Stochastic Dominance of Signals and Reparametrization in Adverse Selection Model

Yonggyun Kim*, Sunghee Lee**

Abstract

This paper investigates how a pair of signals about the type of the agent can be compared in the classical principal agent model with adverse selection. Signal comparison in this model has two distinctive features that make it difficult to directly apply the results from decision theory: timing of the game and the number of incentive compatibility constraint.

The signal in the model takes the form of probability distribution and two popular means of comparing a pair of probability distributions are considered: First Order Stochastic Dominance (FOSD) and Second Order Stochastic Dominance (SOSD). It is straightforwardly shown that FOSD relation implies more informativeness, which guarantees higher profit to the principal.

In contrast, SOSD relation is largely affected by the parametrization of the agent’s type and it might not guarantee more informativeness under some circumstances. Nevertheless, if the parameter of the agent is properly changed so that it reflects the principal’s profit rather than the agent’s cost, SOSD relation may guarantee higher profit to the principal. Under some appropriate conditions, this paper offers a construction algorithm for the reparametrization that SOSD relation implies more informativeness.

JEL Classification: C44, C60, D86

Keywords: Stochastic dominance, value of information, adverse selection, optimal contract, reparametrization.

*Department of Economics and Law, Division of Social Sciences, Korea Military Academy, P.O. Box 77, Gongneung-dong, Nowon-gu, Seoul, Republic of Korea. Tel.: +82 2 2197 2796 (office). e-mail: ygglorious@gmail.com, ygglorious@kma.ac.kr (corresponding author)
**Department of Economics and Law, Division of Social Sciences, Korea Military Academy, P.O. Box 77, Gongneung-dong, Nowon-gu, Seoul, Republic of Korea. Tel.: +82 2 2197 2778 (office). e-mail: shlee@kma.ac.kr

Preprint submitted to ICGT April 7, 2015
1. Introduction

Recent developments in decision theory and information economics have offered intuitive understandings for the information in the wide variety of economic nature. Especially, Chi [5] applies the idea of Lehmann [10] about the statistical ordering to the general economic environments. As a result, signal comparisons in various economic environments with uncertainty can be explained by Chi [5]. As representative applications of Chi [5], there are MIO criterion of Athey and Levin [3] for the monotone decision problem and MPS criterion of Kim [8] for moral hazard model. Nevertheless, there still remain some basic economic models concerning uncertainty, which cannot be viewed as a simple example of Chi [5]. One of the economic models is the classical principal-agent model with adverse selection, which is the base model of this paper.¹

This paper studies the criteria for the signals in the classical principal-agent model with adverse selection.² In this model, the principal delegates a production of a good to the agent. Both the principal and the agent know the value function \(S(q)\) of the principal and the cost function \(C(q, \theta)\) of the agent. However, the principal does not know the agent’s type \(\theta\) exactly so that the principal is uncertain about the agent’s exact cost. The principal’s only information about the agent’s type is a signal, which is a probability distribution of the agent’s type. The principal offers a contract to the agent based on the signal. Then, the agent chooses production level that maximizes his utility and the profit of the principal is realized. In this paper, I investigate the relation between a pair of signals that guarantees higher expected profit by a signal than that by the other signal.

There are features that make difference between the adverse selection model and the models that can be analyzed by Chi [5]. Firstly, as we can see in figure 1.1, the timing of decision is

---

¹For the simplest and intuitive model with two types of agents, see section 2 of Laffont and Martimort [9]. In this paper, I considered principal-agent model with discrete types of agents.

²As a seminal analysis on adverse selection, Akerlof [1] assumed information asymmetry on the types of used car (lemon and peach). The type is defined by the probability of suffering loss. However, in the adverse selection model of Laffont and Martimort [9], which is the base model of this paper, information asymmetry is assumed on the productivity of the agent.
DM chooses a signal $\alpha$ observes an outcome $x$ from $\alpha$ chooses an action $a \in A$ get realized Payoffs $u(a, \theta)$

$S$ and $C$ are drawn
A observes his type P offers a contract to A A chooses production level
P observes a signal $\nu$ (P chooses optimal contract) and payoffs are realized

Figure 1.1: Comparison of timing between Decision making in Chi [5] and principal-agent model with adverse selection.

different. In general decision theory, especially in Bayesian decision theory, the decision maker chooses an action after the revelation of the random variable. For example, in principal-agent model with moral hazard, an action $a$ is decided based on the outcome $x$, which is a realization from signal $\alpha$. However, in principal-agent model with adverse selection, the principal offers a contract before the type is revealed.$^3$

Secondly, because the agent can disguise as another type, incentive compatibility should be considered for all types of the agent. It means that, if the number of types of agent is $n$, there are $n$ groups of incentive compatibility constraints. On the other hand, in decision theory, for example in principal-agent model with moral hazard, it is enough to guarantee that the optimal action maximizes the agent’s expected utility and it is the only incentive compatibility constraint.

For an analysis, I adopt some basic results of Maskin and Riley [12]. Maskin and Riley [12] provided an analysis on second price discrimination of a monopolist, which can be applied to

$^3$This does not mean that all kinds of adverse selection model cannot be analyzed through Chi [5]. Levin [11] analyzed information in the adverse selection model based on Akerlof [1]. In the model Levin [11] considered, the result of Athey and Levin [3] was applied. The main difference of the model of this paper and the model of Levin [11] is whether the contract is involved or not. The model without contract can be analyzed much easier. Specifically, in Levin [11], after the seller’s type is revealed, buyer gets the same value no matter what the signal was. However, in this model, the signal changes the contract and eventually affects to the principal’s payoff after revelation.
this structure through appropriate transformation. The results of Maskin and Riley [12] are restated in subsection 2.2.

As methods of comparing two signals, which actually are probability distribution, I will use two popular relations - First Order Stochastic Dominance (FOSD) and Second Order Stochastic Dominance (SOSD). FOSD relation is generally considered as a robust way of signal comparison. In the adverse selection model I considered, FOSD relation guarantees the more expected profit to the principal. However, SOSD relation fails to work in my model. I will deliver the counterexample that SOSD relation cannot guarantee the more expected profit to the principal.

One of the reason that SOSD does not work well might be the choice of wrong parameter as the type of the agent. The parameter about the agent’s type in this model is related to the cost for the type. However, in order to make SOSD work well, it is needed to find a new parametrization that reflect the principal’s profit appropriately. In this paper, I define SOSD relation for a new parametrization and find some proper reparametrization that SOSD relation guarantee more informativeness under some condition.

The paper proceeds as follows. Section 2 presents the model and basic results. Section 3 relates first and second stochastic domination to informativeness. Section 4 defines reparametrization and constructs an algorithm for an appropriate reparametrization under some conditions. Section 5 concludes.

2. Discrete Type Adverse Selection Model

2.1. Setup

In this section, I introduce the basic model and several results which is the foundation of this analysis. I considered a classical adverse selection model, which assumes the information asymmetry between the principal and the agent on the agent’s type. The types of agents are

---

4See section 2 of Maskin and Riley [12].
5In monopoly and second price discrimination literature, there are several preceding works about information. Maskin and Riley [12] showed that the signal that has a lower hazard rate in the whole interval yields more steep slope on the principal’s profit. Ottaviani and Prat [14] compared public, private, and no information case and showed that the public signal always make the monopolist better off. Saak [16] proved that more private information for the buyer yields more profit to seller in a specific model.
assumed to be discrete in this paper.

The principal wants to delegate a production of a good to an agent. The value that the principal obtains from the production of \( q \) units of a good is defined by a value function \( S(q) \). I assume that the value function of the principal is drawn from the following set

\[
S = \{ S : \mathbb{R}_+ \to \mathbb{R}_+ \mid S \in C^2, S' > 0, S'' \leq 0 \text{ and } S(0) = 0 \},
\]

which implies that the principal is risk averse.

In the discrete type model, the type of the agent \( \theta \) is drawn from \( \Theta = \{ \theta_1, \ldots, \theta_n \} \) and the agent is exactly aware of his or her type. I assume that \( \Theta \subset \mathbb{R} \) and \( \theta_1 < \theta_2 < \cdots < \theta_n \). On the other hand, with signal \( \nu = (\nu_1, \ldots, \nu_n) \in \Delta \), where \( \Delta = \{ (p_1, \ldots, p_n) \in \mathbb{R}_+^n \mid p_1 + \cdots + p_n = 1 \} \), the principal only knows that, for all \( i \in \mathbb{N} = \{1, \ldots, n\} \), the probability that the agent’s type is \( \theta_i \) is \( \nu_i \).

The type \( \theta \) agent’s cost of producing \( q \) units of the good is denoted by a cost function \( C(q, \theta) \). There are several assumptions on the cost function. Firstly, I assume that \( C_q > 0, C_{qq} > 0, C(q, \theta_i) < C(q, \theta_j) \), and \( C_q(q, \theta_i) < C_q(q, \theta_j) \) for all \( 1 \leq i < j \leq n \) and \( q \in \mathbb{R} \), which implies that the cost function is increasing, convex for \( q \) and \( \theta \). I further assume that \( C_{qq}(q, \theta_i) < C_{qq}(q, \theta_j) \) and \( C_q(q, \theta_i) - C_q(q, \theta_{i-1}) \leq C_q(q, \theta_j) - C_q(q, \theta_{j-1}) \). Let \( \mathcal{C} \) be the set of functions that satisfy the preceding assumptions and the cost function of the agent is assumed to be drawn from \( \mathcal{C} \).

In order to avoid the case where the principal would not willing to make a contract, the value from producing \( q \) units (\( S(q) \)) is always greater than the cost (\( C(q, \theta) \)) for all \( \theta \). In other words, a pair of production function and cost function \((S, C)\) is

\[
\mathcal{P} = \{(S, C) \in S \times \mathcal{C} \mid S(q) - C(q, \theta) > 0 \text{ for all } q \in \mathbb{R}_+^+ \text{ and } \theta \in \Theta \}
\]

See Athey and Levin [3] and Chi [5] for the literature that a signal is corresponded to a probability distribution.

The last assumption \( C_q(q, \theta_i) < C_q(q, \theta_j) \) is the discrete version of Spence-Mirrlees condition. Though Araujo and Moreira [2] provided an analysis without Spence-Mirrlees condition, I imposed the condition as in most researches about the adverse selection.

These assumptions will be used in subsection 4.2 and (6.4) as \( \Phi_{qq}(q, i) > 0 \) and \( \Phi_{q}(q, i) < \Phi_{q}(q, j) \) where \( \Phi(q, i) = C(q, \theta_i) - C(q, \theta_{i-1}) \).
Note that the production function $S$ and the cost function $C$ are pieces of public information so that the principal and the agent know the value function and the cost function exactly.

The contract between the principal and the agent is designed by a set of pairs $\{(q(i), t(i))\}_{i \in N}$, where $q$ is the quantity of production, and $t$ is the monetary transfer from the principal to the agent. When the agent decides to take $(q, t)$, the principal’s payoff is $S(q) - t$ and the agent’s payoff is $t - C(q, \theta)$.

Now that I have offered basic notions, I will introduce the timing of this model. The model is constituted by three periods:

- At time 0, the value function $S$ of the principal and the cost function $C$ of the agent are drawn from $(S, C) \in \mathcal{P}$ and they are known to both the principal and the agent. In addition, the agent exactly observes his or her type $\theta$, on the contrary, the principal only observes a signal $\nu$ about the type of the agent.
- At time 1, Based on the signal $\nu$, the principal offers a contract $\{(q(i), t(i))\}_{i \in N}$ to the agent.
- At time 2, Agent chooses how much to produce and the payoff for the agent and the principal is realized.

The main interest of the paper is the evaluation of signal $\nu$ determined in time 0. To assess signal $\nu$, the optimal contract at time 1 should consider and it should maximizes the expected profit of the principal. Moreover, the contract should be designed for the agent with type $\theta_i$ to accept $(q(i), t(i))$. Thus, the agent should not have an incentive to deceive his type or to decide to quit. This property is characterized by the following two constraints.

The contract $\{(q(i), t(i))\}_{i \in N}$ satisfies the Incentive Compatibility Constraint (ICC) if

$$t(i) - C(q(i), \theta_i) \geq t(j) - C(q(j), \theta_j) \quad \forall i, j \in N,$$

which implies that $(q(i), t(i))$ is the best option for the agent with type $\theta_i$.  

6
I also assume that the reservation utility is normalized to 0, so that the contract \( \{(q(i), t(i))\}_{i \in N} \) satisfies the **Participation Constraint** (PC) if

\[
t(i) - C(q(i), \theta_i) \geq 0 \quad \forall i \in N.
\]

With the signal \( \nu \) and the contract \( \{(q(i), t(i))\}_{i \in N} \), the principal’s expected profit is \( \sum_{i=1}^{n} \nu_i \cdot (S(q(i)) - t(i)) \). Thus, the principal’s profit maximization problem is

\[
\max_{\{(q(i), t(i))\}_{i \in N}} \sum_{i=1}^{n} \nu_i \cdot (S(q(i)) - t(i)) \quad \text{s.t. ICC, PC.} \tag{2.1}
\]

Denote the optimal solution of (2.1) as \( \{(q^*_i, t^*_i)\}_{i \in N} \). In the next subsection, the properties of \( \{(q^*_i, t^*_i)\}_{i \in N} \) will be explored.

### 2.2. Basic Results

Now I will report several well known results in adverse selection literature which are baselines of this research. To begin with, I am interested in the principal’s optimal contract design at time 1, which is the optimal solution of (2.1). The optimal contract can be characterized by following propositions.\(^9\) The first part of proposition 2.1 characterizes the monotonicity of the production level on the optimal contract. The second and third part of proposition 2.1 derives the optimal transfer for the each production level. Based on those results, proposition 2.2 fully characterizes the optimal solution \( \{(q^*_i, t^*_i)\}_{i \in N} \).

**Proposition 2.1.** The optimal contract \( \{(q^*_i, t^*_i)\}_{i \in N} \) satisfies the following properties.

1. \( q^*_1 \geq q^*_2 \geq \cdots \geq q^*_n \)
2. \( t^*_i(n) = C(q^*_i(n), \theta_n) \)
3. For \( 1 \leq i \leq n - 1 \),
   \[
   t^*_i(i) = C(q^*_i(i), \theta_i) - C(q^*_i(i + 1), \theta_i) + t(i + 1) = C(q^*_i(i), \theta_i) + \sum_{j=i+1}^{n} \Phi(q^*_j(j), j),
   \]

   where \( \Phi(q, i) = C(q, \theta_i) - C(q, \theta_{i-1}) \).

**Proof.** The first part is equivalent to Proposition 1 of Maskin and Riley [12]. For the second and third parts, see Appendix 6.1

\(^9\)Originally, Maskin and Riley [12] provided these results in second price discrimination of monopolist problem. I restated these results in terms of adverse selection setup.
Then, by using \( \sum_{i=1}^{n} \nu_i \cdot \left\{ \sum_{j=i+1}^{n} \Phi(q(j), j) \right\} = \sum_{i=1}^{n} \nu_i \Phi(q(i), i) \), 2.1 can be transformed as follows:

\[
\sum_{i=1}^{n} \nu_i \cdot (S(q(i)) - t(i)) = \sum_{i=1}^{n} M_i(\nu, q(i)),
\]

where \( M_i(\nu, q(i)) = \nu_i S(q(i)) - \nu_i C(q(i), \theta_i) - \nu_{i-1} \Phi(q(i), i). \)

As another benchmark of the optimal solution, define the second best solution \((\bar{q}_\nu(i))_{i \in N}\) as an unconstrained maximization of \( \sum_{i=1}^{n} M_i(\nu, q(i)) \). The production level of \( i \)th type, \( q_i \), only related to \( M_i(\nu, q(i)) \), not to \( M_j(\nu, q(j)) \) for \( j \neq i \). Thus, the second best solution \( \bar{q}_\nu(i) \) is equal to the production level that maximizes \( M_i(\nu, q(i)) \) and it satisfies

\[
S'(\bar{q}_\nu(i)) = C_q(\bar{q}_\nu(i), \theta_i) + \frac{\nu_{i-1}}{\nu_i} \Phi_q(\bar{q}_\nu(i), i). \tag{2.2}
\]

In order to obtain the optimal solution, constraints for the production level should be imposed and the first part of 2.1 plays role as the constraints. Thus, by solving the maximization problem \( \sum_{i=1}^{n} M_i(\nu, q(i)) \) subject to \( q_\nu(1) \geq q_\nu(2) \geq \cdots \geq q_\nu(n) \), the optimal solution is characterized as following proposition.

**Proposition 2.2.** (Discrete version of Proposition 7 in Maskin and Riley [12] and Theorem 4 in Guesnerie and Laffont [7]) The optimal production level \((q^*_\nu(i))_{i \in N}\) is characterized as follows:

1. \( q^*_\nu(i) \) coincides with \( \bar{q}_\nu(i) \) except on a finite number \( K \) of disjoint sets of consecutive types \( I_k = \{\theta_{a_k}, \cdots, \theta_{b_k}\}, k = 1, \cdots, K, a_k \) increasing with \( k \), where \( q^*_\nu(i) = q^k \) for all \( i \in [a_k, \cdots, b_k] \).
2. For all \( 1 \leq k \leq K \), \( q^k \) satisfies the following equation:

\[
\sum_{a_k}^{b_k} \left( \nu_i S'(q^k) - \nu_i C_q(q^k, \theta_i) - \nu_{i-1} \Phi_q(q^k, i) \right) = 0.
\]
3. If \( q^*_\nu(i) > q^*_\nu(i+1) \), the following inequality holds:

\[
\nu_i S'(q^*_\nu(i)) - \nu_i C_q(q^*_\nu(i), \theta_i) - \nu_{i-1} \Phi_q(q^*_\nu(i), i) \geq 0.
\]

**Proof.** See Appendix 6.2 \( \square \)

When the optimal contract is decided, the principal’s profit can be viewed as the expectation of the principal’s profit with the agent’s realized type \( \theta \). Precisely, define \( P_\nu(i) \) as the principal’s profit when the agent’s type is \( \theta_i \), the probability distribution is \( \nu \) and the optimal contract is
implemented, which is equal to $S(q^*_v(i)) - t^*_v(i)$.\(^{10}\) Define $P(\nu)$ as the principal's profit with the optimal contract when the signal is $\nu$, that is,

$$P(\nu) = \sum_{i=1}^{n} \nu_i \cdot P_\nu(i) = E_\nu [P_\nu(i)].$$

Following proposition and corollary state the property of the function $P_\nu$, which is key for the analysis.

**Proposition 2.3.** For $i < j$, the following equation and inequality hold:

1. If $q^*_v(i) = q^*_v(j)$, $P_\nu(i) = P_\nu(j)$;
2. If $q^*_v(i) > q^*_v(j)$, $P_\nu(i) > P_\nu(j)$.

*Proof.* See Appendix 6.3. \(\square\)

**Corollary 2.4.** $P_\nu(\cdot) : N \to \mathbb{R}$ is a weakly decreasing function, that is, $P_\nu(i) \geq P_\nu(i + 1)$ for all $i \in \mathbb{N}\{n\}$.

Proposition 2.3 and corollary 2.4 are intuitive results: the principal earns more profit if the type of the agent is more efficient. When there exists $\nu \in \Delta$ such that $P_\nu(i) = P_\nu(j)$ for some $i \neq j$, there might be a trouble in the subsection 6.4. By restricting the domain of the signal, the weakly decreasing property of the function $P_\nu$ can be converted into the decreasingness of the function $P_\nu$. We can see that in the following definition, proposition 2.5 and corollary 2.6.

**Definition.** A subset $\Delta^m$ of domain $\Delta$ is defined as follows:

$$\Delta^m \equiv \left\{ \nu \in \Delta \mid \frac{V_i}{\nu_i} \leq \frac{V_{i+1}}{\nu_{i+1}} \text{ for all } 1 \leq i \leq n-1 \right\}. \quad (2.3)$$

**Proposition 2.5.** For $\nu \in \Delta^m$, $\bar{q}_\nu(\cdot) : N \to \mathbb{R}$ is a decreasing function, that is, $\bar{q}_\nu(i) > \bar{q}_\nu(i + 1)$ for all $i \in \mathbb{N}\{n\}$. Thus, $q^*_\nu(\cdot) : N \to \mathbb{R}$ is also a decreasing function.

*Proof.* See Appendix 6.4. \(\square\)

**Corollary 2.6.** For $\nu \in \Delta^m$, $P_\nu(\cdot) : N \to \mathbb{R}$ is a decreasing function, that is, $P_\nu(i) > P_\nu(i + 1)$ for all $i \in \mathbb{N}\{n\}$.

As a final of the basic results, the derivative of $P(\nu)$ is derived and it is familiar form. The following proposition is derived from the multivariate version of envelope theorem.\(^{11}\)

\(^{10}\)Let's abuse the notation $P_\nu(\theta_i) = P_\nu(i) = S(q^*_v(i)) - t^*_v(i)$

\(^{11}\)Milgrom and Segal [13] provided the seminal idea of general envelope theorem for the single parameter.
Proposition 2.7. For all $i \in N$, the following holds:

$$\frac{\partial P(\nu)}{\partial \nu_i} = P(\nu(i)).$$

Proof. See Appendix (6.5). \qed

3. Stochastic Dominance and Profit of the Principal

3.1. Stochastic Dominance Relation and Information Order

The main concern of this paper is the quality of a signal, which is the probability distribution of the type. In this section, I introduce two widely used criteria for the signal comparison - First Order Stochastic Dominance (FOSD) and Second Order Stochastic Dominance (SOSD).

Definition. A signal $\nu'$ is said to be First Order Stochastically Dominant (FOSD) over $\nu$ if

$$V'_1 \geq V_1,$$
$$V'_2 \geq V_2,$$
$$\vdots$$
$$V'_n \geq V_n,$$

where $V_i = \sum_{j=1}^{i} \nu_j$ and $V'_i = \sum_{j=1}^{i} \nu'_j$, and denote $\nu' \succ_{F} \nu$.\textsuperscript{12}

Although FOSD is a good criterion for the signal comparison, FOSD has a limitation on the range of the comparison. For example, FOSD relation implies that the mean of the agent’s type with the signal $\nu'$ is greater than that with the signal $\nu$. It means that any pair of two signals with same average type cannot be compared by FOSD relation.

Definition. A signal $\nu'$ is said to be Second Order Stochastically Dominant (SOSD) over $\nu$ if

$$\sum_{i=1}^{1} V'_i \cdot (\theta_{i+1} - \theta_i) \geq \sum_{i=1}^{1} V_i \cdot (\theta_{i+1} - \theta_i),$$
$$\sum_{i=1}^{2} V'_i \cdot (\theta_{i+1} - \theta_i) \geq \sum_{i=1}^{2} V_i \cdot (\theta_{i+1} - \theta_i),$$
$$\vdots$$
$$\sum_{i=1}^{n-1} V'_i \cdot (\theta_{i+1} - \theta_i) \geq \sum_{i=1}^{n-1} V_i \cdot (\theta_{i+1} - \theta_i),$$

and denote $\nu' \succ_{S} \nu$.\textsuperscript{13}

\textsuperscript{12}In most models, which assume that the higher index implies the better to the principal, FOSD relation is defined by inequalities reverse to (3.1). However, in this setup, the lower index implies the more efficient type of agent, which is better to the principal so that I use the reverse inequality.

\textsuperscript{13}In continuous setup, SOSD is defined as $\int_{\theta_0}^{\theta_1} F(s) ds \geq \int_{\theta_0}^{\theta_1} G(s) ds$ for all $\theta \in [\theta_0, \theta_1]$. 

10
SOSD plays a role as a refinement of FOSD. Some pair of two signals with same means can be compared through SOSD. In SOSD criterion, we can compare two signals with same mean if they satisfy (3.2). When the signal $\nu$ and $\nu'$ yield same mean and $\nu'$ is SOSD over $\nu$, we call $\nu'$ a Mean Preserving Spread of $\nu$.

Remind that the main interest of this research is comparing a pair of signals. The main question requests to define which signal is “more Informative”. If a signal $\nu'$ yields more expected profit than another signal $\nu$ yields no matter what the value function and the cost function are, $\nu'$ is said to be “more informative” than $\nu$. In order to define informative criterion for a pair of signals precisely, it is needed to denote the expected profit function $P$ with respect to the value function $S$ and the cost function $C$ as $P_{S,C}$. Precise definition of “more informativeness” is as follows.

**Definition.** A signal $\nu'$ is said to be more informative than $\nu$ on $P$, if

$$P_{S,C}(\nu') \geq P_{S,C}(\nu),$$

for all $(S, C) \in P$.

In the following subsections, it will be figured out how FOSD and SOSD are related to “more informativeness”.

### 3.2. FOSD and Information Order - Possibility Result

In this subsection, it is shown that FOSD relation is a sufficient condition of more informativeness. The following proposition I use property of FOSD.

**Proposition 3.1.** If $\nu'$ is First Order Stochastically Dominant over $\nu$ ($\nu' \succ_F \nu$) and $f : N \to \mathbb{R}$ satisfies $f(i) \geq f(j)$ for all $i < j$, the following inequality holds:

$$E_{\nu'}[f(\cdot)] = \sum_{i=1}^{n} f(i) \cdot \nu'_i \geq \sum_{i=1}^{n} f(i) \cdot \nu_i = E_{\nu}[f(\cdot)].$$

**Proof.** See Courtault et al. [6].

\[14\] Still, there are also some pairs of two signals with same means that cannot be compared through SOSD.

\[15\] See section 3.4 of Bikhchandani et al. [4] for the proof for continuous case.
Together with decreasingness of \( P \) and the property of FOSD, the definition of \( P \) which is maximized profit.

**Theorem 3.2.** When a signal \( \nu' \) is First Order Stochastically Dominant over another signal \( \nu \) (\( \nu' \succeq_F \nu \)), \( \nu' \) is more informative than \( \nu \) on \( P \), that is, \( P_{S,C}(\nu') \geq P_{S,C}(\nu) \) for all \( (S,C) \in \mathcal{P} \).

**Proof.** For convenience, take \( (S,C) \in \mathcal{P} \) and abuse the notation \( P \) as \( P_{S,C} \).

By 2.3, the function \( P_{\nu} \) is weakly decreasing function. Thus, by proposition 3.1,

\[
P_{\nu} = \sum_{i=1}^{n} \nu_i \cdot P_{\nu}(i) = E_{\nu} [P_{\nu}(\cdot)] \leq E_{\nu'} [P_{\nu'}(\cdot)] = \sum_{i=1}^{n} \nu'_i \cdot P_{\nu}(i).
\]

The right side of the above inequality is equal to \( \sum_{i=1}^{n} \nu'_i \cdot \{S(q^*_i(\nu)) - t^*_i(\nu)\} \). From the definition of \( q^*_i(\nu') \), which maximizes the principal’s expected profit, and the fact that \( \{(q^*_i, t^*_i)\}_{i \in \mathbb{N}} \) satisfies ICC and PC, the following inequality holds:

\[
\sum_{i=1}^{n} \nu'_i \cdot \{S(q^*_i(\nu)) - t^*_i(\nu)\} \leq \sum_{i=1}^{n} \nu'_i \cdot \{S(q^*_i(\nu')) - t^*_i(\nu')\} = P_{\nu'}.
\]

Thus, the inequality \( P_{\nu} \leq P_{\nu'} \) holds.

Theorem 3.2 confirms that FOSD relation guarantees more profit to the principal. In FOSD, the ordinality of \( \theta \) is the only consideration.

As a result of Theorem 3.2, the model with two types can be fully characterized as following corollary.

**Corollary 3.3.** In the model with two types, the principal’s profit \( P_{\nu} \) increases as the probability that the agent is more efficient, \( \nu_1 \), increases.

For the model with two types, any two signals can be compared through FOSD criteria. However, for \( n \geq 3 \), there are many pairs of signals which are not compared by FOSD. Our natural question is whether SOSD, as a widely used refinement of FOSD, can guarantee higher profit to the principal.

### 3.3. Second Order Stochastic Dominance - Counterexample

In this subsection, the answer for the question in the last subsection is provided through a counterexample. I considered a simple example with three types that SOSD cannot guarantee higher profit to the principal.
In this example, value function is $S(q) = 6\sqrt{q}$ and cost function is $C(q, \theta) = \theta q^{3/2}$ for type $\theta \in \Theta = \{1, 1.5, 2\}$. Then, second best solution for this example can be derived through 2.2 as follows:

$$
\bar{q}_1(\nu) = 2, \quad \bar{q}_2(\nu) = \frac{2}{1 + \frac{1}{2^{1/3} 0.5}}, \quad \bar{q}_3(\nu) = \frac{2}{1 + \frac{1}{2^{1/3} 0.5}}. \quad (16)
$$

Then, consider three signals $\nu_A = (0.28, 0.44, 0.28), \nu_B = (0.29, 0.42, 0.29), \nu_C = (0.30, 0.40, 0.30)$. They are not comparable by FOSD, however, can be ordered through SOSD: $\nu_C \succ_S \nu_B \succ_S \nu_A$.

The optimal contract and the profit of the principal for each signal can be seen in the table 1. In all three cases, the second best solution is weakly decreasing so that the optimal solution is equal to the second best solution.

If SOSD relation can guarantee higher profit to the principal as FOSD does in theorem 3.2, $P(\nu_C) \geq P(\nu_B) \geq P(\nu_A)$ should hold. However, as we can see in table 1, $P(\nu_A) > P(\nu_C) > P(\nu_B)$.

**Proposition 3.4.** In discrete type model, SOSD relation cannot guarantee higher profit to the principal.

Why does this result happen? In SOSD, not as in FOSD, the dominance relation is affected not only the ordinality but also the cardinality of $\theta$ critically. However, the cardinality of $\theta$ does not have proper information about the agent with type $\theta$. For example, consider two

---

\[\text{Table 1: The numerical example that SOSD does not guarantee higher profit.}\]

<table>
<thead>
<tr>
<th>Type ($\theta$)</th>
<th>Signal ($\nu$)</th>
<th>$q^*$</th>
<th>$t^*$</th>
<th>$P_\nu(\theta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.280</td>
<td>2.000</td>
<td>3.643</td>
<td>4.843</td>
</tr>
<tr>
<td>1.5</td>
<td>0.440</td>
<td>1.100</td>
<td>1.968</td>
<td>4.325</td>
</tr>
<tr>
<td>2</td>
<td>0.280</td>
<td>0.609</td>
<td>0.950</td>
<td>3.731</td>
</tr>
<tr>
<td>Principal’s Expected Profit ($P(\nu_A)$)</td>
<td><strong>4.30364</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Case B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.290</td>
<td>2.000</td>
<td>3.637</td>
<td>4.848</td>
</tr>
<tr>
<td>1.5</td>
<td>0.420</td>
<td>1.084</td>
<td>1.937</td>
<td>4.310</td>
</tr>
<tr>
<td>2</td>
<td>0.290</td>
<td>0.620</td>
<td>0.977</td>
<td>3.748</td>
</tr>
<tr>
<td>Principal’s Expected Profit ($P(\nu_A)$)</td>
<td><strong>4.30314</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Case C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.300</td>
<td>2.000</td>
<td>3.630</td>
<td>4.855</td>
</tr>
<tr>
<td>1.5</td>
<td>0.400</td>
<td>1.067</td>
<td>1.903</td>
<td>4.293</td>
</tr>
<tr>
<td>2</td>
<td>0.300</td>
<td>0.632</td>
<td>1.004</td>
<td>3.764</td>
</tr>
<tr>
<td>Principal’s Expected Profit ($P(\nu_B)$)</td>
<td><strong>4.30319</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[16\] The first best solution for this example is $(q_{FB}^P, t_{FB}^P) = (2/\theta, 2\sqrt{2}/\sqrt{\theta})$. 

13
cost functions $C(q, \theta) = \theta q^{3/2}$ and $\tilde{C}(q, \tilde{\theta}) = \sqrt{\tilde{\theta} q^{3/2}}$, where $\theta \in \Theta = \{1, 1.5, 2\}$ and $\tilde{\theta} \in \tilde{\Theta} = \{1, 2.25, 4\}$. Practically, $C$ and $\tilde{C}$ are same function, however, the parameter $\theta$ and $\tilde{\theta}$ give totally different mean preserving spreads. Moreover, the parameter $\theta$ only reflects the property of the cost function, while the parameter should reflect the principal’s profit in order to make SOSD work well.

Then, the next question is whether and how we can find the proper parameter $\alpha$ that not only preserves the order of $\theta$ but also satisfies SOSD criteria.

4. reparametrization

4.1. $\alpha$-Reparametrized SOSD

This section deals with the reparametrization of the agent’s type and second order stochastic dominance. Firstly, second order stochastic dominance relation is redefined on the newly parametrized type of the agent. Secondly, the condition that makes SOSD on the reparametrization work well will be investigated.

Definition. For a vector $\alpha \in \mathbb{R}_+^n$, a signal $\nu'$ is said to be $\alpha$-Reparametrized Second Order Stochastically Dominant ($\alpha$-RSOSD) over $\nu$ if

$$\sum_{i=1}^{1} V_i' \cdot (\alpha_{i+1} - \alpha_i) \geq \sum_{i=1}^{1} V_i \cdot (\alpha_{i+1} - \alpha_i),$$
$$\sum_{i=2}^{2} V_i' \cdot (\alpha_{i+1} - \alpha_i) \geq \sum_{i=2}^{2} V_i \cdot (\alpha_{i+1} - \alpha_i),$$
$$\vdots$$
$$\sum_{i=1}^{n} V_i' \cdot (\alpha_{i+1} - \alpha_i) \geq \sum_{i=1}^{n} V_i \cdot (\alpha_{i+1} - \alpha_i).$$

In order to find a new parametrization $\alpha = (\alpha_1, \ldots, \alpha_n)$ with $\alpha_1 < \alpha_2 < \cdots < \alpha_n$ that $\alpha$-RSOSD criterion works well, we need to check the most basic form of mean preserving spread.

Following Rothschild and Stiglitz [15], for $i_1 < i_2 \leq i_3 < i_4$, consider $\gamma_{i_1}, \gamma_{i_2}, \gamma_{i_3}, \gamma_{i_4}$ such that $\gamma_{i_1} = -\gamma_{i_2} > 0$ and $-\gamma_{i_3} = \gamma_{i_4} > 0$ and

$$\alpha_{i_1} \cdot \gamma_{i_1} + \alpha_{i_2} \cdot \gamma_{i_2} + \alpha_{i_3} \cdot \gamma_{i_3} + \alpha_{i_4} \cdot \gamma_{i_4} = 0.$$  \hspace{1cm} (4.2)

Consider a vector $\gamma = (0, \cdots, \gamma_{i_1}, \cdots, \gamma_{i_2}, \cdots, \gamma_{i_3}, \cdots, \gamma_{i_4}, \cdots, 0)$. To satisfy SOSD criterion, mean preserving spread toward $\gamma$ direction should yield more payoff to the principal.

\textsuperscript{17}All components except $i_1, i_2, i_3, i_4$th components are zero. If $i_2 = i_3$, define $i_3$th component as $\gamma_{i_2} + \gamma_{i_3}$.
Then, by proposition 2.7,
\[ \nabla \gamma P(\nu) = \gamma \cdot \nabla P(\nu) = \sum_{j=1}^{4} \gamma_{ij} \cdot P_{\nu}(i_{j}) \]
(4.3)
is nonnegative.

Let \( \delta = \gamma_{i_1} = -\gamma_{i_2} > 0 \). From (4.2), \( \left( \frac{\alpha_{i_4} - \alpha_{i_3}}{\alpha_{i_4} - \alpha_{i_1}} \right) \cdot \delta = -\gamma_{i_3} = \gamma_{i_4} > 0 \). Then, from the nonnegativity of (4.3), the following inequality holds:
\[ (P_{\nu}(i_1) - P_{\nu}(i_2)) \cdot (\alpha_{i_2} - \alpha_{i_1}) \geq (P_{\nu}(i_3) - P_{\nu}(i_4)) \cdot (\alpha_{i_4} - \alpha_{i_3}). \]
(4.4)

Note that the right side of (4.4) is nonnegative by corollary 2.4 and \( \alpha_{i_4} > \alpha_{i_3} \).

4.2. Construction of appropriate reparametrization \( \alpha^* \) on \( \triangle^m \)

If there exists \( \nu \in \triangle \) and \( i < j \) such that \( \bar{q}_{\nu}(i) \leq \bar{q}_{\nu}(j) \), then \( q^*_{\nu}(i) = q^*_{\nu}(j) \) and \( P_{\nu}(i) = P_{\nu}(j) \) by proposition 2.3. Then, for any \( (k, l) \) satisfying \( j \leq k < l \), by (4.4) and the nonnegativity of the right side of (4.4), \( q^*_{\nu}(k) > q^*_{\nu}(l) \) implies that \( \alpha_{k} = \alpha_{l} \), which is not desirable to the reparametrization.

Thus, if we cannot assure that \( q^*_{\nu}() \) is a decreasing function, there would be no proper \( \alpha \) that makes \( \alpha - RSOSD \) guarantee the higher profit to the principal. In this sense, I restrict the domain of a signal to \( \triangle^m \subset \triangle \), which assures that \( q^*_{\nu}() \) is decreasing by proposition 2.5.

Then, how can we construct appropriate reparametrization? There are two important considerations for the construction. Firstly, as I considered earlier, increasingness of \( \alpha \) should be considered. Secondly, (4.4) should be satisfied for all signal \( \nu \in \triangle^m \) and quadruple \((i_1, i_2, i_3, i_4)\) with \( i_1 < i_2 \leq i_3 < i_4 \).

For example, let us construct \( \alpha^*_{3} \) when \( \alpha^*_{1} \) and \( \alpha^*_{2} \) are given as 1 and 2. By (4.4), for all \( \nu \in \triangle^m \), \( \alpha^*_{3} \) should satisfy
\[ (P_{\nu}(1) - P_{\nu}(2)) \cdot (\alpha^*_{2} - \alpha^*_{1}) \geq (P_{\nu}(2) - P_{\nu}(3)) \cdot (\alpha^*_{3} - \alpha^*_{2}), \]
which is equivalent to
\[ \frac{P_{\nu}(1) - P_{\nu}(2)}{P_{\nu}(2) - P_{\nu}(3)} \cdot (\alpha^*_{2} - \alpha^*_{1}) + \alpha^*_{2} \geq \alpha^*_{3}, \]
(4.4)
for all $\nu \in \Delta^m$.\footnote{Note that $P_\nu(2) - P_\nu(3)$ is always positive since $\nu \in \Delta^m$ so that dividing the inequality by $P_\nu(2) - P_\nu(3)$ does not change the sign.}

Define $\alpha^*_3$ as follows:

$$\alpha^*_3 \equiv \inf \left\{ \left( \frac{P_\nu(1) - P_\nu(2)}{P_\nu(2) - P_\nu(n)} \right) + \alpha^*_2 \mid \nu \in \Delta^m \right\}. \quad (4.5)$$

This construction inherently satisfies (4.4). Moreover, unless $\inf \{ P_\nu(1) - P_\nu(2) \mid \nu \in \Delta^m \} = 0$, $\alpha^*_3 > \alpha^*_2$ satisfies. Thus, (4.5) defines $\alpha^*_3$ properly so that $\alpha^*$-RSOSD works well.

As we can see in the preceding argument, $\nu \in \Delta^m$ guarantees $P_\nu(i) > P_\nu(j)$ for $i < j$, however, the infimum of $P_\nu(i) - P_\nu(j)$ cannot be guaranteed as positive. Thus,

**Definition.** $(S, C) \in \mathcal{P}$ is called to satisfy **Nondegenerate Condition** when

$$\inf \{ P_{S,C,\nu}(i) - P_{S,C,\nu}(j) \mid \nu \in \Delta^m \} > 0,$$

for all $1 \leq i < j \leq n$.

Based on the preceding construction of $\alpha^*_3$, under nondegenerate condition, the appropriate $\alpha^*$ can be constructed through the following inductive construction. We can see the result on Theorem 4.1.

**Definition.** Inductive Construction

1. $\alpha^*_1 = 1$, $\alpha^*_2 = 2$.
2. Given $\alpha^*_1, \ldots, \alpha^*_l$, define $\alpha^*_{l+1}$ as follows:

$$\alpha^*_{l+1} \equiv \inf \left\{ \left( \frac{P_\nu(i) - P_\nu(j)}{P_\nu(k) - P_\nu(n)} \right) \cdot (\alpha^*_j - \alpha^*_i) + \alpha^*_k \mid \nu \in \Delta^m, 1 \leq i < j < k \leq l \right\}. \quad (4.6)$$

**Theorem 4.1.** When $(S, C) \in \mathcal{P}$ satisfies nondegenerate condition, through inductive construction, we can construct $\alpha^*$ with $\alpha^*_1 < \alpha^*_2 < \cdots < \alpha^*_n$ such that if $\nu'$ is $\alpha^*$-RSOSD over $\nu$, $P(\nu) \leq P(\nu')$.

**Proof.** First, I will show that (4.4) satisfies for any $1 \leq i < j \leq k \leq l$.

By (4.6) for $l + 1$ and $P_\nu(l + 1) > P_\nu(n)$,

$$\frac{P_\nu(i) - P_\nu(j)}{P_\nu(k) - P_\nu(l + 1)} \cdot (\alpha^*_j - \alpha^*_i) + \alpha^*_k > \alpha^*_{l+1},$$

which implies (4.4).

Second, I will show that $\alpha^*_{l+1} > \alpha^*_l$. 

18
When $k = l$, it is clear that
\[
\left( \frac{P_\nu(i) - P_\nu(j)}{P_\nu(k) - P_\nu(n)} \right) \cdot (\alpha_j^* - \alpha_i^*) + \alpha_k^* > \alpha_l^*.
\]
since $\alpha_k^* = \alpha_l^*$.

When $k < l$, the definition of $\alpha_j^*$ implies that for $1 \leq i < j \leq k \leq l - 1$
\[
\left( \frac{P_\nu(i) - P_\nu(j)}{P_\nu(k) - P_\nu(n)} \right) \cdot (\alpha_j^* - \alpha_i^*) + \alpha_k^* > \alpha_l^*.
\]

5. Conclusion

In this paper, I explored information criteria for the principal-agent model with adverse selection, which cannot be analyzed directly through Chi [5] by difference in the timing of decision and incentive structure. Based on the results of preceding works, it is proved that first order stochastic dominance relation for a pair of signals implies more informativeness. However, it is shown that second order stochastic dominance relation cannot guarantee higher profit to the principal through a counterexample. Nevertheless, under some conditions, through reparametrization, second order stochastic dominance can play a role as an information criterion. Under the restricted domain $\Delta^m$ and with nondegenerate condition, I offered a construction of reparametrization that makes second order stochastic dominance guarantee higher profit to the principal.
6. Appendix

6.1. Proof of Proposition 2.1

The second and third part of proposition 2.1 is proved by the following two lemmas.

Lemma 6.1. Among Participation Constraints, the only binding constraint is \( t(n) - C(q(n), \theta_n) \geq 0 \).

Proof. For all \( i \in \{ 1, 2, \ldots, n-1 \} \), \( t(i) - C(q(i), \theta_i) \geq t(n) - C(q(n), \theta_i) \) by the ICC of type \( i \). Then, the right side of the inequality is equal to \( t(n) - C(q(n), \theta_n) + C(q(n), \theta_i) - C(q(n), \theta_i) \), which is always positive by the PC of type \( n \) and \( C(q(n), \theta_n) > C(q(n), \theta_i) \). Thus, for \( 2 \leq i \leq n \), the PC of type \( i \) does not bind.

Next, let’s show that the PC of type \( n \) binds. Suppose that the optimal monetary transfer \( (t(i))_{i \in N} \) does not bind the PC of type \( n \). Then, consider \( t'(i) = t(i) - \epsilon \) for all \( i \in N \), where \( \epsilon = t(n) - C(q(n), \theta_n) \). Then, \( (t'(i))_{i \in N} \) still satisfy participation constraints and incentive compatibility constraints and raise the principal’s profit \( \epsilon \) more than \( (t(i))_{i \in N} \), which is contradictory to the optimality of \( (t(i))_{i \in N} \). Thus, the PC of type \( n \) binds.

Lemma 6.2. For \( i \neq j \), incentive compatibility constraint \( t(i) - C(q(i), \theta_i) \geq t(j) - C(q(j), \theta_i) \) binds if and only if \( q^*(i) = q^*(j) \) or there exists \((i, j')\) such that \( q^*(i) = q^*(i') \), \( q^*(j) = q^*(j') \) and \( j' = i' + 1 \).

Proof. The following four cases encompasses all possible cases.

1. If \( q^*(i) = q^*(j) \), from ICC for \( i \) and \( j \), \( t(i) = t(j) \). Thus, ICC binds.

2. When there exists \((i', j')\) such that \( q^*(i) = q^*(i') \), \( q^*(j) = q^*(j') \) and \( j' = i' + 1 \), suppose that \( \epsilon = \{ t(i) - C(q(i), \theta_i) \} - \{ t(j) - C(q(j), \theta_i) \} > 0 \).

Then, the following monetary transfer still satisfies ICC and PC and raises the principal’s expected profit:

\[
t'(k) = \begin{cases} 
  t(k) & \text{if } k \leq j', \\
  t(k) - \epsilon & \text{if } k \geq i'.
\end{cases}
\]

Thus, ICC binds.

3. When \( i > j \) and \( q^*(i) < q^*(j) \), suppose that \( t(j) - C(q(j), \theta_j) \geq t(i) - C(q(i), \theta_j) \) binds. Then,

\[
t(j) = t(i) + C(q(j), \theta_j) - C(q(i), \theta_j)
\]

\[
  = t(i) + \int_{q(i)}^{q(j)} C_q(q, \theta_j) dq.
\]

\[\text{From the above paragraph, the participation constraints for other than type } n \text{ always does not bind. Incentive compatibility constraints for } (t'_i)_{i \in N} \text{ also works well because they can be derived by subtracting } \epsilon \text{ to the both sides of the incentive compatibility constraints for } (t_i)_{i \in N}.\]
The right side of equation is greater than \( t(i) + \int_{q(i)}^{q(j)} C_q(q, \theta_i) dq \), which is equal to \( t(i) + C_q(q(j), \theta_i) - C_q(q(i), \theta_i) \). This does not satisfy ICC for type \( i \). Thus, in this case, ICC should not bind.

4. When there exists \( k \) such that \( j > k > i \) and \( q^*(j) < q^*(k) < q^*(i) \), from ICC for \( i \) and \( k \),

\[
t(i) - C(q(i), \theta_i) \geq t(k) - C(q(k), \theta_i),
\]

and from ICC for \( k \) and \( j \),

\[
t(k) - C(q(k), \theta_k) \geq t(j) - C(q(j), \theta_k).
\]

Then, from \( C(q', \theta) - C(q'', \theta) = \int_{q'}^{q''} C_q(q, \theta) dq \) and \( C_q(q, \theta_k) > C_q(q, \theta_i) \),

\[
C(q(k), \theta_k) - C(q(j), \theta_k) > C(q(k), \theta_i) - C(q(j), \theta_i)
\]

and by plugging this inequality into (6.1), we obtain

\[
t(k) - C(q(k), \theta_i) > t(j) - C(q(j), \theta_i).
\]

Thus, ICC does not bind.

6.2. Proof of Proposition 2.2

- Maximization Problem :

\[
\max_{(q(i))_{i \in N}} \sum_{i=1}^{n} \left( \nu_i S(q(i)) - \nu_i C(q(i), \theta_i) - V_{i-1} \Phi(q(i), i) \right)
\]

subject to \( q(1) \geq q(2), \ q(2) \geq q(3), \cdots, \ q(n-1) \geq q(n) \).

- Lagrange Multiplier \( \lambda_i \geq 0 \) for \( q(i) \geq q(i+1) \).

- F.O.C. for \( i = 1 \)

\[
\nu_1 S'(q^*(1)) - \nu_1 C(q^*(1), \theta_1) + \lambda_1 = 0.
\]

- F.O.C. for \( 2 \leq i \leq n-1 \)

\[
\nu_i S'(q^*(i)) - \nu_i C(q^*(i), \theta_i) - V_{i-1} \Phi(q^*(i), i) + \lambda_i - \lambda_{i-1} = 0.
\]
• F.O.C. for \( i = n \)

\[
\nu_n S'(q^*_\nu(n)) - \nu_n C_q(q^*_\nu(n), \theta_n) - V_{n-1} \Phi_q(q^*_\nu(n), n) - \lambda_{n-1} = 0.
\]

• First, we can take \( I_1 = \{\theta_{a_1}, \ldots, \theta_{b_1}\} \) by taking \( a_1 = \min_i \{i \in N \mid q^*_\nu(i) = q^*_\nu(i + 1)\} \) and 
\( b_1 = \min_i \{i \in N \mid q^*_\nu(a_1) > q^*_\nu(i + 1)\} \).

• Second, take \( I_2 = \{\theta_{a_2}, \ldots, \theta_{b_2}\} \) by taking \( a_2 = \min_i \{i \in N \setminus \{1, 2, \ldots, b_1\} \mid q^*_\nu(i) = q^*_\nu(i + 1)\} \) and 
\( b_2 = \min_i \{i \in N \setminus \{1, 2, \ldots, b_1\} \mid q^*_\nu(a_2) > q^*_\nu(i + 1)\} \). Inductively so on.

• If \( \theta_i \notin \bigcup_{k=1}^{K} I_k \), \( \theta_i = \theta_1 \) or \( \theta_i = \theta_n \) or \( q^*_\nu(i - 1) > q^*_\nu(i) > q^*_\nu(i + 1) \).

  - When \( \theta_i = \theta_1 \), \( q^*_\nu(1) > q^*_\nu(2) \) and \( \lambda_1 = 0 \). Thus, from the F.O.C. for \( i = 1 \),

\[
\nu_1 S'(q^*_\nu(1)) - \nu_1 C_q(q^*_\nu(1), \theta_1) = 0,
\]

which implies that \( q^*_\nu(1) = \bar{q}_\nu(1) = q^*_\nuF(1) \).

  - When \( \theta_i = \theta_n \), \( q^*_\nu(n - 1) > q^*_\nu(n) \) and \( \lambda_{n-1} = 0 \). Thus, from the F.O.C. for \( i = n \),

\[
\nu_n S'(q^*_\nu(n)) - \nu_n C_q(q^*_\nu(n), \theta_n) - V_{n-1} \Phi_q(q^*_\nu(n), n) = 0,
\]

which implies that \( q^*_\nu(n) = \bar{q}_\nu(n) \).

  - When \( q^*_\nu(i - 1) > q^*_\nu(i) > q^*_\nu(i + 1) \), \( \lambda_{i-1} = \lambda_i = 0 \). Thus, from the F.O.C. for \( i \),

\[
\nu_i S'(q^*_\nu(i)) - \nu_i C_q(q^*_\nu(i), \theta_i) - V_{i-1} \Phi_q(q^*_\nu(i), i) = 0,
\]

which implies that \( q^*_\nu(i) = \bar{q}_\nu(i) \).

6.3. Proof of Proposition 2.3

Proof. It is sufficient to show that for \( i \) and \( i + 1 \). From the definition of \( P_i(\nu) \) and proposition 2.1, it can be derived that

\[
P_i(\nu) - P_i(\nu + 1) = \{S(q^*_\nu(i) - S(q^*_\nu(i + 1)))\} - \{C(q^*_\nu(i), \theta_i) - C(q^*_\nu(i + 1), \theta_i)\}.
\] (6.2)

• If \( q^*_\nu(i) = q^*_\nu(i + 1) \), by the above equation, \( P_i(\nu) \) is equal to \( P_i(\nu + 1) \).
• If \( q^*_v(i) > q^*_v(i+1) \), by the concavity of \( S \),
\[
S(q^*_v(i)) - S(q^*_v(i+1)) > S'(q^*_v(i)) \cdot (q^*_v(i) - q^*_v(i+1)).
\]
Also, by the convexity of \( C(\cdot, \theta_i) \),
\[
-C(q^*_v(i), \theta_i) - C(q^*_v(i+1), \theta_i) > -C(q^*_v(i), \theta_i) \cdot (q^*_v(i) - q^*_v(i+1)).
\]

Then, by the third part of proposition 2.2, the following inequality holds:
\[
P_v(i) - P_v(i+1) > (S'(q^*_v(i)) - C(q^*_v(i), \theta_i)) \cdot (q^*_v(i) - q^*_v(i+1))
\geq \frac{V_{i-1}}{\nu_i} \Phi(q^*_v(i), i) \cdot (q^*_v(i) - q^*_v(i+1)) > 0.
\]

\[\square\]

6.4. Proof of Proposition 2.5

Proof. From (2.2), (2.3) and the property of cost function \( C \),
\[
S'(\bar{q}_v(i)) < C(q_v(i, \theta_{i+1}) + \frac{V_{i}}{\nu_{i+1}} \cdot \Phi(q_v(i, i+1)).
\]

Consider a function
\[
F(q) = S'(q) - C(q, \theta_{i+1}) - \frac{V_{i}}{\nu_{i+1}} \cdot \Phi(q, i+1).
\]

Then, by \( S'' < 0, C_{qq} > 0 \) and \( \Phi_{qq} > 0 \), it can be derived that
\[
F'' < 0
\]
and \( F(\bar{q}_v(i)) < 0. \) Since \( F(\bar{q}_v(i+1)) \) is equal to zero, the above results imply that \( \bar{q}_v(i) > \bar{q}_v(i+1). \)

\[\square\]

6.5. Proof of Proposition 2.7

Proof. From \( \frac{\partial P_v(i)}{\partial \nu_i} = P_v(i) + \sum_{j=1}^{n} \nu_j \cdot \frac{\partial P_v(j)}{\partial \nu_i} \), it is sufficient to show that
\[
\sum_{j=1}^{n} \nu_j \cdot \frac{\partial P_v(j)}{\partial \nu_i} = 0.
\]

By using the result of proposition 2.1 and the definition of \( P_v(j) \), the following equation holds:
\[
\frac{\partial P_v(j)}{\partial \nu_i} = \left\{ S'(q^*_v(j)) - C(q^*_v(j), \theta_j) \right\} \cdot \frac{\partial q^*_v(j)}{\partial \nu_i} - \sum_{l=j+1}^{n} \Phi(q^*_v(l), l) \cdot \frac{\partial q^*_v(l)}{\partial \nu_i}.
\]
From the above equation and
\[
\sum_{j=1}^{n} \nu_j \cdot \left\{ \sum_{l=j+1}^{n} \Phi_q(q^*_v(l), l) \cdot \frac{\partial q^*_v(l)}{\partial \nu_i} \right\} = \sum_{j=1}^{n} \nu_{j-1} \cdot \Phi_q(q^*_v(j), j) \cdot \frac{\partial q^*_v(j)}{\partial \nu_i},
\]
it can be derived that
\[
\sum_{j=1}^{n} \nu_i \cdot \frac{\partial P_v(j)}{\partial \nu_i} = \sum_{j=1}^{n} N_j \cdot \frac{\partial q^*_v(j)}{\partial \nu_i}, \tag{6.3}
\]
where \(N_j = \nu_j \cdot \left\{ S'(q^*_v(j)) - C_q(q^*_v(j), \theta_j) \right\} - \nu_{j-1} \cdot \Phi_q(q^*_v(j), j)\).

By the result of proposition 2.2, the following facts can be derived:

- When \(\theta_j \notin \bigcup_{k=1}^{K} I_k\), since \(q^*_v(j) = \bar{q}_v(j)\), \(N_j = \nu_j \cdot S'(q^*_v(j)) - \nu_j \cdot C_q(q^*_v(j), \theta_j) - \nu_{j-1} \cdot \Phi_q(q^*_v(j), j) = 0\) from the definition of \(\bar{q}_v(j)\).

- For all \(\theta_j, \theta_l \in I_k\), \(q^*_v(j) = q^*_v(l) = q^k\) and the minute change of \(\nu_i\) will not change the fact that they are equal. Thus, \(\frac{\partial q^*_v(j)}{\partial \nu_i} = \frac{\partial q^*_v(l)}{\partial \nu_i} = \frac{\partial q^k}{\partial \nu_i}\). Then, \(\sum_{j=a_k}^{b_k} N_j \cdot \frac{\partial q^*_v(j)}{\partial \nu_i} = \left(\sum_{j=a_k}^{b_k} N_j\right) \cdot \frac{\partial q^k}{\partial \nu_i} = 0\), by the second part of proposition 2.2.

Then, (6.3) can be decomposed as follows:
\[
\sum_{\theta_j \notin \bigcup_{k=1}^{K} I_k} N_j \cdot \frac{\partial q^*_v(j)}{\partial \nu_i} + \sum_{k=1}^{K} \left\{ \left(\sum_{j=a_k}^{b_k} N_j\right) \cdot \frac{\partial q^k}{\partial \nu_i} \right\},
\]
and it is equal to 0, by the above facts. \(\square\)


