indivisible. Any change of one of these three conditions may cause differences of consequence. For example, $\mathbb{E}\Delta - \sum \underline{u}_i(\underline{\theta}_i^{\#}) \geq 0$ happens when initial endowment is fairly symmetric (Crampton, Gibbons and Klemperer, 1986), even if utility is still linear[13]. Furthermore, if the agent's utility is concave, then existences appear when the initial endowment is either fairly symmetric or extreme[14], as we will see below. However, once we take the ex post IR into consideration, the results are significantly changed. The following propostion describes several impossibility results.

Proposition 3.7: *If utility is identical $v(x_i, \theta_i)$ with i.i.d. θ_i, and total endowment is fixed, in the any of following situations, there does not exist an ex post socially efficient mechanism:*

(i) the utility $v(x_i, \theta_i)$ is linear $\theta_i x_i$, for any initial endowment allocation;

(ii) the lowest type $v(x_i, \underline{\theta}) = 0$, for any $v(x_i, \theta)$ and extreme initial endowment allocation.

(Proof see Appendix A6)

Part (i) says that the possibility of trade will dispear again once ex post IR is considered in a partner dissolving game where each bidder initially owns some lottery or share of the object. Part (ii) indicates that the lowest type agent's value plays some subtle role, interacting with the endowment allocation. If $v(x_i, \underline{\theta}) = 0$, the designer is not able to punish the lowest type agent, therefore whole incentive scheme will be affected.

To detect the existence of an ex post socially efficient mechanism, we provide the following sufficient condition based on M3..

Proposition 3.8: *If the utility is symmetric and quasi-linear $v(x_i, \theta_i) = \theta_i \phi(x_i)$, there exist an ex post socially efficient mechanism if one of the traders' initial endowment satisfies the following conditions:*

$$\phi^{-1}\left[2\phi(\frac{1}{2}) - \frac{r^1(\theta_i)}{\theta_i}\right] \geq \bar{x}_i \geq \frac{1}{2} \tag{3.11}$$

[13]MacAfee (1992) shows that if endowment is uncertain, then there exists an interim socially efficient mechanism too.

[14]When initial endownment is extremely distributed, there exists an interim socially efficient trade mechanism if and only if $\mathbb{E}\Delta + \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\bar{\theta}} \frac{\partial}{\partial \tau} v(x_S^*(\tau, \theta_j), \tau) d\tau dF(\theta_j) \geq v(1, \bar{\theta})$.

where $r^1(\theta_i)$ is defined as formula (11). **(Proof see Appendix A8)**

Proof. Based on M3, ex post IR requires two inequalities

$$\begin{aligned} S(\theta_i,\theta_j) - r^1(\theta_j) &\geq \underline{u}_i(\theta_i) \text{ if } \theta_i \leq \theta_j \\ r^1(\theta_j) &\geq \underline{u}_j(\theta_j) \text{ if } \theta_i \leq \theta_j \end{aligned}$$

Note that $S(\theta_i,\theta_j) - \underline{u}_i(\theta_i)$ decreases with θ_i, the first inequality holds if and only if $S(\theta_j,\theta_j) - r^1(\theta_j) \geq \underline{u}_i(\theta_j)$; and the second inequality holds if and only if $\frac{1}{2}S(\theta_j,\theta_j) \geq \underline{u}_j(\theta_j)$. Substituting $x^*(\theta_j,\theta_j) = \frac{1}{2}$ into these inequalitis, we obtain (11). Q.E.D. ■

The window of initial endowment allocation needs to meet two conditions; on one hand, it should not too asymmetric to meet interim IR constraint, on the other hand, it should not be too symmetric, allowing ex post IR to hold. The following explicit example demonstrates the intuition.

Example 3.5. *From example 4, under M3, ex post IR requires two inequalities*

$$\begin{aligned} S(\theta_i,\theta_j) - r^1(\theta_j) &\geq \underline{u}_i(\theta_i) \text{ if } \theta_i \leq \theta_j \\ r^1(\theta_j) &\geq \underline{u}_j(\theta_j) \text{ if } \theta_i \leq \theta_j \end{aligned}$$

which requires

$$\theta_i + \theta_j - \frac{\theta_i\theta_j}{\theta_i+\theta_j} - \frac{\ln 4 + 1}{3}\theta_j \geq \theta_i(2x_i - x_i^2)$$

The necessary and sufficient condition is

$$\begin{aligned} \frac{3}{2} - \frac{\ln 4 + 1}{3} &\geq (2x_i - x_i^2) \\ \frac{\ln 4 + 1}{3} &\geq (2x_j - x_j^2) = 1 - x_i^2 \end{aligned}$$

The endowment satisfies the above requirement is

$$0.456 \cong 1 - \frac{1}{6}\sqrt{6}\sqrt{2\ln 4 - 1} \geq x_i \geq \sqrt{1 - \frac{\ln 4 + 1}{3}} \cong 0.452$$

3.6 Conclusion and discussion

The basic findings of the present paper can be summarized as follows. First, in private good environments, we prove that the existence of an ex post IR socially efficient mechanism if and only if the VCG mechanism runs expected social surplus, which is the same as the condition for the existence of interim socially efficient mechanisms. Interestingly, we prove that our mechanism is the generically unique Bayesian mechanism satisfying ex post budget balance, ex post IR and ex post payoff monotonicity, which maximizes the probability of ex post IC. Our mechanism can be implemented through a specific auction for social surplus, which can be regarded as an auction for an entilement associating with post-auction redistributions. Compared with standard auctions, this auction enables the seller to earn a risk-free revenue, and the bidders to be ex post individually rational. It is also worth pointing out that we are able to characterize the bidding strategy explicitly even if the distribution is asymmetric, which is in general hard to solve in a standard first price auction.

Second, in public good environments, we find the flexibility of supply matters. If the supply of quantity is flexible, and when the number of agent is $n = 2$, there exist ex post efficient mechanisms whenever the interim socially efficient mechanisms exist, it does depend on whether or not the IR constraint is type-dependent. If the supply of quantity is fixed like a mutli-unit auction or a divisible good auction, the conclusion is only true for a type-independent outside reservation. In the fixed quantity case, our proposed auction generates a risk-free revenue the same as the expected social surplus of a VCG, which is always efficient, but the seller's revenue might be higher or lower than the uniform price auction. When the IR constraint is type-dependent like a bilaterial trade, there does not exist an ex post socially efficient mechanism even though its interim counterpart exists. For example, there does not exists an ex post socially efficient partner dissolving mechanism even though the initial endowment is symmetric, in contrast to Crampton, Klemperer and Gibbons (1986). We show non-existence of any ex post socially efficient trade if either utility is linear or the lowest type agent gain zero utility with an extreme initial endowment allocation. We also provide a sufficient condition for the existence of ex post socially efficient mechanims and an explicit example. This observation is helpful in understanding no-trade possibility in a more general context, where the trader's preference and the distribution of initial endowment have been taken into consideration.

Last but not least, we raise a concrete example where the number of players matters in determing the existence of an ex post individually rational mechanism. Due to the existence of externality, when $n > 2$, it will be more expensive to fully incorporate externatities as the numbers of players gets larger. Therefore the punishment of low boundary type will be too heavy (or the rewards to the upper bound type are too high), which may break ex post IR. There are several important extensions, including inter-dependent value, affiliation, correlated type or information acquisition in our prescribed auction, but we leave them for future discovery.

3.7 Appendix

3.7.1 A1. Proof of Proposition 1

Proof. (i) Note that

$$\begin{aligned} v_i(x_i^*(\theta_i, \theta_{-i}), \theta_i) + S_{-i}(\theta_i, \theta_{-i}) &\geq v_i(x_i^*(\theta_i', \theta_{-i}), \theta_i) + S_{-i}(\theta_i', \theta_{-i}) \\ v_i(x_i^*(\theta_i', \theta_{-i}), \theta_i') + S_{-i}(\theta_i', \theta_{-i}) &\geq v_i(x_i^*(\theta_i, \theta_{-i}), \theta_i') + S_{-i}(\theta_i, \theta_{-i}) \end{aligned}$$

therefore, $x_i^*(\theta_i, \theta_{-i})$ is non-decreasing in θ_i due to the supermodularity of $v_i(., \theta_i)$. Moreover, note that,

$$\begin{aligned} & v_i(x_i^*(\theta_i, \theta_j, \theta_{-ij}), \theta_i) + \sum_{k \neq i} v_k(x_k^*(.), \theta_k) - c(\mathbf{x}^*(\theta_i, \theta_j, \theta_{-ij})) \\ \geq\ & v_i(x_i^*(\theta_i, \theta_j', \theta_{-ij}), \theta_i) + \sum_{k \neq i} v_k(x_k^*(.), \theta_k) - c(x_i^*(\theta_i, \theta_j', \theta_{-ij}), x_j^*(\theta_i, \theta_j, \theta_{-ij}), x_{-ij}^*) \end{aligned}$$

and

$$\begin{aligned} & v_i(x_i^*(\theta_i, \theta_j', \theta_{-ij}), \theta_i) + \sum_{k \neq i} v_k(x_k^*(.), \theta_k) - c(\mathbf{x}^*(\theta_i, \theta_j', \theta_{-ij})) \\ \geq\ & v_i(x_i^*(\theta_i, \theta_j, \theta_{-ij}), \theta_i) + \sum_{k \neq i} v_k(x_k^*(.), \theta_k) - c(x_i^*(\theta_i, \theta_j, \theta_{-ij}), x_j^*(\theta_i, \theta_j, \theta_{-ij}), x_{-ij}^*) \end{aligned}$$

we have,

$$\begin{aligned}
& c(x_i^*(\theta_i,\theta_j,\theta_{-ij}), x_j^*(\theta_i,\theta_j,\theta_{-ij}), x_{-ij}^*) + c(x_i^*(\theta_i,\theta_j',\theta_{-ij}), x_j^*(\theta_i,\theta_j',\theta_{-ij}), x_{-ij}^*) \\
\leq\ & c(x_i^*(\theta_i,\theta_j',\theta_{-ij}), x_j^*(\theta_i,\theta_j,\theta_{-ij}), x_{-ij}^*) + c(x_i^*(\theta_i,\theta_j,\theta_{-ij}), x_j^*(\theta_i,\theta_j',\theta_{-ij}), x_{-ij}^*)
\end{aligned}$$

If $\theta_j' \geq \theta_j$, $x_j^*(\theta_i,\theta_j',\theta_{-ij}) \geq x_j^*(\theta_i,\theta_j,\theta_{-ij})$, the above inequality implies $x_i^*(\theta_i,\theta_j',\theta_{-ij}) \geq x_i^*(\theta_i,\theta_j,\theta_{-ij})$ by the submodularity of $c(\mathbf{x})$. The same logic applies to $\theta_j' \leq \theta_j$. Therefore, $S(\theta_i,\theta_j,\theta_{-ij})$ is supermodular.

(ii) We can derive the similar property based on the Lagrangian. Note that

$$\begin{aligned}
& v_i(x_i^*(\theta_i,\theta_{-i}),\theta_i) + v_j(x_j^*(\theta_i,\theta_j,\theta_{-ij}),\theta_j) \\
& + \sum_{k\neq i,j} v_k(x_k^*(.),\theta_k) + \lambda^*(\theta_i,\theta_j,\theta_{-ij})[\bar{c} - c(\mathbf{x}^*(\theta_i,\theta_j,\theta_{-ij}))] \\
\geq\ & v_i(x_i^*(\theta_i',\theta_{-i}),\theta_i) + v_j(x_j^*(\theta_i',\theta_j,\theta_{-ij}),\theta_j) \\
& + \sum_{k\neq i,j} v_k(x_k^*(.),\theta_k) + \lambda^*(\theta_i,\theta_j,\theta_{-ij})[\bar{c} - c(x_i^*(\theta_i',\theta_j,\theta_{-ij}), x_j^*(\theta_i',\theta_j,\theta_{-ij}), x_{-ij}^*)]
\end{aligned}$$

and

$$\begin{aligned}
& v_i(x_i^*(\theta_i',\theta_{-i}),\theta_i') + v_j(x_j^*(\theta_i',\theta_j,\theta_{-ij}),\theta_j) \\
& + \sum_{k\neq i,j} v_k(x_k^*(.),\theta_k) + \lambda^*(\theta_i',\theta_j,\theta_{-ij})[\bar{c} - c(\mathbf{x}^*(\theta_i',\theta_j,\theta_{-ij}))] \\
\geq\ & v_i(x_i^*(\theta_i,\theta_{-i}),\theta_i') + v_j(x_j^*(\theta_i,\theta_j,\theta_{-ij}),\theta_j) \\
& + \sum_{k\neq i,j} v_k(x_k^*(.),\theta_k) + \lambda^*(\theta_i,\theta_j,\theta_{-ij})[\bar{c} - c(x_i^*(\theta_i,\theta_j,\theta_{-ij}), x_j^*(\theta_i,\theta_j,\theta_{-ij}), x_{-ij}^*)]
\end{aligned}$$

Because the constraint is still binding, then we obtain inequality

$$v_i(x_i^*(\theta_i,\theta_{-i}),\theta_i) + v_i(x_i^*(\theta_i',\theta_{-i}),\theta_i') \geq v_i(x_i^*(\theta_i',\theta_{-i}),\theta_i) + v_i(x_i^*(\theta_i,\theta_{-i}),\theta_i')$$

which implies that $x_i^*(\theta_i,\theta_{-i})$ increases with θ_i. At the same time, note that when the quantity is constrained, if $x_j^*(\theta_i,\theta_j,\theta_{-ij}) > x_j^*(\theta_i,\theta_j',\theta_{-ij})$, there must be at least for some k,

$x_k^*(\theta_i, \theta_j, \theta_{-ij}) < x_k^*(\theta_i, \theta_j', \theta_{-ij})$. Therefore, we have

$$\begin{aligned}
\frac{\partial}{\partial x_i} v_i(x_i^*(\theta_i, \theta_j', \theta_{-ij}), \theta_i) &= \frac{\partial}{\partial x_k} v_k(x_k^*(\theta_i, \theta_j', \theta_{-ij}), \theta_k) \\
&\geq \frac{\partial}{\partial x_k} v_k(x_k^*(\theta_i, \theta_{-i}), \theta_k) = \frac{\partial}{\partial x_k} v_k(x_k^*(\theta_i, \theta_{-i}), \theta_k)
\end{aligned}$$

implying $x_i^*(\theta_i, \theta_j, \theta_{-ij}) \geq x_i^*(\theta_i, \theta_j', \theta_{-ij})$. Futhermore, $S(\theta_i, \theta_j, \theta_{-ij})$ is submodular. Q.E.D. ■

3.7.2 A2. Proof of Lemma 3.

Proof. By contradiction. Suppose that there is an incentive compatible and budget balance mechanism M^* resulting in

$$\sum U_i^*(\underline{\theta}_i) > \mathbb{E}\Delta$$

We can rewrite $\mathbb{E}\Delta$ as follows:

$$\begin{aligned}
\mathbb{E}\Delta &= \int_{\boldsymbol{\theta}} S(\boldsymbol{\theta}) d\mathbf{F}(\boldsymbol{\theta}) - \sum \int_{\theta_{-i}} \left(\int_{\underline{\theta}_i}^{\bar{\theta}_i} (1 - F_i(\theta_i)) \frac{\partial S(\theta_i, \theta_{-i})}{\partial \theta_i} d\theta_i \right) d\mathbf{F}_{-i}(\boldsymbol{\theta}_{-i}) \\
&= \int_{\boldsymbol{\theta}} S(\boldsymbol{\theta}) d\mathbf{F}(\boldsymbol{\theta}) - \sum \int_{\boldsymbol{\theta}} \frac{1}{\lambda_i(\theta_i)} \frac{\partial S(\theta_i, \theta_{-i})}{\partial \theta_i} d\mathbf{F}(\boldsymbol{\theta})
\end{aligned}$$

where $\frac{1-F_i(\theta_i)}{f_i(\theta_i)} = \frac{1}{\lambda_i(\theta_i)}$. If x_i is continuously differentiable so that $x_i^*(z_i, \theta_{-i})$ is an interior point, then

$$\frac{\partial x_i^*(z_i, \theta_{-i})}{\partial z_i} \left[\sum \frac{\partial v_i(x_i^*(z_i, \theta_{-i}), \theta_i)}{\partial x_i} - \frac{\partial}{\partial x_i} c(\mathbf{x}^*(z_i, \theta_{-i})) \right] = 0$$

If $x_i^*(z_i, \theta_{-i})$ is not continuous, say, a discrete variable, then $x_i^*(z_i, \theta_{-i})$ is a step function specified by (z_i, θ_{-i}), therefore, under every interval that $x_i^*(z_i, \theta_{-i})$ is being applied,

$$\frac{\partial}{\partial z_i} \left[\sum \frac{\partial v_i(x_i^*(z_i, \theta_{-i}), \theta_i)}{\partial x_i} - \frac{\partial}{\partial x_i} c(\mathbf{x}^*(z_i, \theta_{-i})) \right] = 0$$

Therefore, in any case $\frac{\partial S(\theta_i, \theta_{-i})}{\partial \theta_i} = \frac{\partial v_i(x_i^*(\theta_i, \theta_{-i}), \theta_i)}{\partial \theta_i}$ is true. Meanwhile, if $\mathbf{x}$ is exogeneously given, we write $S(\boldsymbol{\theta}) = \max_{\mathbf{x}, \boldsymbol{\lambda}} \sum v_i(x_i, \theta_i) + \lambda[\bar{c} - c(\mathbf{x})]$, it will also be the case $\frac{\partial S(\theta_i, \theta_{-i})}{\partial \theta_i} =$

$\frac{\partial v_i(x_i^*(\theta_i,\theta_{-i}),\theta_i)}{\partial\theta_i}$ since $\bar{c} = c(\mathbf{x})$ is binding at the optimum. Moreover, note that

$$m_i(\tau_i) = m_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\tau_i} \left[\frac{\partial}{\partial z_i}\int_{\theta_{-i}} v_i(x_i^*(z_i,\theta_{-i}),\theta_i) f_{-i}(\boldsymbol{\theta}_{-i}) d\boldsymbol{\theta}_{-i}\right]_{z_i=\theta_i} d\theta_i$$

And integrating by parts, we have

$$\begin{aligned}
\mathbb{E}m_i(\theta_i) &= m_i(\underline{\theta}_i) - \mathbb{E}_{\theta_{-i}} v_i(x_i^*(\underline{\theta}_i,\theta_{-i}),\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\bar{\theta}_i}\int_{\theta_{-i}} v_i(x_i^*(\tau_i,\theta_{-i}),\tau_i) f_{-i}(\boldsymbol{\theta}_{-i}) d\boldsymbol{\theta}_{-i} dF_i(\theta_i) \\
&\quad - \int_{\underline{\theta}_i}^{\bar{\theta}_i} \frac{1}{\lambda_i(\theta_i)}\left(\int_{\theta_{-i}} \frac{\partial v_i(x_i^*(\theta_i,\theta_{-i}),\theta_i)}{\partial\theta_i} f_{-i}(\boldsymbol{\theta}_{-i}) d\boldsymbol{\theta}_{-i}\right) dF_i(\theta_i)
\end{aligned}$$

Therefore, the total money collected under mechanism M^* is

$$\begin{aligned}
\sum \mathbb{E}m_i(\theta_i) &= -\sum U_i(\underline{\theta}_i) + \sum \mathbb{E}_{\boldsymbol{\theta}}[v_i(x_i^*(\theta_i,\theta_{-i}),\theta_i) - \frac{1}{\lambda_i(\theta_i)}\frac{\partial v_i(x_i^*(\theta_i,\theta_{-i}),\theta_i)}{\partial\theta_i}] \\
&< -\mathbb{E}\Delta + \sum \mathbb{E}_{\boldsymbol{\theta}}[v_i(x_i^*(\theta_i,\theta_{-i}),\theta_i) - \frac{1}{\lambda_i(\theta_i)}\frac{\partial v_i(x_i^*(\theta_i,\theta_{-i}),\theta_i)}{\partial\theta_i}] \\
&= \mathbb{E}c(\mathbf{x}^*(\boldsymbol{\theta}))
\end{aligned}$$

This means under M^*, VCG must runs expected deficit. Given this fact, we may construct another mechanism $M^{\#}$ to subsidize each individual to make $M^{\#}$ individually rational (but not budget balanced), i.e.,

$$U_i^{\#}(\underline{\theta}_i) = U_i^*(\underline{\theta}_i) + k_i = \underline{u}_i(\underline{\theta}_i)$$

The total subsidy is

$$\sum k_i = \sum \underline{u}_i(\underline{\theta}_i) - \sum U_i^*(\underline{\theta}_i) < \sum \underline{u}_i(\underline{\theta}_i) - \mathbb{E}\Delta$$

which contridicts lemma 1, given that $|\mathbb{E}\Delta - \sum \underline{u}_i(\theta_i)|$ is the smallest amount of subsidy to make an incentive compatible mechanism to be interim socially efficient, since VCG maximizing the payment within incentive compatible mechanisms. Putting differently, when $\sum k_i = 0$, M^* is an interim socially efficient mechanism, but $\sum \underline{u}_i(\underline{\theta}_i) - \mathbb{E}\Delta > 0$ suggests that there does not exist any interim socially efficient mechanism, a contradiction. Q.E.D. ■

3.7.3 A3. Proof of Theorem 3

Proof. Step 1: payment only depends on the winner's type $\theta^{n:n}$

By contradiction. Suppose in general the payment function $M(\theta_1, \theta_2, ..., \theta_n)$ depends on the ordered type, $\theta_1 \geq \theta_2 \geq ... \geq \theta_n$. Here, $M(\theta_1, \theta_2, ..., \theta_n)$ need not to be differentiable, but $M(\theta_1, \theta_2, ..., \theta_n)$ must be integrable so that $\mathbb{E}_{\theta_{-i}}[M(\theta, \theta_{-i}) \backslash \theta > \max_{j \neq 1} \theta_j]$ is differentiable.

We use $f^{(n)}(\theta)$ to denote the joint pdf of all n order statistics as

$$f^{(n)}(\boldsymbol{\theta}^{(n)}) \equiv f_{1,2....,n:n}(\theta_1, ..., \theta_n) = n! \prod_{i=1}^{n} f(\theta_i) \quad \bar{\theta} \geq \theta_1 \geq \theta_2 \geq ... \geq \theta_n \geq \underline{\theta}$$

Note that the efficient allocation must enable the highest type agent to win and pay, then the expected payment is,

$$\begin{aligned} & m_i(\theta) \\ = \; & \Pr(\theta > \max_{j \neq i} \theta_j) \mathbb{E}_{\theta_{-i}}[M(\theta, \theta_{-i}) \backslash \theta > \max_{j \neq 1} \theta_j] \\ & -\frac{\Pr(\theta < \max_{j \neq i} \theta_j) \mathbb{E}_{\theta_{-i}} \left[\mathbb{E}_{\theta_{-ij}} \sum_{j=0}^{n-2} M(\theta_1, ..., \theta, ..., \theta_n) \backslash \theta = \theta_{n-j-2:n-2}) \backslash \theta < \max_{j \neq i} \theta_j \right]}{n-1} \\ = \; & \int_{\underline{\theta}}^{\theta} \int_{\underline{\theta}}^{\theta_2} ... \int_{\underline{\theta}}^{\theta_{n-1}} M(\theta, \theta_{-i}) f^{(n-1)}(\boldsymbol{\theta}^{(n-1)}) d\theta_{-1} - \frac{1}{n-1} R(\theta) \end{aligned}$$

where

$$R(\theta) = \int_{\theta}^{\bar{\theta}} \{ \sum_{j=1}^{n-1} \underbrace{\int_{\theta}^{\theta_1} ... \int_{\theta}^{\theta_{j-1}}}_{j-1} \underbrace{\int_{\underline{\theta}}^{\theta} \{ \int_{\underline{\theta}}^{\theta_{j+2}} \int_{\underline{\theta}}^{\theta_{j+3}} ... \int_{\underline{\theta}}^{\theta_{n-1}}}_{n-j-1} M(\theta_1, .., \theta_j, \theta, \theta_{j+2}, ., \theta_n) f^{(n-1)} ... \} d\theta_2 \} d(\theta_1)$$

Note that $M(\boldsymbol{\theta})$ is non-decreasing in each argument, we have

$$\begin{aligned}
R(\theta) &\leq \int_{\theta}^{\bar{\theta}} \{\sum_{j=1}^{n-1} \frac{(n-1)!F(\theta)^{n-j-1}}{(n-j-1)!} \underbrace{\int_{\theta}^{\theta_1} \ldots \int_{\theta}^{\theta_{j-1}}}_{j-1} M(\theta_1, ., \theta_j, \theta, \theta, ., \theta) f(\theta_{j-1}) \ldots\} d\theta_2\} d\theta_1 \\
&= \int_{\theta}^{\bar{\theta}} \{\sum_{j=1}^{n-1} \frac{(n-1)!F(\theta)^{n-j-1}(F(\theta_1)-F(\theta))^{j-1}}{(j-1)!(n-j-1)!} M(\underbrace{\theta_1, ., \theta_1}_{j}, \theta, \theta, ., \theta) f(\theta_1) d\theta_1 \\
&\leq (n-1) \int_{\theta}^{\bar{\theta}} F(\theta_1)^{n-2} M(\underbrace{\theta_1, ., \theta_1, \theta_1}_{n-2}, \theta, \theta) f(\theta_1) d\theta_1
\end{aligned}$$

Therefore,

$$m_i(\theta) \leq \int_{\underline{\theta}}^{\theta} \int_{\underline{\theta}}^{\theta_2} \ldots \int_{\underline{\theta}}^{\theta_{n-1}} M(\theta, \theta_{-1}) f^{(n-1)}(\boldsymbol{\theta}^{(n-1)}) d\theta_{-1} - \int_{\theta}^{\bar{\theta}} F(\theta_1)^{n-2} M(\underbrace{\theta_1, ., \theta_1, \theta_1}_{n-2}, \theta, \theta) dF(\theta_1) \tag{3.12}$$

Note that when $\theta \rightarrow \bar{\theta}$,

$$m_i(\bar{\theta}) = \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta_2} \ldots \int_{\underline{\theta}}^{\theta_{n-1}} M(\bar{\theta}, \theta_{-i}) f^{(n-1)}(\boldsymbol{\theta}^{(n-1)}) d\theta_2 \ldots d\theta_n$$

Since inequality (3.12) should hold for any θ, and $\bar{\theta}$, it must be the case,

$$\begin{aligned}
m_i'(\theta) \leq\ & f(\theta)(n-1) \int_{\underline{\theta}}^{\theta} \int_{\underline{\theta}}^{\theta_3} \ldots \int_{\underline{\theta}}^{\theta_{n-1}} M(\theta, \theta, \theta_{-12}) f^{(n-2)}(\boldsymbol{\theta}^{(n-2)}) d\theta_{-12} \\
&+ \int_{\underline{\theta}}^{\theta} \int_{\underline{\theta}}^{\theta_2} \ldots \int_{\underline{\theta}}^{\theta_{n-1}} \frac{\partial}{\partial \theta}[M(\theta, \theta_{-1})] f^{(n-1)}(\boldsymbol{\theta}^{(n-1)}) d\theta_{-1} \\
&+ F(\theta)^{n-2} f(\theta) M(\theta, \ldots, \theta, \theta) - \int_{\theta}^{\bar{\theta}} F(\theta_1)^{n-2} \{\frac{\partial}{\partial \theta}[M(\underbrace{\theta_1, ., \theta_1, \theta_1}_{n-2}, \theta, \theta)]\} dF(\theta_1)
\end{aligned}$$

where $\frac{\partial}{\partial \theta}[M(\theta, \theta_{-1})]$ is piecewise if $M(\underbrace{\theta_1, ., \theta_1, \theta_1}_{n-2}, \theta, \theta)$ is not differentiable in θ (but $M(\underbrace{\theta_1, ., \theta_1, \theta_1}_{n-2}, \theta, \theta)$ is still weakly differentiable).

Note that $\int_{\theta}^{\bar{\theta}} F(\theta_1)^{n-2}\{\frac{\partial}{\partial\theta}[M(\underbrace{\theta_1,.,\theta_1,\theta_1}_{n-2},\theta,\theta)]\}dF(\theta_1) \geq 0$, therefore, we have

$$\begin{aligned} S_i(\theta) \leq\ & \frac{\int_{\underline{\theta}}^{\theta}\int_{\underline{\theta}}^{\theta_3}\cdots\int_{\underline{\theta}}^{\theta_{n-1}} M(\theta,\theta,\theta_{-12})f^{(n-2)}(\boldsymbol{\theta}^{(n-2)})d\theta_{-12}}{F(\theta)^{n-2}} \\ & +\frac{1}{n-1}M(\theta,...,\theta,\theta) \\ & +\frac{F(\theta)}{f(\theta)}\frac{\int_{\underline{\theta}}^{\theta}\int_{\underline{\theta}}^{\theta_2}\cdots\int_{\underline{\theta}}^{\theta_{n-1}} \frac{\partial}{\partial\theta}\delta[M(\theta,\theta_{-1})]f^{(n-1)}(\boldsymbol{\theta}^{(n-1)})d\theta_{-1}}{F(\theta)^{n-1}}. \end{aligned}$$

When $\theta \to \underline{\theta}$, we have $\frac{n-1}{n}S_i(\underline{\theta}) \leq M(\underline{\boldsymbol{\theta}})$. If $\frac{n-1}{n}S_i(\underline{\theta}) < M(\underline{\boldsymbol{\theta}})$, payment rule $M(\boldsymbol{\theta})$ will not lead to ex post monotonicity. So it has to be $M(\underline{\boldsymbol{\theta}}) = \frac{n-1}{n}S_i(\underline{\theta})$. Because

$$\lim_{\theta\to\underline{\theta}}\frac{F(\theta)}{f(\theta)}\frac{\int_{\underline{\theta}}^{\theta}\int_{\underline{\theta}}^{\theta_2}\cdots\int_{\underline{\theta}}^{\theta_{n-1}} \frac{\partial}{\partial\theta}\delta[M(\theta,\theta_{-1})]f^{(n-1)}(\boldsymbol{\theta}^{(n-1)})d\theta_{-1}}{F(\theta)^{n-1}} = 0,$$

and the above inequality should hold for any θ and $\underline{\theta}$ uniformly, therefore,

$$S_i(\theta) \leq \frac{\int_{\underline{\theta}}^{\theta}\int_{\underline{\theta}}^{\theta_3}\cdots\int_{\underline{\theta}}^{\theta_{n-1}} M(\theta,\theta,\theta_{-12})f^{(n-2)}(\boldsymbol{\theta}^{(n-2)})d\theta_{-12}}{F(\theta)^{n-2}} + \frac{1}{n-1}M(\theta,...,\theta,\theta)$$

should be true for all θ.

If $n \geq 2$, and $M(\theta,\theta,\theta_{-12})$ depends on θ_j for $j \geq 3$, we have

$$S_i(\theta) < \frac{n}{n-1}M(\theta,...,\theta,\theta),$$

which fails ex post payoff monotonicity.

If $n = 2$, or $M(\boldsymbol{\theta}) = M(\theta_1,\theta_2)$, we have $S_i(\theta) \leq \frac{n}{n-1}M(\theta,\theta)$. Because $M(\theta,\theta) = \frac{n-1}{n}S_i(\theta)$ can not be incentive compatible, there must be $M(\theta,\theta) > \frac{n-1}{n}S_i(\theta)$ for some measurable set of θ. To see this, note that if $M(\theta,\theta) = \frac{n-1}{n}S_i(\theta)$ is incentive compatible, we will have $\frac{\partial}{\partial\theta}M(\theta,\theta) = 0$ for all θ, a contradiction. Therefore, $M(\boldsymbol{\theta})$ must depends on $\theta^{n:n}$ only.

Step 2: derivation of payment rule as a function of $\theta^{n:n}$ only.

If payment depends on $\theta^{n:n}$, we have the following integration eqution:

$$M(\theta)G(\theta) - \frac{1}{n-1}\int_{\theta}^{\bar{\theta}} M(\tau)dG(\tau) = m(\theta)$$

Note that $\int_{\theta}^{\bar{\theta}} M(\tau)dG(\tau)$ should be differentiable even if $M(\theta)$ is not differentiable. We thus can solve the above integration equation, yielding,

$$M(\theta) = \frac{n-1}{n}\frac{\int_{\underline{\theta}}^{\theta} S(\tau)dF(\tau)^n}{F(\theta)^n}.$$

which is consistent with M^F. Q.E.D. ■

3.7.4 A4. Proof of Proposition 4

Proof. (i) The incentive compatibility is met by construction, and so is budget balance. We only need to check ex post IR constraint. First, we need to prove the payment rule is non-negative. Observing that from the equation (6), we have

$$\begin{aligned}\sum M_i' &= -\left(\sum M_i\right)\left(\sum q_k\right) - \frac{n}{n-1}\sum q_k M_k + \frac{n}{n-1}\sum q_i M_i + S(\theta_i)(n-1)\sum q_k \\ &= -\left(\sum M_i\right)\left(\sum q_k\right) + S(\theta_i)(n-1)\sum q_k\end{aligned}$$

thus,

$$\sum M_i(\theta) = (n-1)\frac{\int_{\underline{\theta}}^{\theta} S(\tau)d\Pi(\tau)}{\Pi(\theta)} \tag{3.13}$$

Using this formula, we can show that if $M_i(\theta) \leq \min_{j\neq} M_j(\theta)$ for any θ, then $M_i'(\theta) > 0$. To see this, note that for any θ, if $M_i \leq \min_{j\neq} M_j$,

$$M_k = \sum M_j - \sum_{j\neq k} M_j \leq \sum M_j - \sum_{j\neq i} M_j$$

therefore

$$q_k M_k \leq q_k(\sum M_j - \sum_{j\neq i} M_j)$$

Thus,

$$
\begin{aligned}
M_i' &= -M_i \sum_{k \neq i} q_k - \frac{1}{n-1} \sum_{k \neq i} q_k M_k + S(\theta_i) \sum_{k \neq i} q_k \\
&\geq -M_i \sum_{k \neq i} q_k - \frac{1}{n-1} (\sum M_j - \sum_{j \neq i} M_j) \sum_{k \neq i} q_k + S(\theta_i) \sum_{k \neq i} q_k \\
&\geq \left(\sum_{k \neq i} q_k \right) \left(S(\theta_i) - \frac{\int_{\underline{\theta}}^{\theta} S(\tau) d\Pi(\tau)}{\Pi(\theta)} \right) \\
&> 0
\end{aligned}
$$

By the above inequality, payment increases with type. Meanwhile, note that $M_i(\underline{\theta}) = \frac{1}{n} S(\underline{\theta}) \geq 0$, therefore, if $M_i(\theta)$ is the lowest payment, then $M_i(\theta) > 0$; if $M_i(\theta)$ is not lowest, of course $M_i(\theta) > 0$. Therefore, the loser's pay-off is non-negative.

Now it is easy to show that the winner's ex post pay-off is also non-negative, due to

$$
S(\theta_i) - M_i(\theta_i) \geq M_j(\theta_i) \geq \frac{1}{n-1} M_j(\theta_i)
$$

since M_i is non-negative.

(ii) Note that,

$$
\begin{aligned}
\sum U_i(\underline{\theta}) &= -\sum m_i(\underline{\theta}) \\
&= \frac{1}{n-1} \sum_{j=1} \sum_{j \neq i} \int_{\underline{\theta}}^{\bar{\theta}} M_j(z) \left[\Pi_{k \neq j,i} F_k(z_j) \right] dF_j(z) \\
&= \sum_{i=1} \mathbb{E}_{\theta_{-i}} S(\underline{\theta}_i, \theta_{-i}) - \sum_{i=1} \frac{1}{n-1} \mathbb{E}_{\theta_{-i}} \sum_{j \neq i} \mathbb{E}_{\theta_{-j}} [S_{-j}(\theta_j, \theta_{-j})] \\
&= \mathbb{E}\Delta
\end{aligned}
$$

where the third step comes from plugging in equation (5). We get the conclusion. Q.E.D. ■

3.7.5 A5. Proof of Proposition 5

Proof. Note that for any incentive compatible mechanism, the first order condition

$$
m_i'(\theta) = f_j(\theta) S_i(\theta)
$$

is always the case. Thus, we obtain two equations:

$$S(\theta_i)f_j(\theta_i) = f_jM_i + F_jM_i' + M_jf_j$$

$$S(\theta_j)f_i(\theta_j) = f_iM_j + F_iM_j' + M_if_i$$

Therefore, we have

$$0 = f_iF_jM_i' - f_jF_iM_j'$$

Taking derivative w.r.t. θ_i, we have

$$M_j' = S'(\theta_i) - M_i' - (\frac{F_j}{f_j})'M_i' - (\frac{F_j}{f_j})M_i''$$

Therefore we obtain a second order ODE regarding a single unknown:

$$\frac{f_j}{F_j}S'(\theta_i) = M_i'\left(\frac{f_j}{F_j} + \frac{f_j}{F_j}(\frac{F_j}{f_j})' + \frac{f_i}{F_i}\right) + M_i''$$

The general solution is,

$$M_i' = \frac{f_j}{F_iF_j^2}C_1 + \frac{f_j\int_{\underline{\theta}}^{\theta}F_iF_jS'd\theta}{F_iF_j^2}$$

hence,

$$M_i = M_i(\underline{\theta}_i) + C_1\int_{\underline{\theta}}^{\theta_i}\frac{f_j}{F_iF_j^2}d\theta + \int_{\underline{\theta}}^{\theta_i}\frac{f_j\int_{\underline{\theta}}^{\theta}F_iF_jS'd\tau}{F_iF_j^2}d\theta$$

Note that in any case M_i should be non-negative and not be infinite, so $C_1 = 0$, otherwise,

$$\int_{\underline{\theta}}^{\theta_i}\frac{f_j}{F_iF_j^2}d\theta \geq \int_{\underline{\theta}}^{\theta_i}\frac{f_j}{F_j^2}d\theta = \frac{1}{F_j(\underline{\theta}_i)} - \frac{1}{F_j(\theta_i)} \to \infty.$$

It can be shown that $\int_{\underline{\theta}}^{\theta_i}\frac{f_j\int_{\underline{\theta}}^{\theta}F_iF_jS'd\tau}{F_iF_j^2}d\theta$ is bounded. Even for $\theta \to \underline{\theta}$, according to L'Hospital Law,

$$\lim_{\theta\to\underline{\theta}}\frac{1}{F_j}\left(\frac{\int_{\underline{\theta}}^{\theta}F_iF_jS'd\tau}{F_iF_j}\right) = \frac{1}{2}\lim_{\theta\to\underline{\theta}}\frac{1}{f_j}S'(\underline{\theta})$$

Therefore, we have

$$\begin{aligned} M_i &= M_i(\underline{\theta}_i) + \int_{\underline{\theta}}^{\theta_i} \frac{f_j \int_{\underline{\theta}}^{\theta} F_i F_j S' d\tau}{F_i F_j^2} d\theta \\ &= M_i(\underline{\theta}_i) + \frac{\int_{\underline{\theta}}^{\theta_i} S d(F_i F_j)}{F_i F_j} - \int_{\underline{\theta}}^{\theta_i} \frac{1}{F_j} \frac{\int_{\underline{\theta}}^{\theta} F_i F_j S' d\tau}{F_i^2} dF_i \end{aligned}$$

To check the incentive compatibility, we have

$$\begin{aligned} & F_j M_i - \int_{\theta_i}^{\bar{\theta}} M_j dF_j \\ =\ & F_j \left(M_i(\underline{\theta}_i) + \int_{\underline{\theta}}^{\theta_i} \frac{f_j \int_{\underline{\theta}}^{\theta} F_i F_j S' d\theta}{F_i F_j^2} d\theta \right) - \int_{\theta_i}^{\bar{\theta}} \left(M_j(\underline{\theta}_j) + \int_{\underline{\theta}}^{\theta_j} \frac{f_i \int_{\underline{\theta}}^{\theta} F_i F_j S' d\theta}{F_j F_i^2} d\theta \right) dF_j \\ =\ & M(\underline{\theta})(2F_j - 1) - F_j(\theta_i) \left(\frac{1}{F_i F_j} \int_{\underline{\theta}}^{\theta} F_i F_j S' d\tau \Big|_{\underline{\theta}}^{\theta_i} - \int_{\underline{\theta}}^{\theta_i} S' d\theta \right) \\ & - \int_{\underline{\theta}}^{\bar{\theta}} \frac{f_i \int_{\underline{\theta}}^{\theta} F_i F_j S' d\tau}{F_j F_i^2} d\theta - \int_{\theta_i}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} F_i F_j S' d\tau d\frac{1}{F_i} \\ =\ & M(\underline{\theta})(2F_j - 1) - \int_{\underline{\theta}}^{\bar{\theta}} \frac{f_i \int_{\underline{\theta}}^{\theta} F_i F_j S' d\tau}{F_j F_i^2} d\theta - \int_{\underline{\theta}}^{\bar{\theta}} F_i F_j S' d\tau + F_j(\theta_i) \int_{\underline{\theta}}^{\theta_i} S' d\theta + \int_{\theta_i}^{\bar{\theta}} F_j S' d\theta \\ =\ & m_i(\underline{\theta}) + \int_{\underline{\theta}}^{\theta_i} S(\theta) dF_j(\theta) \end{aligned}$$

where

$$m(\underline{\theta}) = M(\underline{\theta})(2F_j - 1) - \int_{\underline{\theta}}^{\bar{\theta}} \frac{f_i \int_{\underline{\theta}}^{\theta} F_i F_j S' d\tau}{F_j F_i^2} d\theta - \int_{\underline{\theta}}^{\bar{\theta}} F_i F_j S' d\tau - \int_{\underline{\theta}}^{\bar{\theta}} S(\theta) dF_j(\theta) + S(\bar{\theta}) - F_j(\theta_i) S(\underline{\theta})$$

We set $M(\underline{\theta}) = \frac{1}{2} S(\underline{\theta})$, which means that for the lowest type, it is indifferent for him to lose or win. Then,

$$m_i(\underline{\theta}) = \int_{\underline{\theta}}^{\bar{\theta}} S(\theta) d(F_i F_j) - \int_{\underline{\theta}}^{\bar{\theta}} S(\theta) dF_j(\theta) - \int_{\underline{\theta}}^{\bar{\theta}} \frac{f_i \int_{\underline{\theta}}^{\theta} F_i F_j S' d\tau}{F_j F_i^2} d\theta - \frac{1}{2} S(\underline{\theta})$$

which is a constant independent of θ.

We can verify that under this payment rule, truth-telling is an equilibrium. For an agent i

to deviate from θ_i, the resulting profit difference will be:

$$\begin{aligned} U_i(\theta_i,\theta_i) - U_i(\theta_i,\tilde{\theta}_i) &= [\Pr(\theta_j \le \theta_i) - \Pr(\theta_j \le \tilde{\theta}_i)]S(\theta_i) - [m_i(\theta_i) - m_i(\theta_i)] \\ &= \int_{\tilde{\theta}_i}^{\theta_i} [S(\theta_i) f_j(\tau) - m_i'(\tau)] d\tau \end{aligned}$$

When $\theta_i > \tilde{\theta}_i$, $S(\theta_i)f_j(\tau) - m_i'(\tau) > S(\tau)f_j(\tau) - m_i'(\tau) = 0$; when $\theta_i < \tilde{\theta}_i$, $S(\theta_i)f_j(\tau) - m_i'(\tau) < S(\tau)f_j(\tau) - m_i'(\tau) = 0$, therefore, in any case, $U_i(\theta_i,\theta_i) - U_i(\theta_i,\tilde{\theta}_i) > 0$ for any $\theta_i \ne \tilde{\theta}_i$.

For the ex post monotonicity, it is easy to see that the above payment rule is monotone in θ and non-negative. We only need to verify

$$S(\theta_i) - M_i(\theta_i) \ge M_j(\theta_j) \text{ for } \theta_i \ge \theta_j.$$

by noting that

$$\begin{aligned} & M_i(\theta) + M_j(\theta) \\ = \; & S(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \frac{f_j \int_{\underline{\theta}}^{z} F_i F_j S' d\tau}{F_i F_j^2} d\theta + \int_{\underline{\theta}}^{\theta} \frac{f_i \int_{\underline{\theta}}^{\theta} F_j F_i S' d\tau}{F_j F_i^2} d\theta \\ = \; & S(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} \left(\int_{\underline{\theta}}^{z} F_i F_j S' d\tau \right) d\frac{1}{F_i F_j} \\ = \; & S(\theta) - \frac{\int_{\underline{\theta}}^{\theta} F_i F_j S' d\tau}{F_i F_j} \\ = \; & \frac{\int_{\underline{\theta}}^{\theta} S d(F_i F_j)}{F_i F_j} \le S(\theta) \end{aligned}$$

Q.E.D. ■

3.7.6 A6. Proof of Theorem 4

Proof. (i) It is obvious to see that **M2** is ex post budget balance by construction. The incentive compatibility could be checked as follows. Note that the structure of expected payment $m(\theta)$

can be written as:

$$
\begin{aligned}
m(\theta) &= (1-k)G(\theta)\beta^k(\theta) + k\int_{\underline{\theta}}^{\theta}\beta^k(\tau)dG(\tau) - \int_{\theta_{-i}\leq\theta} S_{-i}(\theta,\theta_{-i})d\mathbf{F}_{-i}(\theta_{-i}) \\
&\quad -\frac{1}{n-1}\left(k\begin{bmatrix} (n-1)F(\theta)^{n-2}(1-F(\theta))\beta^k(\theta) \\ +(n-1)\int_{\theta}^{\bar{\theta}}(1-F(\tau))\beta^k(\tau)dF(\tau)^{n-2}\end{bmatrix} + (1-k)\int_{\theta}^{\bar{\theta}}\beta^k(\tau)dG(\tau)\right) \\
&\quad +\sum_{j=2}^{n}\Pr(\theta \text{ is } j\text{-th highest order statistic})\mathbb{E}[v(x^*(\theta,\theta_{-i}),\theta)/\theta \text{ is } j\text{-th highest order statistic}] \\
&= (1-k)G(\theta)\beta^k(\theta) + k\int_{\underline{\theta}}^{\theta}\beta^k(\tau)dG(\tau) - \int_{\theta_{-i}\leq\theta} S_{-i}(\theta,\theta_{-i})d\mathbf{F}_{-i}(\theta_{-i}) \\
&\quad -k\left[F(\theta)^{n-2}(1-F(\theta))\beta^k(\theta) + \int_{\theta}^{\bar{\theta}}(1-F(\tau))\beta^k(\tau)dF(\tau)^{n-2}\right] - \frac{(1-k)}{n-1}\int_{\theta}^{\bar{\theta}}\beta^k(\tau)dG(\tau) \\
&\quad +\sum_{j=1}^{n}\bar{v}_i^{(j)}(x_i^*(..,\theta,.),\theta) - \bar{v}_i^{(1)}(x_i^*(..,\theta,.),\theta)
\end{aligned}
$$

(Here $v_i(x_i^*(..\theta_{j-1},\theta,.\theta_{j+1},.),\theta)$ means individual i's type is θ, which is j-th highest order statistic among n). Taking derivative w.r.t θ and simplifying the above equation, we obtain a differential equation regarding $\beta^k(\theta)$:

$$
\begin{aligned}
& m'(\theta) + \frac{d}{d\theta}\left(\int_{\theta_{-i}\leq\theta} S_{-i}(\theta,\theta_{-i})f^{(n-1)}(\theta_{-i})d\theta_{-i} - \sum_{j=1}^{n}\bar{v}_i^{(j)}(x_i^*(..,\theta,.),\theta)\right) \\
&= (1-k)G(\theta)\beta'(\theta) - kF(\theta)^{n-2}(1-F(\theta))\beta^{k\prime}(\theta) + (1-k)g(\theta)\beta^k(\theta + kg(\theta)\beta^k(\theta) \\
&\quad -kF(\theta)^{n-3}((n-2)-(n-1)F(\theta))\beta^k(\theta)f(\theta) \\
&\quad +k(n-2)(1-F(\theta))\beta^k(\theta)F(\theta)^{n-3}f(\theta) + \frac{(1-k)}{n-1}g(\theta)\beta^k(\theta) \\
&= \beta^{k\prime}(\theta)(F(\theta)-k)F(\theta)^{n-2} + nF(\theta)^{n-2}f(\theta)\beta^k(\theta)
\end{aligned}
$$

$$
\begin{aligned}
&= (1-k)G(\theta)\beta^k(\theta) + k\int_{\underline{\theta}}^{\theta}\beta(\tau)dG(\tau) - \int_{\theta_{-i}\leq\theta} S_{-i}(\theta,\theta_{-i})f^{(n-1)}(\theta_{-i})d\theta_{-i} \\
&\quad -k\left[F(\theta)^{n-2}(1-F(\theta))\beta^k(\theta) + \int_{\theta}^{\bar{\theta}}(1-F(\tau))\beta^k(\tau)dF(\tau)^{n-2}\right] - \frac{(1-k)}{n-1}\int_{\theta}^{\bar{\theta}}\beta^k(\tau)dG(\tau)
\end{aligned}
$$

Plugging $m'(\theta)$ into the above equation, the ODE becomes

$$
\begin{aligned}
& \beta^{k\prime}(\theta)(F(\theta)-k)F(\theta)^{n-2} + nF(\theta)^{n-2}f(\theta)\beta^k(\theta) \\
= & \frac{d}{d\theta}\left(\int_{\theta_{-i}\leq\theta} S(\theta,\theta_{-i})d\mathbf{F}_{-i}(\theta_{-i})\right) - \frac{d}{d\theta}\sum_{j=1}^{n}\bar{v}_i^{(j)}(x_i^*(..,\theta,.),\theta) \\
& + \left[\frac{\partial}{\partial z_i}\int_{\theta_{-i}} v_i(x_i^*(z_i,\theta_{-i}),\theta)d\mathbf{F}_{-i}(\theta_{-i})\right]_{z_i=\theta}
\end{aligned}
$$

The solution turns out to be

$$
\beta^k(\theta) = \frac{\int_{F^{-1}(k)}^{\theta}\left(\frac{d}{d\tau}\bar{S}^{(1)}(\tau) - \sum_{j=1}^{n}\frac{d}{d\tau}\bar{v}_i^{(j)}(x_i^*(\tau,.),\tau) + m'(\tau)\right)\frac{(F(\tau)-k)^{n-1}}{F(\tau)^{n-2}}d\tau}{(F(\theta)-k)^n}
$$

which means **M2** is incentive compatible. To check the second order condition, just apply the proof of lemma 3.

$$
\begin{aligned}
& U(\theta,\theta) - U(\theta,\tilde{\theta}) \\
= & \int_{\theta_{-i}} [v(x(\theta,\theta_{-i}),\theta) - v(x(\tilde{\theta},\theta_{-i}),\theta)]dF_{-i}(\theta_{-i}) - (m(\theta) - m(\tilde{\theta})) \\
= & \int_{\tilde{\theta}}^{\theta}\left(\frac{\partial}{\partial z}\int_{\theta_{-i}} v(x(z,\theta_{-i}),\theta)d\mathbf{F}_{-i}(\boldsymbol{\theta}_{-i})\right)dz - \int_{\tilde{\theta}}^{\theta} m'(z)dz
\end{aligned}
$$

From the IC constraint,

$$
m'(z) = \left[\frac{\partial}{\partial \tau}\int_{\theta_{-i}}\frac{\partial}{\partial x}v(x(\tau,\theta_{-i}),z)d\mathbf{F}_{-i}(\boldsymbol{\theta}_{-i})\right]_{\tau=z}
$$

Note that $x(z,\theta_{-i})$ is increasing function of z, and $\frac{\partial}{\partial x}v(x(z,\theta_{-i}),\theta)$ is increasing function of θ, therefore,

$$
\begin{aligned}
& U(\theta,\theta) - U(\theta,\tilde{\theta}) \\
= & \int_{\tilde{\theta}}^{\theta}\left[\frac{\partial}{\partial z}\int_{\theta_{-i}} v(x(z,\theta_{-i}),\theta)d\mathbf{F}_{-i}(\boldsymbol{\theta}_{-i}) - \left[\frac{\partial}{\partial \tau}\int_{\theta_{-i}}\frac{\partial}{\partial x}v(x(\tau,\theta_{-i}),z)d\mathbf{F}_{-i}(\boldsymbol{\theta}_{-i})\right]_{\tau=z}\right]d\mathbf{F}_{-i}(\theta_{-i})dz \\
\geq & \ 0
\end{aligned}
$$

(ii) To show this, it is convenient to apply the revenue equivalence principle. With assistance of $\beta^0(\theta)$, we know

$$m(\underline{\theta}) = -\frac{1}{n-1}\int_{\underline{\theta}}^{\bar{\theta}} \beta^k(\theta)dG(\theta) + \int_{\theta_{-i}} v(x^*(\underline{\theta}, \theta_{-i}), \underline{\theta})d\mathbf{F}_{-i}(\theta_{-i})$$

And the lowest type agent's payoff:

$$U(\underline{\theta}) = \mathbb{E}v(x^*(\underline{\theta}, \theta_{-i}), \underline{\theta})] - m(\underline{\theta}) = \frac{1}{n-1}\int_{\underline{\theta}}^{\bar{\theta}} \beta^0(\theta)dG(\theta)$$

Plugging the above equality into the payment function:

$$\begin{aligned}
& \frac{1}{n-1}\int_{\underline{\theta}}^{\bar{\theta}} \frac{\int_{\underline{\theta}}^{\theta} F(\tau)d\int_{\theta_{-i}\leq\tau} S(\tau, \theta_{-i})d\mathbf{F}_{-i}(\theta_{-i})}{F(\theta)^n} dG(\theta) \\
= & -\int_{\underline{\theta}}^{\bar{\theta}} \left(\int_{\underline{\theta}}^{\theta} F(\tau)d\int_{\theta_{-i}\leq\tau} S(\tau, \theta_{-i})d\mathbf{F}_{-i}(\theta_{-i})\right) d\frac{1}{F(\theta)} \\
= & \int_{\underline{\theta}}^{\bar{\theta}}\int_{\theta_{-i}\leq\theta} S(\theta, \theta_{-i})d\mathbf{F}_{-i}(\theta_{-i})dF(\theta)
\end{aligned}$$

The third line of the above simplification comes from integrating by parts. Similarly,

$$\begin{aligned}
& -\frac{1}{n-1}\int_{\underline{\theta}}^{\bar{\theta}} \frac{\int_{\theta_{-i}}\left(\int_{\underline{\theta}}^{\theta} F(\tau)\frac{\partial}{\partial\tau}v(x^*(\tau, \theta_{-i}), \tau)d\tau\right) d\mathbf{F}_{-i}(\theta_{-i})}{F(\theta)^n} dG(\theta) \\
= & \int_{\theta_{-i}}\left[\int_{\underline{\theta}}^{\bar{\theta}}\left(\int_{\underline{\theta}}^{\theta} F(\tau)\frac{\partial}{\partial\tau}v(x^*(\tau, \theta_{-i}), \tau)d\tau\right) d\frac{1}{F(\theta)}\right] d\mathbf{F}_{-i}(\theta_{-i}) \\
= & -\int_{\theta_{-i}}\int_{\underline{\theta}}^{\bar{\theta}}\left(\frac{1-F(\theta)}{f(\theta)}\frac{\partial}{\partial\theta}v(x^*(\theta, \theta_{-i}), \theta)dF(\theta)\right) d\mathbf{F}_{-i}(\theta_{-i})
\end{aligned}$$

Therefore, the lowest type agent's utility turns out to be:

$$\begin{aligned}
U(\underline{\theta}) &= \frac{1}{n-1}\int_{\underline{\theta}}^{\bar{\theta}} \beta^0(\theta)dG(\theta) \\
&= \int_{\underline{\theta}}^{\bar{\theta}}\int_{\theta_{-i}\leq\theta} S(\theta, \theta_{-i})d\mathbf{F}_{-i}(\theta_{-i})dF(\theta) - \int_{\theta_{-i}}\int_{\underline{\theta}}^{\bar{\theta}}\left(\frac{1-F(\theta)}{f(\theta)}\frac{\partial}{\partial\theta}v(x^*(\theta, \theta_{-i}), \theta)dF(\theta)\right) d\mathbf{F}_{-i}(\theta_{-i})
\end{aligned}$$

Note that[15]

$$\frac{1}{n}\int_{\underline{\theta}}^{\bar{\theta}}\int_{\theta_{-i}} S(\theta,\theta_{-i})d\mathbf{F}_{-i}(\theta_{-i})dF(\theta) = \int_{\underline{\theta}}^{\bar{\theta}}\int_{\theta_{-i}\le\theta} S(\theta,\theta_{-i})d\mathbf{F}_{-i}(\theta_{-i})dF(\theta)$$

So $U(\underline{\theta}) = \frac{1}{n}\mathbb{E}\Delta$. Q.E.D. ■

3.7.7 A7. Derivation of M3.

Let $\Psi(\tau) = 1-(1-F(\tau))^{n-1}$ be the distribution of $Z_i = \min_{i\neq j}\theta_j$, the expected payment can be written as

$$\begin{aligned}
m(\theta) &= (1-k)(1-\Psi(\theta))r^k(\theta) + k\int_{\theta}^{\bar{\theta}} r^k(\tau)d\Psi(\tau) - \int_{\theta_{-i}\ge\theta} S_{-i}(\theta,\theta_{-i})d\mathbf{F}_{-i}(\theta_{-i}) \\
&\quad -\frac{1}{n-1}\left(k\left[\begin{array}{c}(n-1)F(\theta)(1-F(\theta))^{n-2}r^k(\theta) \\ +(n-1)(n-2)\int_{\underline{\theta}}^{\theta} r^k(\tau)F(\tau)(1-F(\tau))^{n-3}dF(\tau)\end{array}\right] + (1-k)\int_{\underline{\theta}}^{\theta} r(\tau)d\Psi(\tau)\right) \\
&\quad +\sum_{j=2}^{n}\Pr(\theta \text{ is } j\text{-th smallest order statistic})\mathbb{E}[v(x^*(\boldsymbol{\theta}),\theta)/\theta \text{ is } j\text{-th smallest order statistic}] \\
&= (1-k)(1-\Psi(\theta))r^k(\theta) + k\int_{\theta}^{\bar{\theta}} r^k(\tau)d\Psi(\tau) - \int_{\theta_{-i}\ge\theta} S_{-i}(\theta,\theta_{-i})d\mathbf{F}_{-i}(\theta_{-i}) \\
&\quad -k\left[F(\theta)(1-F(\theta))^{n-2}r^k(\theta) + (n-2)\int_{\underline{\theta}}^{\theta} r^k(\tau)F(\tau)(1-F(\tau))^{n-3}dF(\tau)\right] \\
&\quad -\frac{(1-k)}{n-1}\int_{\underline{\theta}}^{\theta} r(\tau)d\Psi(\tau) + \int_{\theta_{-i}} v(x^*(\theta,\theta_{-i}),\theta)d\mathbf{F}_{-i}(\theta_{-i}) - \int_{\theta_{-i}\ge\theta} v(x^*(\theta,\theta_{-i}),\theta)d\mathbf{F}_{-i}(\theta_{-i})
\end{aligned}$$

---

[15]Note that the following formula is true by symmetricity of $S(\theta,\theta_{-i})$:

$$\begin{aligned}
&\int_{\underline{\theta}}^{\bar{\theta}}\int_{\theta_{-i}} S(\theta,\theta_{-i})d\mathbf{F}_{-i}(\theta_{-i})dF(\theta) \\
=& \sum_{i=1}\Pr(i \text{ is the i-th highest order})\mathbb{E}S(\theta_i,\theta_{-i})/i \text{ is the i-th highest order statistics}] \\
=& \mathbb{E}S(\theta_i,\theta_{-i})/i \text{ is the i-th highest order statistics}] \\
=& \int_{\underline{\theta}}^{\bar{\theta}} \frac{\int_{\theta_{-i}\le\theta} S(\theta,\theta_{-i})d\mathbf{F}_{-i}(\theta_{-i})}{F(\theta)^{n-1}} dF(\theta)^n \\
=& n\int_{\underline{\theta}}^{\bar{\theta}}\int_{\theta_{-i}\le\theta} S(\theta,\theta_{-i})d\mathbf{F}_{-i}(\theta_{-i})dF(\theta)
\end{aligned}$$

The above equality can also come from Fubini's theorem.

Taking derivative w.r.t θ to simplify the above equition, we obtain a differential equation regarding $r^k(\theta)$:

$$
\begin{aligned}
& m'(\theta) + \frac{d}{d\theta}\left(\int_{\theta_{-i}\geq\theta} S(\theta,\theta_{-i})d\mathbf{F}_{-i}(\theta_{-i}) - \sum_{j=1}^{n}\frac{d}{d\tau}\bar{v}_i^{(j)}(x_i^*(\theta,.),\theta)\right) \\
= \; & (1-k)(1-\Psi(\theta))r^{k\prime}(\theta) - (1-k)\Psi'(\theta)r^k(\theta) - kr^k(\theta)\Psi'(\theta) \\
& -k\left[\begin{array}{c} F(\theta)(1-F(\theta))^{n-2}r^{k\prime}(\theta) - (n-2)F(\theta)(1-F(\theta))^{n-3}f(\theta)r^k(\theta) \\ +(1-F(\theta))^{n-2}f(\theta)r^k(\theta) + (n-2)(1-F(\theta))^{n-3}F(\theta)f(\theta)r^k(\theta)] \end{array}\right] \\
& -\frac{(1-k)}{n-1}r^k(\theta)\Psi'(\theta) \\
= \; & r^{k\prime}(\theta)[(1-k)(1-\Psi(\theta)) - kF(\theta)(1-F(\theta))^{n-2}] \\
& -r^k(\theta)[\Psi'(\theta) + k(1-F(\theta))^{n-2}f(\theta) + \frac{(1-k)}{n-1}\Psi'(\theta)] \\
= \; & r^{k\prime}(\theta)[(1-k-F(\theta)](1-F(\theta))^{n-2} - r^k(\theta)n(1-F(\theta))^{n-2}f(\theta)
\end{aligned}
$$

The solution for this ODE is

$$
r^k(\theta) = \frac{\int_{F^{-1}(1-k)}^{\theta}\frac{[k-1+F(\tau)]^{n-1}}{(1-F(\tau))^{n-2}}(\sum_{j=1}^{n}\frac{d}{d\tau}\bar{v}_i^{(j)}(x_i^*(\tau,.),\tau) - m'(\tau) - \frac{d}{d\tau}\bar{S}^{(n)}(\tau))d\tau}{[k-1+F(\theta)]^n}
$$

Thus, if $k=0$,

$$
\begin{aligned}
U(\underline{\theta}) &= \mathbb{E}S(\underline{\theta},\theta_{-i}) - r^0(\underline{\theta}) \\
&= \mathbb{E}S(\underline{\theta},\theta_{-i}) - \int_{\underline{\theta}}^{\bar{\theta}}(1-F(\tau))\left(\int_{\theta_{-i}}\frac{\partial}{\partial\tau}v(x^*(\tau,\theta_{-i}),\tau)d\mathbf{F}_{-i}(\theta_{-i}) - \frac{d}{d\tau}\int_{\theta_{-i}\geq\tau}S(\tau,\theta_{-i})d\mathbf{F}_{-i}(\theta_{-i})\right)d \\
&= \int_{\underline{\theta}}^{\bar{\theta}}\int_{\theta_{-i}\geq\tau}S(\tau,\theta_{-i})d\mathbf{F}_{-i}(\theta_{-i}))dF(\tau) - \int_{\theta_{-i}}\int_{\underline{\theta}}^{\bar{\theta}}\frac{1}{\lambda(\tau)}\frac{\partial}{\partial\tau}v(x^*(\tau,\theta_{-i}),\tau)dF(\tau)d\mathbf{F}_{-i}(\theta_{-i}) \\
&= \mathbb{E}\Delta
\end{aligned}
$$

3.7.8 A8. Proof of Proposition 7

Proof. (i) Let q_i be the initial endowment and $\theta_i^* = F^{-1}(q_i)$. Without loss of generality, suppose $q_j \geq q_i$, therefore $\theta_j^* \geq \theta_i^*$, thus $U_i(\theta_i^*) - \theta_i^* q_i \geq U_j(\theta_j^*) - \theta_j^* q_j$. Ex post IR requires the following conditions:

$$\begin{aligned}
(1-q_i)\theta_i - b(\theta_i) - K_j &\geq 0 \quad (\text{if } \theta_i \geq \theta_j) \\
b(\theta_j) - K_j &\geq \theta_i q_i \quad (\text{if } \theta_i \leq \theta_j) \\
(1-q_j)\theta_j - b(\theta_j) + K_j &\geq 0 \quad (\text{if } \theta_i \leq \theta_j) \\
b(\theta_i) + K_j &\geq \theta_j q_j \quad (\text{if } \theta_i \geq \theta_j) \\
\int_{\underline{\theta}}^{\overline{\theta}} \tau F(\tau) dF(\tau) - \int_{\theta_j^*}^{\overline{\theta}} \tau dF(\tau) &\leq K_j \leq \int_{\theta_i^*}^{\overline{\theta}} \tau dF(\tau) - \int_{\underline{\theta}}^{\overline{\theta}} \tau F(\tau) dF(\tau)
\end{aligned}$$

where K_j is possible transfer from i to j.

From the first two inequalities, the necessary and sufficient condition is that

$$(1-q_i)\theta_i - b(\theta_i) - K_j \geq 0 \ \& \ b(\theta_i) - K_j \geq \theta_i q_i$$

This requires $K_j \leq \frac{1}{2}(1-2q_i)\theta_i$. Looking at the third and fourth inequality, however, it requires $K_j \geq \frac{1}{2}(2q_j-1)\theta_j = \frac{1}{2}(1-2q_i)\theta_j$. It is impossible to have $\frac{1}{2}(1-2q_i)\theta_j \leq K_j \leq \frac{1}{2}(1-2q_i)\theta_i$ for all θ_i and θ_j (as long as the type space is not trivially separating).

(ii) Suppose that i=S owns 1 unit of endowment, like a seller and we let i=B denote the buyer. For any incentive compatible payment rule $M(\theta_i, \theta_j)$, ex post individual rationality requires the following inequalities:

$$\begin{aligned}
S(\theta_B, \theta_S) - M(\theta_B, \theta_S) - K_S &\geq 0 \quad (\text{if } \theta_B \leq \theta_S) \\
M(\theta_B, \theta_S) - K_S &\geq 0 \quad (\text{if } \theta_B \geq \theta_S) \\
S(\theta_S, \theta_B) - M(\theta_S, \theta_B) + K_S &\geq v(1, \theta_S) \quad (\text{if } \theta_B \geq \theta_S) \\
M(\theta_S, \theta_B) + K_S &\geq v(1, \theta_S) \quad (\text{if } \theta_B \leq \theta_S) \\
v(1, \bar{\theta}) - \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\bar{\theta}} \frac{\partial}{\partial \tau} v(x_S^*(\tau, \theta_j), \tau) d\tau dF(\theta_j) &\leq \frac{1}{2}\mathbb{E}\Delta + K_S \\
0 &\leq \frac{1}{2}\mathbb{E}\Delta - K_S
\end{aligned}$$

Looking at the first two inequalities, it is necessary to have

$$K_S \leq \frac{1}{2}S(\underline{\theta}_S, \underline{\theta}_B)$$

since $K_S \leq \min\{S(\theta_B, \theta_B) - M(\theta_B, \theta_S), M(\theta_B, \theta_S)\}$.

Moreover, from the third and fourth inequality, we have

$$\frac{1}{2}S(\underline{\theta}_S, \underline{\theta}_B) \geq K_S \geq v(1, \bar{\theta}_S) - \frac{1}{2}S(\bar{\theta}_S, \bar{\theta}_S)$$

If $v(x, \bar{\theta}_S)$ is linear in x, then we go back to (i). If $v(x, \bar{\theta}_S)$ is strictly concave in x, then $v(1, \bar{\theta}_S) - \frac{1}{2}S(\bar{\theta}_S, \bar{\theta}_S) > 0$, a contradiction with $K_S \leq 0$. Q.E.D. ■

Bibliography

[1] C. d'Aspremont, L.-A. Gerard-Varet, Incentives and incomplete information, J. Public Econ. 11 (1979) 25–45.

[2] L. Ausubel, An efficient ascending-bid auction for multiple objects, American Econ. Rev. 94. (2004) No.5 pp 1452-1475.

[3] D. Bergemann, S. Morris, Robust Mechanism Design, Econometric, 73, No.7, (2005), 1771-1813.

[4] K. Chatterjee,W. Samuelson, Bargaining under incomplete information, Oper. Res. 31 (1983) 835–851.

[5] E., Clarke, Multipart pricing of public goods, Journal Public Choice, 11, (1971), 17-33

[6] P. Cramton, R. Gibbons, P. Klemperer, Dissolving a partnership efficiently, Econometrica 55 (1987) 615–632.

[7] P. Cramton, T. Palfrey, Cartel enforcement with uncertainty about costs, Int. Econ. Rev. 31 (1990) 17–47.

[8] J. Cremer, C. d'Aspremont, L.-A. Gerard-Varet, Correlation, independence, and Bayesian implementation, Soc. Choice Welfare 21 (1999) 281–310.

[9] J. Cremer, R. McLean, Full extraction of the surplus in Bayesian and dominant strategy mechanisms, Econometrica 56 (1988) 1247–1258.

[10] K. Chung, J. Ely, Ex post incentive compatible mechanism design, Working paper, (2003), Department of Economics, Northwestern Unversity.

[11] M. Dudek, T. Kim, J. Ledyard, First best Bayesian privatization mechanisms, Social Science Working Paper #896 California Institute of Technology (1995).

[12] P. Eso, G. Futo, Auction design with a risk averse seller, Economics Letters, 65 (1999), 71-74.

[13] Green J., J.J., Laffont, Characterization of strongly individually incentive compatible mechanisms for the revelation of preferences for public goods, Econometric, (1977), 45, 427-438.

[14] T. Gresik, Incentive efficient equilibria of two-party sealed-bid bargaining games, J. Econ. Theory 68 (1996) 26–48.

[15] T., Groves, Incentives in Teams, Econometric, (1973), 41., 617-631.

[16] M. Hellwig, Public good provision with many participants, Rev. Econ. Stud. 70 (2003) 589–614.

[17] B. Holmstrom, On Incentives and Control in Organization, Part 2, December, (1977), Ph.D. Dissertation, Stanford University.

[18] B. Holmstrom, R. Myerson, Efficient and durable decision rules with incomplete information, Econometrica 51 (1983) 1799–1819.

[19] M. Jackson, H. Moulin, Implementing a public project and distributing its cost, J. Econ. Theory 57 (1992) 125–140.

[20] V. Krishna, M. Perry, Efficient mechanism design, (1998), Working paper.

[21] V. Krishna, Auction Theory, Academic Press, New York, 2002.

[22] J.-J. Laffont, E. Maskin, The theory of incentives: an overview, in: W. Hildenbrand (Ed.), Advances in Economic Theory, Cambridge University Press, Cambridge, 1982, pp. 31–94.

[23] D. Laussel, T. Palfrey, Efficient equilibria in the voluntary contributions mechanism with private information, J. Public Econ. Theory 5 (2003) 449–478.

[24] J. Ledyard, T. Palfrey, Voting and lottery drafts as efficient public goods mechanisms, Rev. Econ. Stud. 61 (1994) 327–355.

[25] J. Ledyard, T. Palfrey, A characterization of interim efficiency with public goods, Econometrica 67 (1999b) 435–848.

[26] J. Ledyard, T. Palfrey, The approximation of efficient public good mechanisms by simple voting schemes, J. Public Econ. 83 (2002) 153–172.

[27] J. Ledyard,T. Palfrey, Ageneral characterization of interim efficient mechanisms for independent linear environments, J. Econ. Theory, 133 (2007) 441–466

[28] P. McAfee, Efficient allocation with continuous quantities, J. Econ. Theory, 53, (1991), 51-74.

[29] G. Mailath, A. Postlewaite, Asymmetric information bargaining problems with many agents, Rev. Econ. Stud. 57 (1990) 351–367.

[30] P. Milgrom, Putting Auction Theory to Work, Cambridge University Press, 2004.

[31] D. Mookherjee, S. Reichelstein, Dominant Strategy implementation of Baysian incentive compatible allocation rules, J. Econ. Theory, 56, (1992), 378-399.

[32] R. Myerson, Optimal auction design, Math. Oper. Res. 6 (1981) 58–73.

[33] R. Myerson, M. Satterthwaite, Efficient mechanisms for bilateral trading, J. Econ. Theory 28 (1983) 265–281.

[34] J. Riley,W. Samuelson, Optimal auctions, Quart. J. Econ. 71 (1981) 381–392.

[35] W. Vickrey, Counterspeculation, auctions, and competitive sealed tenders, J. Finance 19 (1961) 8–37.

[36] J. Wang, J.F. Zender, Auctioning divisible goods, Econ. Theory, 19 (2002), 673-705.

[37] R. Wilson Auctions of Shares. The Quarterly Journal of Economics, Vol. 93, (1979), pp. 675-689

[38] R.Wilson, Incentive efficiency of double auctions, Econometrica 53 (1985) 1101–1116.