

Reiffen, David; Robe, Michel A.

**Working Paper**

## Demutualization and enforcement incentives at self-regulatory financial exchanges

CFS Working Paper, No. 2008/44

**Provided in Cooperation with:**

Center for Financial Studies (CFS), Goethe University Frankfurt

*Suggested Citation:* Reiffen, David; Robe, Michel A. (2008) : Demutualization and enforcement incentives at self-regulatory financial exchanges, CFS Working Paper, No. 2008/44, Goethe University Frankfurt, Center for Financial Studies (CFS), Frankfurt a. M., <https://nbn-resolving.de/urn:nbn:de:hebis:30-62239>

This Version is available at:

<https://hdl.handle.net/10419/43223>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



---

No. 2008/44

**Demutualization and Enforcement Incentives  
at Self-regulatory Financial Exchanges**

David Reiffen and Michel Robe

---





## Center for Financial Studies

The *Center for Financial Studies* is a nonprofit research organization, supported by an association of more than 120 banks, insurance companies, industrial corporations and public institutions. Established in 1968 and closely affiliated with the University of Frankfurt, it provides a strong link between the financial community and academia.

The CFS Working Paper Series presents the result of scientific research on selected topics in the field of money, banking and finance. The authors were either participants in the Center's Research Fellow Program or members of one of the Center's Research Projects.

If you would like to know more about the *Center for Financial Studies*, please let us know of your interest.

Prof. Dr. Jan Pieter Krahen

Prof. Volker Wieland, Ph.D.



CFS Working Paper No. 2008/44

## **Demutualization and Enforcement Incentives at Self-regulatory Financial Exchanges\***

David Reiffen<sup>1</sup> and Michel Robe<sup>2</sup>

May 18, 2008

### **Abstract:**

In the last few years, many of the world's largest financial exchanges have converted from mutual, not-for-profit organizations to publicly-traded, for-profit firms. In most cases, these exchanges have substantial responsibilities with respect to enforcing various regulations that protect investors from dishonest agents. We examine how the incentives to enforce such regulations change as an exchange converts from mutual to for-profit status. In contrast to oft-stated concerns, we find that, in many circumstances, an exchange that maximizes shareholder (rather than member) income has a greater incentive to aggressively enforce these types of regulations.

**JEL Classification:** G28, D02, K23

**Keywords:** Demutualization, Ownership Structure, Regulation of Financial Institutions, Enforcement Delegation, Customer Protection Rules.

\* We are grateful to Jiro Kondo for helpful discussions and suggestions. We also thank Jennifer Elliott, Jo Grammig, Robert Hauswald, Jim Moser, Mike Penick, Ria Steiger, and participants to seminars at the SEC, the CFTC, the IMF, American University, the 2008 Meeting of the American Law and Economics Association at Columbia University, the 2007 Meeting of the European Finance Association in Ljubljana, the 2007 Conference on Institutional Foundations for Industry Self-Regulation at Harvard University, the 2006 Meeting of the European Association for Law and Economics in Madrid, and the First RS-DeGroot Conference on Market Structure and Integrity in Toronto, for useful comments. This paper reflects the opinions of its authors only, and not those of the CFTC, the Commissioners, or any of the authors' colleagues upon the Commission staff. All errors and omissions, if any, are the authors' sole responsibility.

1 Corresponding author: U.S. Commodity Futures Trading Commission (CFTC), 1155 21st Street, NW, Washington, DC 20581; E-mail: dreiffen@cftc.gov; Telephone: (+1) 202-418-5602.

2 U.S. CFTC and Kogod School of Business, American University, 4400 Massachusetts Avenue NW, Washington, DC 20816-8044, USA; E-mail: mrobe@american.edu

“[T]he profit motive of a shareholder-owned SRO (self-regulatory organization) could detract from proper self-regulation. For instance, shareholder-owned SROs may commit insufficient funds to regulatory operations.”

*U.S. Securities and Exchange Commission, SEC Concept Release, Fall 2004*

“(W)hen operated by a management team whose main goal is to create a profit, an exchange may have less interest in devoting resources to its regulatory functions.”

*IMF Financial Sector Assessment: A Handbook, 2005*

“Might a for-profit, publicly-traded SRO attempt to attract volume or increase its profits through lax self-regulation?”

*U.S. Commodity Futures Trading Commission, 2005 Request for Comments*

## **I. Introduction**

Following the example of the Stockholm Stock Exchange in the early 1990’s, most of the world’s major financial exchanges have converted from mutual, not-for-profit organizations to publicly-traded, for-profit firms. Since 2000, institutions such as the Deutsche Börse, the London, Tokyo, Hong Kong and Toronto Stock Exchanges, and the Sydney Futures Exchange have demutualized. In the United States, the two largest stock markets (the New York Stock Exchange and NASDAQ) and the three main futures exchanges – the Chicago Mercantile Exchange (CME), the Chicago Board of Trade (CBOT), and the New York Mercantile Exchange (NYMEX) – have all adopted the for-profit form.<sup>1</sup>

In most cases, the mutual exchanges had substantial self-regulatory (SR) authority. Significantly for investors (i.e., for the customers whose trades are executed on exchanges), these exchanges had legal authority to establish and enforce a variety of rules governing the behavior of economic agents interacting at the exchanges. While some of the newly demutualized entities have established independent subsidiaries for regulatory operations or even completely outsourced them,<sup>2</sup> many for-profit exchanges have retained these self-regulatory responsibilities.

As the pace of demutualization has accelerated, concerns have grown that for-profit exchanges might neglect their self-regulatory responsibilities. In particular, because enforcement activities are costly, “self-enforcement” could become “too little enforcement” if demutualized exchanges commit insufficient resources to regulatory operations in a bid to maximize profits.

---

<sup>1</sup> The CBOT and CME merged in July 2007, several years after each had demutualized and gone public. In March 2008, the NYMEX announced that it had agreed to be acquired by the CME.

<sup>2</sup> For example, NASD Regulation oversees and regulates all trading on NASDAQ and in the over-the-counter (OTC) markets, as well as trades in New York Stock Exchange- and Amex-listed securities reported to NASDAQ. Late in 2003, the National Futures Association entered into a “regulatory services agreement” with Eurex US to provide market surveillance and trade practice surveillance services.

Even if an exchange contracts out these duties to a subsidiary or third party, the same basic fear remains that the exchange may have incentives to under-fund the enforcement activities of the contracted party. This concern is articulated in documents released by agencies such as the U.S. Securities and Exchange Commission (SEC) and Commodity Futures Trading Commission (CFTC), the International Monetary Fund (IMF), and the International Organization of Securities Commissions (IOSCO). It is also found in statements of many academic commentators on the effects of demutualization (e.g., Karmel, 2002; Macey and O’Hara, 2005).

In this paper, we examine the relationship between self-regulation and SRO ownership structure. Precisely, we look at the enforcement of “trade practice regulations,” i.e., of the rules governing how the end-investors’ agents (stock specialists, dealers, futures commission merchants, etc.) carry out their customers’ trades.

We start from the observation that market surveillance and enforcement activities exist *because* investors’ agents have incentives to misbehave. If investors are aware of these incentives, they will need reassurance that the exchanges where their trades are carried out adequately monitor agents and enforce penalties for wrongdoing. Hence, an exchange that cuts surveillance and enforcement expenditures runs the risk that investors will decline to trade on that exchange. In other words, cost-cutting could itself be costly.

To capture these features, we analyze whether for-profit exchanges have greater or lesser incentives to enforce trade practice regulations in the context of a model in which agents have better information than do their customers about the outcomes of trades. Agents can exploit that information to their advantage. The exchange can investigate suspected misrepresentations and mete out penalties to the wrongdoers it identifies, but such monitoring is costly. This type of costly state verification model (CSV) has been used to evaluate principal-agent problems in the contexts of debt contracts,<sup>3</sup> various agency relationships, and in self-regulatory organizations (DeMarzo, Fishman & Hagerty, 2005 – hereafter, DFH). We extend this prior work in several ways. First, we model SRO decision-making under alternative ownership structures, which allows us to assess how enforcement policies differ across ownership structures. Second, we consider the effects of agent wealth and, more importantly, agent heterogeneity. This extension, which captures the reality that exchange members differ, allows us to model cases in which some agents commit trade practice violations in equilibrium. Finally, we analyze the interaction of

---

<sup>3</sup> See, e.g., Townsend (1979), Gale & Hellwig (1985), Boyd & Smith (1994), and references cited therein.

these effects – i.e., to the extent that violations take place, whether their frequency differs across ownership structures.

Our principal finding is that, contrary to oft-expressed fears, for-profit SROs have greater incentives to enforce trade practice rules than do mutual SROs. Intuitively, the goal of a mutual SRO is to maximize agent (i.e., member) income, and so it adopts an enforcement policy that creates positive incentives for agents to report honestly (carrots). By contrast, a for-profit SRO is less interested in agent income, and as a result it relies to a greater extent on punishment for dishonesty (sticks) in the form of a greater likelihood of investigating agents to insure honest reporting. Our conclusion that demutualization need not lead to laxer enforcement of customer protection rules complements theoretical demonstrations that, also contrary to widespread concerns, futures exchange competition need not lead to a “race to the bottom” with respect to margin requirements (Santos and Scheinkman, 2001) and, similarly, that merging stock exchanges need not always adopt “lowest common denominator” standards for listing companies (Chemmanur and Fulghieri, 2006).

Our model generates additional implications for the enforcement of trade practice rules. We show that greater agent wealth allows SROs (whether mutual or for-profit) to reduce the frequency of investigations without inducing misreporting. What is more important, in the extension of the model in which agents are heterogeneous, we show that, *ceteris paribus*, an equilibrium in which some misreporting is tolerated is more likely when the SRO is a mutual exchange. This result suggests an additional dimension in which demutualization might lead to a more, rather than less, rigorous trading environment.

Two qualifications to our results are worth noting. First, our results imply that demutualized exchanges should earn lower aggregate profit than mutual exchanges, since increased enforcement expenditures reduce profits (holding the number of violations fixed). Therefore, it must be that the decision to demutualize in the first place is the result of economic considerations that are unrelated to the enforcement of trade practice rules. Indeed, previous literature suggests that the recent demutualization wave can be rationalized as a means of responding to structural market changes when exchange members differ in their objectives or in their abilities to adapt to these changes (e.g., Hart and Moore, 1996; Pirrong, 2000; Steil, 2002). In other words, our analysis implicitly assumes that the increased cost of enforcing trade practice

rules are small compared to other gains from demutualization (so that demutualization is *ex-ante* profitable, even when the higher enforcement costs are rationally anticipated).<sup>4</sup>

Second, trade practice rules are only one aspect of the rules enforced by self-regulatory exchanges. Other kinds of regulations include standards for listing companies; rules meant to ensure the integrity of market prices, such as bans on insider trading and price manipulations; and statutes designed to guard the financial integrity of the exchange or clearing house, such as margin rules and member net-worth requirements. Nevertheless, a significant portion of enforcement activities at many financial exchanges stems from regulations related to the behavior of exchange members with respect to their customers, such as rules against front-running, wash trading, and bucketing. This is particularly true of U.S. futures exchanges, which have no analogue to the listing requirements or insider trading statutes that apply to U.S. (and many international) securities exchanges. To be sure, rules such as bans on price manipulation and member net-worth requirements do apply to these futures exchanges. As an empirical matter, however, trade practice violations have generated the majority of enforcement cases at U.S. futures exchanges over the past decade.<sup>5</sup> Thus, our focus on trade practice rules reflects a key aspect of self-regulatory activity at some of the world's largest financial exchanges.

The remainder of the paper proceeds as follows. Section II summarizes our contribution to the literature. Section III outlines our stylized model of self-regulating organizations (SROs). Section IV characterizes the optimal strategies for customers, agents and the SRO, when all agents are homogenous. Section V introduces agent heterogeneity. Section VI considers two other extensions of the basic model, including relaxing liability limits. Section VII concludes.

## **II. Related Work**

Our paper is part of a sizeable body of research on organizational ownership structure. More specifically, it contributes to the literature that focuses on the causes and effects of the choice between the open-stock and mutual forms of organization by financial institutions such as

---

<sup>4</sup> To wit, Hasan et al. (2003) provide evidence that demutualized exchanges use resources more efficiently. The resulting savings could offset the losses due to higher enforcement costs arising from demutualization.

<sup>5</sup> Price manipulation cases at U.S. futures exchanges have typically been handled by the government rather than by these exchanges, which may reflect the government's greater ability to handle such cases.



thrifts, insurance companies, or exchanges. Some of that work is theoretical;<sup>6</sup> much is empirical.<sup>7</sup>

Because financial exchanges play a major role in market economies, the demutualization trend has generated a substantial amount of discussion in law reviews<sup>8</sup> as well as in policy circles.<sup>9</sup> In the present paper, we provide the first formal model of the implications of financial exchange demutualization on enforcement activities. We show that maximizing the income of an exchange's shareholders need not conflict with – in fact, depends on – the exchange's vigorously enforcing trade practice rules.

Our endeavor is also related to the (theoretical) economics literature that analyzes the organizational choices of financial exchanges, as well as to the (empirical) finance literature that assesses the consequences of exchange demutualization.

On the theoretical side, Hart and Moore (1996, 1998) and Pirrong (2000) analyze an exchange's choice between for-profit stock ownership and mutual organization. In these models, trader heterogeneity in terms of size (Hart and Moore) or efficiency (Pirrong) is the main driver of ownership structure choice.<sup>10</sup> Unlike those papers, our focus is not on a financial exchange's decision to demutualize in the first place. Rather, we take a self-regulatory exchange's decision to demutualize as a given, and then investigate the impact of the resulting separation between trading rights and ownership rights on enforcement incentives.

---

<sup>6</sup> Fama and Jensen (1983a, 1983b, 1985) provide general analyses of organizational choice with specific discussions of financial mutuals. In the case of insurance companies, see e.g. Mayers and Smith (1981, 1988); Smith and Stutzer (1990a, 1995); Doherty (1991); Doherty and Dionne (1993); and Ligon and Thistle (2005). There is relatively less work on organizational choices by thrifts (Esty, 1997a) or mutual financing companies (Smith and Stutzer, 1990b).

<sup>7</sup> A number of empirical studies look at the reasons for (de)mutualization at thrifts (Masulis, 1987; Mester, 1991) and at insurance firms (Fitzgerald, 1973; Mayers and Smith, 1986, 1994, 2005; Viswanathan and Cummins, 2003; Viswanathan, 2006). Studies of the aftermath of demutualization by insurance companies show that it leads to better product-market and financial performance (McNamara and Rhee, 1992) but also greater risk-taking (Lamm-Tennant and Starks, 1993) and increased reliance on trials rather than settlements to manage customer claims (Kerr, 2005). In the case of thrifts, demutualization improves financial performance (Cole and Mehran, 1998) but also leads to greater risk-taking (Cordell, MacDonald and Wohar, 1993; Esty, 1997a, 1997b).

<sup>8</sup> See, e.g., Karmel (2002) and Fleckner (2006) on demutualization, and Mahoney (1997) and Macey and O'Hara (2005) on the respective roles of government and SROs amidst changes in exchanges' competitive environment.

<sup>9</sup> For regulatory and economic policy papers on the issues raised by demutualization and other structural changes in the environment in which exchanges operate, see e.g. Elliott (2002); Claessens, Lee and Zechner (2003); Lee (2003); Carson (2003); Chaddad and Cook (2004); and, references cited therein. See also SEC (2004), CFTC (2005), IMF (2005), and IOSCO (2006). For an international perspective, see Akhtar (2002).

<sup>10</sup> See Steil (2002, pp. 68-70) for a comparison of these two models and a discussion of the relevant issues. We are aware of only two other theoretical papers on exchange structure choice. Pirrong (1999) presents a theory of the organization of financial exchange markets that rationalizes the choice of mutual structure as a means for exchange members to earn rents, and then tests his model's predictions about the resulting equilibrium market structure. That paper, however, does not analyze the issue of exchange demutualization. In a recent working paper on the causes of demutualization, Serifsoy and Tyrell (2006) build a dynamic model of the impact of competition between exchanges on the fragility of mutual vs. stock forms of ownership.

On the empirical side, several recent studies analyze the consequences of exchange demutualization. Using trading data for 40 stocks listed on two Indian exchanges in the 1990s, Krishnamurti, Sequeira & Fangjian (2003) conclude that demutualization lowers transactions costs. Treptow (2006) analyzes the effect of demutualization on trading volume and spreads. Using data for 156 stocks that were dual listed on the New York Stock Exchange and 12 non-U.S. exchanges, that author studies whether the market share of a home exchange rises (relative to the NYSE's share of the total trading volume) after that exchange demutualizes. He finds evidence that demutualization brings about higher volume, as well as lower bid-ask spreads. In a similar vein, we focus on an outcome of demutualization that matters for the investors who trade on the exchange.<sup>11</sup> Rather than an empirical study of market liquidity or spreads, however, we carry out a theoretical analysis of the implications of ownership structure for the vigor with which the exchange will enforce customer protection rules.

### **III. A Model of Self-Regulation in Financial Markets**

Our goal is to model the decisions of agents in financial markets regarding whether to honestly represent the interests of their clients, and the impact of an exchange's objectives (i.e., its ownership structure) on these decisions and responses. Specifically, we consider agents who carry out their clients' wishes to trade on organized exchanges. In doing so, we model clients and exchanges as rationally anticipating the behavior of agents given the reward schedule. In the United States and elsewhere, many of these exchanges are self-regulatory organizations, whereby the exchanges enforce rules about permissible trading behavior by customers' agents.<sup>12</sup>

On U.S. exchanges, for instance, federal regulations prohibit a variety of trade practices such as "front-running," "non-competitive trading," "changing prices," etc. These regulations

---

<sup>11</sup> In case studies of demutualization by the London Stock Exchange and Borsa Italiana, Hazarika (2005) concludes that demutualization is associated with increased order flow but that its impact on trading costs varies depending on the cause of demutualization (reaction to a more competitive environment vs. government mandate). Other empirical studies of demutualized exchanges either compare their post-IPO financial performance with that of other firms, or compare demutualized stock exchanges pre- and post-IPO operating performance and value enhancement (O'Hara and Mendiola, 2003; Agarwal and Dahiya, 2006; Serifsoy, 2007). As these papers point out, of course, any evidence of changes in performance should be seen as preliminary because demutualization is a recent phenomenon.

<sup>12</sup> In many countries, a government agency (such as the SEC or the CFTC in the United States) has oversight of the SRO's enforcement practices, including the ability to conduct additional inspections and to potentially sanction the SRO for failing to enforce rules. The potential for government intervention, which can bring about a more stringent enforcement policy by a mutual SRO (see DFH), does not affect our main conclusions.

are collectively known as “trade practice” rules.<sup>13</sup> The element common to the prohibited practices is that they allow the agent to misrepresent the best available price, to the detriment of the customer. For example, front-running a trade for a client who wants to establish a long position can result in the customer’s paying a higher price to establish that position than if the customer’s trade had been made prior to the agent’s trade.<sup>14</sup>

### ***A. Model Overview***

We employ a stylized model of this environment to evaluate how investors respond to the potential for dishonest agents and how the SRO chooses an enforcement policy. There are three kinds of parties in this model: investors, i.e., customers; agents, who conduct trades on behalf of these investors; and the exchange (SRO) on which the trading takes place. Following DFH, we focus on a single exchange. A customer does not route a trade to this exchange unless she expects to at least achieve an exogenous reservation utility level,  $\alpha$ . One could interpret  $\alpha$  as the customer’s expected utility from transacting on an alternative trading platform, with greater competition across trading venues captured by a larger  $\alpha$ .

We assume that there are many agents competing for each customer’s business on the exchange, so that customers have all the bargaining power when contracting with agents. In this environment, a customer hires an agent to carry out a trade by offering a contract that maximizes the customer’s surplus from trading, subject to the agent’s receiving non-negative profits. At the same time, however, policies set by the exchange influence the share of the overall surplus from trading that each category of participant (customer, agent, exchange) can expect to earn.

We posit that the best price an agent can get when carrying out a customer’s trade is not observable by the agent’s customer. Instead, the customer only receives the agent’s report of the cash-flow (net of the agent’s trading costs) generated by the trade. We assume that the true cash-flow  $W$  takes on one of two values:  $w_2$  with probability  $\pi_2$ , or  $w_1$  with probability  $\pi_1$ , where  $w_2 >$

---

<sup>13</sup> “Front-running” refers to trades made by the agent (broker or futures commission merchant) on the same side of a market prior to executing an order that has already been placed by a client. “Non-competitive” or fictitious trading includes wash trading, bucketing, or other schemes that give an appearance of trading when no *bona fide* competitive trading has occurred. “Changing prices” refers to an agent’s misrepresenting the actual trading price. See, e.g., Johnson and Hazen (2004) for a detailed discussion of trade practice rules. See also Cumming and Johan (2006) for an empirical investigation of surveillance activities in 25 of the world’s largest financial exchanges.

<sup>14</sup> For trade practice violations, the benefit to the exchange of enforcing the rules is internalized. In contrast, the enforcement of other kinds of rules (e.g., anti-manipulation) can have positive externalities to rival exchanges. Macey and O’Hara (2005) articulate concerns about an exchange’s incentive to enforce rules benefiting its rivals.

$w_1$ , greater meaning that it is more advantageous for the customer. For example, suppose that the customer is selling a share of stock. Then  $w_2$  is the high realization of the price received, and  $w_1$  is the low realization. This captures the general notion of unobservable states of the world in a tractable model. To avoid trivial cases, we assume that  $w_1 < \alpha < w_2$  – i.e., the customer is willing to trade, but only if she expects to receive a payment above  $w_1$ .

The SRO oversees agent behavior, in the sense that it observes agents’ reports of the realized states, can choose to investigate (at a cost) whether a given report was accurate and, if it was not, the size of the fine  $X$  to impose on the reporting agent. We assume that agents have limited liability so that the fine cannot exceed an agent's resources, i.e., the sum of the ill-gotten gains plus the agent's pre-trade wealth,  $\gamma$ .<sup>15</sup> Any fine thus levied on an agent is paid to that agent’s injured customer.

### B. Timing

**Figure 1: Sequence of Events**

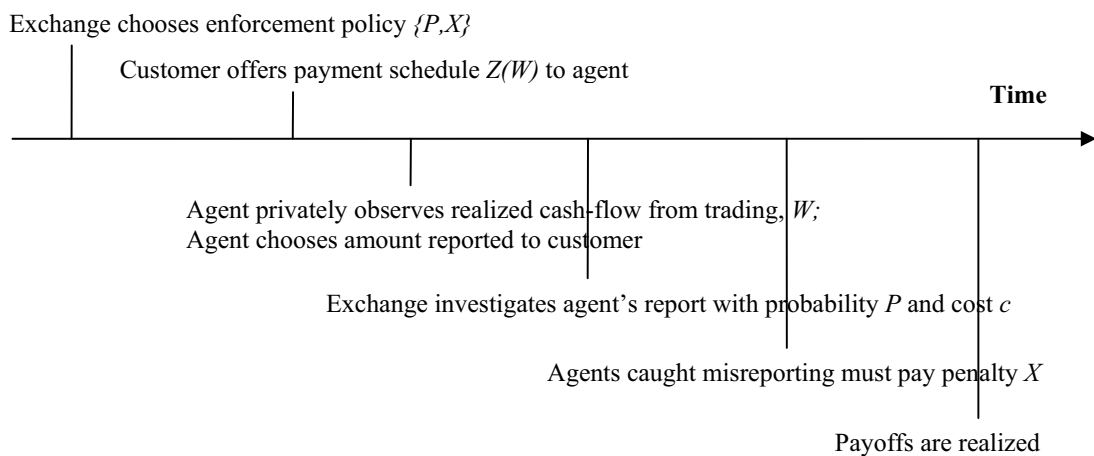


Figure 1 depicts the sequence of events in the model. The SRO first sets its enforcement policy  $\{P, X\}$  – where  $P$  stands for the probabilities  $p_i$  ( $i=1,2$ ) that the SRO will review a transaction that is reported to be in state  $i$  while  $X$  stands for the penalties  $x_{ij}$  ( $i,j=1,2$ ) to be meted out if the agent reports state  $w_j$  and the SRO determines that the true state was  $w_i$ . The customer

<sup>15</sup> In a dynamic environment, the SRO would also have the option of suspending the agent’s trading privileges, which has an effect similar to that of a fine. The static framework adopted here captures the essence of the trade-off inherent in the agent’s decision of whether to honestly report the realized state. In the dynamic environment,  $\gamma$  could represent the future earnings loss to a trader whose trading privileges are suspended.

then chooses with which agent to trade and the terms of the contract, given  $\{P, X\}$ . In this context, the contract consists of a schedule specifying the state-contingent transfer that the agent will make to the customer,  $Z(W)$  and thus the agent's fee,  $W-Z(W)$ . Given the enforcement parameters and the fee schedule, the agent decides whether to accept the customer's contract. If the agent chooses to accept the contract, he then executes the transaction, chooses a report to make to the customer about the transaction, and makes the associated transfer to the customer,  $z(w_i)$ . Finally, the SRO has the opportunity to examine the agent's report and to impose the promised penalty in case of misreporting.

### ***C. The Exchange's Objective***

The SRO sets its enforcement policy to maximize its own objective function, in anticipation of the behavior of customers and agents. A key goal of our analysis is to compare the enforcement policy of a not-for-profit, mutual exchange with that of a for-profit, demutualized exchange. To do so, we take the self-regulatory exchange's decision to demutualize as a given, and then investigate the implications of that decision on the SRO's optimal enforcement policy.

We posit, in line with DFH, that a mutual SRO seeks to maximize agent income (subject to customers' expecting to receive their reservation utility  $\alpha$ ) and sets its transaction fee  $t_{NFP}$  to cover its expected enforcement costs:  $t_{NFP} = (P_{NFP} \Pi)c$ , where  $P_{NFP}$  is the mutual exchange's vector of investigation probabilities,  $\Pi$  is the vector of states-of-the-world probabilities, and  $c$  is the unit investigation cost.

In contrast, we assume that a demutualized exchange seeks to maximize its shareholders' income (trading fees net of expected investigation costs), subject to customers and agents expecting to receive their respective reservation utilities. This choice of objective function:  $t_{FP} - (P_{FP} \Pi)c$  captures the concern that a for-profit exchange has incentives to curtail its enforcement expenditures. The assumption that the exchange collects revenues by charging per-unit-trade fees reflects practices at the major U.S. futures exchanges.

The next section assesses whether this fear is warranted, given that an exchange must optimize its objective function subject to the constraints it faces – in particular, the constraint that customers will not trade on that exchange unless their expected gains from trading there, which depend on its enforcement policies, are high enough.

## IV. Enforcement Policies under Alternative Ownership Structures

As noted above, we model the behavior of customers who offer fee schedules to agents representing them in the execution of transactions on an exchange. What is relevant to the customer when choosing these fees is the exchange's enforcement policy  $\{P, X\}$ , but not the reason why the exchange chose  $P$  and  $X$  (i.e., the exchange's objective function). We therefore first derive the contract that the customer optimally offers to the agent, taking  $P$  and  $X$  as given. Given the fee schedule chosen by the customer and the enforcement policy of the exchange, agents decide which message to send (e.g., which price to report), conditional on the exogenous true state. The SRO chooses its enforcement policy  $\{P, X\}$  in anticipation of the behavior of customers and agents. The remainder of this section derives the equilibrium of this game.

### *A. The Customer's Optimization*

As is standard in CSV models, a customer contracts with an agent to make a trade (e.g., sell a share of stock) and the customer cannot observe the realization of the trade. All the agents are risk-neutral and otherwise identical (we relax this assumption in Section V below). The customer, who takes the exchange's enforcement policy as given, wants to set a fee schedule that induces the agent to tell the truth about the realized trade. Specifically, the risk-neutral customer sets a schedule of fees to maximize her expected income from the trade, subject to the constraints that (i) the agent tells the truth (agent incentive compatibility constraint – AIC); (ii) the agent is better off serving the customer than not (agent individual rationality, or participation, constraint – AIR); (iii) the agent earn a non-negative return whenever he correctly reports the true state (no loss condition – NLC).

Formally, the customer's problem is to:<sup>16</sup>

$$\text{Max}_{\{z(w_1), z(w_2)\}} \pi_2 z(w_2) + \pi_1 z(w_1) - t$$

subject to

$$\text{AIC} \quad (1) \quad w_2 - z(w_2) \geq p_1(\max\{w_2 - z(w_1) - x_{12}, -\gamma\}) + (1-p_1)(w_2 - z(w_1))$$

---

<sup>16</sup> Condition (3) implies that the AIR condition (2) is redundant if  $A_\theta = 0$ , as in DFH. We impose both constraints in this Section to set up the parallel with the weaker version of condition (3) that we analyze in Section V.B. There, we consider optimal contracts in a more general model, in which condition (2) is not redundant even if  $A_\theta = 0$ . The qualitative results of the present section carry through to that alternative model.

$$\text{AIR} \quad (2) \quad \pi_1(w_1 - z(w_1)) + \pi_2(w_2 - z(w_2)) \geq A_0$$

$$\text{NLC} \quad (3) \quad w_i - z(w_i) \geq 0; i = 1, 2$$

where:  $\pi_i$  is the likelihood of state  $i$ , with  $1 > \pi_2 = 1 - \pi_1 > 0$ ;  
 $z_i \equiv z(w_i)$  is the customer's return in state  $i$  (e.g., how much the customer receives from the sale of a stock when the agent announces that the state is state  $i$ );  
 $t$  is the transaction fee charged by the SRO;  
 $w_i - z(w_i)$  is the agent's fee in state  $i$  (i.e., how much the customer pays the agent when the latter reports *state*  $i$ );  
 $p_i$  is the probability that the SRO will review a transaction that is reported to be in state  $i$  ( $i = 1, 2$ );  
 $x_{ij}$  is the penalty to the agent if he announces state  $i$ , but state  $j$  is the true state and the SRO catches the agent misreporting;  
 $A_0 \geq 0$  is the agent's opportunity cost; following DFH, we set  $A_0 = 0$ ;  
 $\gamma \geq 0$  is the agent's pre-transaction wealth.

Implicit in the statement of the AIC constraint is the fact that the agent has no incentive to misrepresent the poor outcome,  $w_1$ . Embedded in the statement of the AIR constraint is the result that the SRO will set penalties  $x_{ij} = 0$  for all  $i, j$  except perhaps for  $x_{12}$  (i.e., when the high return  $w_2$  is realized but the agent pretends the low return  $w_1$  has been realized). This penalty structure is optimal for both the mutual and for-profit SROs.<sup>17</sup> Consequently, to ensure that there are gains from trading for all  $0 \leq p_1 \leq 1$ , we assume that  $\pi_1 w_1 + \pi_2 w_2 > \alpha + \pi_1 c$ .

The customer's problem is similar to that in DFH, except that we allow the agent wealth  $\gamma \geq 0$ , rather than restrict  $\gamma$  to 0. Figure 2 illustrates graphically three of the constraints faced by the customer in this environment. The agent's incentive-compatibility constraint (1) is depicted by the upward-sloping AIC line. The line, drawn for the case where  $w_2 - z(w_1) + \gamma \geq x_{12}$ , can be written as  $z(w_2) = z(w_1) + p_1 x_{12}$ .<sup>18</sup> To ensure incentive compatibility,  $z(w_2)$  must lie on or below this line. The agent's individual rationality constraint (2) is depicted by the downward-sloping AIR line,  $z(w_2) = w_2 + \pi_1 w_1 / \pi_2 - \pi_1 z(w_1) / \pi_2$ . To ensure the agent's participation,  $z(w_2)$  must lie on or below this line. Finally, the limited liability constraint (3) is shown for  $i = 1$  by the vertical

<sup>17</sup> It is straightforward to extend arguments in DFH to show that this result also holds for the mutual SRO with the weaker version of (3) used in Section IV.A. As discussed below, the same result holds for the for-profit SRO.

<sup>18</sup> If it were optimal for the SRO to set  $x_{12} > w_2 - z(w_1) + \gamma$ , then the AIC would take the form  $z(w_2) = (1 - p_1) z(w_1) + p_1 (w_2 + \gamma)$ . In equilibrium, however, it does not matter whether the SRO selects  $x_{12} > w_2 - z(w_1) + \gamma$  or  $x_{12} = w_2 - z(w_1) + \gamma$ , because the  $z(w_i)$  chosen are the same in either case. We therefore focus here on the latter case.

NLC line at  $z(w_1) = w_1$ ;  $z(w_1)$  must lie to the left of this line. The shaded five-sided area depicts the combinations of  $z(w_1)$  and  $z(w_2)$  that meet all three constraints.

The customer's income in Figure 2 is just the mirror image of the agent's income, in that the expected aggregate income to the two parties is always  $\Sigma\pi_i w_i$  (minus the SRO fee  $t$ ). In other words, the customer's iso-income lines are parallel to the agent participation constraint (AIR). The solid AIR line represents the maximal customer income (gross of enforcement costs) that is consistent with the agent-participation constraint (2). Whether the AIR constraint is binding or not (i.e., whether maximized customer income falls short of this amount or not) depends on the parameters  $P$ ,  $X$ , and  $\gamma$  – all of which are taken as exogenous by the consumer. In any case, the customer's constrained optimization is to set  $z^*(w_1) = w_1$ . Setting instead  $z(w_1) < w_1$  would not only lower the customer's payment in state 1, but would also lower the maximum  $z(w_2)$  that is incentive-compatible – i.e., that is consistent with (1). Given this constraint, the highest income the customer can obtain is depicted by the downward sloping line going through the intersection of the incentive compatibility constraint (1) and the vertical NLC line representing the no-loss constraint (3). At this point, constraint (2) is not binding (for  $A_0 = 0$ ). Thus, as in DFH, the customer will set  $z^*(w_2, \gamma) = p_1 x_{12} + w_1$ . Lemma 1 summarizes these results:

**Lemma 1:** Given the exchange's enforcement policy (i.e., given investigation likelihood  $p_1$  and penalty for wrongdoing  $x_{12}$ ), the customer sets  $z^*(w_1) = w_1$  and  $z^*(w_2) = p_1 x_{12} + w_1$ .

### ***B. The Mutual SRO's Optimization***

Given this behavior by the customer, a mutual SRO (MSRO) seeks to maximize agent income using the  $p_i$ 's and  $x_{ij}$ 's as instruments, subject to the constraint that customers expect to earn their reservation levels of income (customer individual rationality – CIR).

Formally, the MSRO's problem (MP) is to maximize

$$\pi_2 (w_2 - z(w_2)) + \pi_1 (w_1 - z(w_1)) \quad (\text{MP})$$

with respect to the enforcement parameters (the  $x_{ij}$ 's and  $p_i$ 's) subject to the agent's truth-telling and no-loss constraints, (1) and (3), and to the customer's expecting an income of at least  $\alpha$ :

$$\text{CIR} \quad (4) \quad \pi_2 z(w_2) + \pi_1 z(w_1) - t \geq \alpha$$



where,  $t$  is the fee that the exchange charges customers per transaction. Reflecting the non-profit nature of an MSRO, we assume that  $t$  is set equal to the expected number of inspections,  $\pi_1 p_1$ , times the unit investigation cost,  $c$ .

As discussed above, the customer's choice of  $Z(W)$  reflects constraints (1) through (3). Furthermore,  $t = \pi_1 p_1 c$ . Thus, we can replace the above constraints with the single constraint:

$$\text{CIR}' \quad (4') \quad \pi_2 z^*(w_2) + \pi_1 z^*(w_1) - \pi_1 p_1 c \geq \alpha$$

where  $z^*(w_i, \gamma)$  reflects the optimized value of  $z(w_i)$  subject to constraints (1) through (3).

We have assumed, following DFH, that  $\alpha > w_1$ .<sup>19</sup> It is worth noting that this assumption imposes bounds on how large the investigation cost  $c$  can be. Specifically, the customer receives  $\pi_2 z(w_2) + \pi_1 z(w_1) - t = p_1 \pi_2 (w_2 - w_1 + \gamma) + w_1 - \pi_1 p_1 c$ , which must be greater than  $\alpha$  for the CIR to hold. Given  $\alpha > w_1$ , no solution exists unless  $c \leq (w_2 - w_1 + \gamma) \pi_2 / \pi_1$ .<sup>20</sup>

Lemma 2 derives the MSRO's optimal enforcement parameters.

**Lemma 2:** The mutual SRO sets  $p_1 \equiv P_M = \frac{\alpha - w_1}{\pi_2 (w_2 - w_1 + \gamma) - \pi_1 c}$ ;  $p_2 = 0$ ;  $x_{12} = w_2 - w_1 + \gamma$ ; and,  $x_{ij} = 0$  for all other  $i, j$ .

**Proof:** See Appendix.

Intuitively, Lemma 2 shows that the MSRO creates an enforcement environment – *via* positive fines for misreporting (which are paid to the customer) and a positive probability of detection – such that customers choose to give their agents some incentive for honesty. This SRO policy maximizes agent income, subject to the constraint that customers expect an income of  $\alpha$ . For any given expected fine (i.e., for any  $p_1 x_{12}$ ), it is optimal for the MSRO to set  $x_{12}$  as high as possible (i.e., set  $x_{12} = w_2 - w_1 + \gamma$ ) because the concomitant decrease in the probability of investigation,  $p_1$ , reduces enforcement expenditures – which, in turn, allows for higher agent fees in equilibrium. Lemma 2 thus also implies that  $\partial P_M / \partial \gamma < 0$ ; that is, higher agent wealth allows the exchange to select higher penalties, which allows it to reduce  $p_1$  while holding  $p_1 x_{12}$  fixed.

<sup>19</sup> In the trivial case where  $\alpha \leq w_1$ , it is optimal for the SRO to set  $p_1 = 0$  – in which case the agent always reports that state 1 has occurred and the customer receives  $w_1$ , regardless of which state actually occurred.

<sup>20</sup> This result strengthens the conclusion in Proposition 2 of DFH. These authors find that, when  $\gamma = 0$ , an investor-income maximizing SRO would either set  $p_1 = 0$  or  $p_1 = 1$  depending on the sign of  $c - (w_2 - w_1) \pi_2 / \pi_1$ . The analysis here implies that an SRO seeking to maximize customer surplus would set  $p_1 = 1$  when  $\gamma = 0$ .

The condition that  $\pi_2 (w_2 - w_1 + \gamma) - \pi_1 c \geq \alpha - w_1$  in Lemma 2 (which is implicit in DFH, for  $\gamma = 0$ , and hence holds for all  $\gamma < \alpha - w_1$ ) requires that there be some  $p_1$  that allows the MSRO to provide an enforcement regime in which customers can earn an income of at least  $\alpha$ . The interpretation of this inequality is that  $\alpha - w_1$  would be the customer's loss in income from trading if  $p_1$  were equal to 0 (recalling that  $z^*(w_2) = z^*(w_1) = w_1$  if  $p_1 = 0$ ). The left-hand side,  $\pi_2 (w_2 - w_1 + \gamma) - \pi_1 c$ , is the gain in customer income from a unit increase in  $p_1$ . Thus, if  $\pi_2 (w_2 - w_1 + \gamma) - \pi_1 c < \alpha - w_1$ , increases in  $p_1$  are insufficient to make up for the entire income loss.

Figure 3 shows how changing the investigation probability  $p_1$  affects the customer's expected income. At  $p_1 = p'$ , the customer can only reach the income level  $C_1$ . If  $C_1 < \alpha$ , then condition (4) is not satisfied. In order to meet the customer's participation constraint, the MSRO needs to increase  $p_1$  to a level, say  $p''$ , that enables customers to reach income level  $C_0 = \alpha$ .

Figure 4 shows the relation between the customer's expected income and the exogenous agent wealth,  $\gamma$ . As  $\gamma$  rises (for a fixed  $w_1$ ), the customer can move to higher income, from  $C_0$  to  $C_2$ . If  $p_1$  were held constant at  $p''$ , this move would reduce agent income. The MSRO therefore lowers  $p_1$  (to  $p'$ ) as  $\gamma$  increases, so that  $px$  is kept constant and the customer's expected income remains equal to  $\alpha$ . The agent's income correspondingly rises. Furthermore, because combined net income (i.e., agent income plus customer income minus enforcement costs) rises as  $p_1$  falls, the agent and thus the MSRO are strictly better off with higher  $\gamma$ .

### ***C. The Profit-maximizing SRO's Optimization***

The analysis of a mutual exchange in Section IV.B is similar to DFH, save for the minor generalization that the agent's wealth  $\gamma$  can be strictly positive. In this Section, we characterize the optimal enforcement policy of a demutualized, profit-maximizing exchange (PSRO) that earns its revenues through the transaction fee,  $t$ . Given risk-neutrality, the PSRO's problem is to

$$\text{Max}_{\{t, p, x\}} t - \Pi P c$$

subject to the same all-in-one customer individual rationality constraint (4').

Again,  $Z^*(W)$  takes into account constraints (1) to (3). As shown above,  $z^*(w_1) = w_1$  and  $z^*(w_2) = w_1 + p_1 x_{12}$ . The CIR' constraint must bind as well, otherwise  $t$  could be increased (thereby raising the PSRO's objective) without inducing customers or agents to exit. Hence:

**Lemma 3:** The for-profit SRO sets  $p_1 \equiv P_F = \frac{w_2 - w_1}{w_2 - w_1 + \gamma}$ ;  $p_2 = 0$ ;  $x_{12} = w_2 - w_1 + \gamma$ ;

$x_{ij} = 0$  for all other  $i, j$ ; and,  $t = w_1 + \pi_2 (w_2 - w_1) - a$ .

**Proof:** See Appendix.

Intuitively, the PSRO chooses values for the enforcement parameters,  $P$  and  $X$ , and the transaction fee,  $t$ , so that both agents and customers only receive their reservation values.

The PSRO sets the penalty  $x_{12}$  as high as it can given agent-liability limits (i.e.,  $x = w_2 - w_1 + \gamma$ ), as otherwise  $x_{12}$  could be increased and  $p_1$  reduced, which would lower SRO costs. It then chooses  $p_1$  so that, when customers optimally choose  $z(w_2)$  and  $z(w_1)$ , an agent's expected income is 0 (these choices of  $p_1$  and  $x_{12}$  lead to  $z^*(w_i) = w_i$ ,  $i = 1, 2$ ).

As was the case for the MSRO,  $\partial P_F / \partial \gamma < 0$ . The intuition for  $\partial P_F / \partial \gamma < 0$  is similar to that for the MSRO: higher agent wealth increases the maximum penalty that can be levied on the agent, which allows the PSRO to satisfy the AIC with a smaller investigation probability  $p_1$ . Note that, when  $\gamma = 0$ ,  $P_F = 1$ : because the agent's payment is 0 in both states, the agent would have no incentive to honestly report for any  $p < 1$ .

Lemma 3 implies that the fee  $t$  is independent of  $\gamma$  in equilibrium. Still, because the PSRO spends less on enforcement to obtain the same  $t$  as  $\gamma$  increases, its profits rise with  $\gamma$ . Hence, like the MSRO, the PSRO prefers agents to have higher wealth.

Figure 5 illustrates the PSRO's decision.  $C_0$  is the income level associated with the customer's participation constraint (4).  $A_0$  is the income level associated with the agent's participation constraint (2). As before, the AIC line  $z^*(w_2) = z^*(w_1) + p_1 (w_2 - w_1 + \gamma)$  represents the agent's incentive compatibility constraint (1), and the vertical NLC line at  $w_1$  reflects the no-loss constraint (3) in state 1. Figure 5 shows that, if the exchange sets  $p_1 = p' < P_F$ , then agent's expected income would be  $A_1 > A_0$ , and  $t$  would equal  $A_1 - C_0$ , the value that maximizes exchange profits when  $p_1 = p'$ . By increasing  $p_1$  towards  $P_F$ , the exchange makes the agent's income fall towards his reservation level  $A_0$ , and  $t$  can be increased without violating condition (4). In equilibrium, the fee  $t$  is the vertical distance between  $A_0$  and  $C_0$ , achieved when  $p_1 = P_F$ . As  $\gamma$  increases,  $x_{12}$  rises so that the AIR shifts upward, and the value of  $P_F$  that leads to the agent income level  $A_0$  falls; however, because the distance between  $A_0$  and  $C_0$  is independent of  $\gamma$  (both lines have the same slope:  $\pi_2/\pi_1$ ), the PSRO fee  $t$  remains unaffected.

In sum, the for-profit exchange uses  $p_I$  and  $t$  to extract surplus from both customer and agent. By increasing  $p_I$ , the exchange reduces the agent's surplus. Were  $t$  set equal to  $\pi_I pc$  (as was the case for mutual SROs), the customer would thus attain higher levels of expected income. Consequently, the for-profit SRO sets  $t > \pi_I pc$  to extract those rents from the customer.

#### ***D. Comparison of Ownership Structures***

The principal conclusion that follows from the foregoing analysis is that the PSRO devotes more resources to enforcement than does the MSRO. Formally, we have:

**Proposition 1:** The for-profit SRO spends at least as much on enforcement as does the mutual SRO.

*Proof:* See Appendix.

The logic behind Proposition 1 is that, because the agent's compensation equals  $E[W] - \alpha - \pi_I t P_M$  under the mutual form, there would be no revenue for the owners of the for-profit SRO if agent compensation were not lower with a for-profit SRO. With either a for-profit or a mutual SRO, rents are earned by agents only if state 2 occurs. Hence, it follows that a for-profit SRO must reduce agent compensation in the high state (state 2) and, thus, the difference between the agent's state-1 income and state-2 income must be lower at a for-profit SRO. This, in turn, implies that  $P_F$  must be higher than  $P_M$  in order to induce honest reporting.

An implication of Proposition 1 is that PSROs are generally more willing than are MSRO to invest in technology that reduces the per-unit cost of enforcement,  $c$ . For example, suppose there was an innovation that lowers the cost per inspection by  $\Delta c$  but requires an up-front outlay of  $F$ . If  $\Delta c$  is sufficiently large, then it is straightforward to show that the PRSO will be willing to pay more to acquire this new technology. The intuition follows directly from Proposition 1: for any given  $c$ , the PRSO carries out more inspections than does the MSRO and, hence, its total cost savings from lowering the per-inspection cost  $c$  are higher.

#### **V. Agent Heterogeneity**

In Section IV, customers as well as agents are assumed homogeneous in all respects. While this simplifying assumption puts the emphasis on the intuition behind our main results, it abstracts from the differences across customers and agents that characterize financial markets.

To evaluate the impact of customer heterogeneity, DFH introduce differences in investors' alternatives to trading on an exchange by allowing customers to have different reservation utilities ( $\alpha$ ). Such an extension of the basic model could be incorporated into our analysis. In that case, changes in the probability of investigation  $p$  would have welfare effects.

For the present analysis, we focus instead on differences among customers' agents. Precisely, we examine the impact of cross-agent differences in wealth or future profitability (captured by variations in  $\gamma$ ) on self-regulation at mutual versus demutualized exchanges.

There are several reasons why allowing for agent heterogeneity is especially relevant in the context of our paper. First, previous research suggests that it is a major reason why exchanges have traditionally used the mutual, not-for-profit form of organization (e.g., Hart and Moore, 1996, 1998; Pirrong, 2000). In a related vein, there is evidence that the recent wave of exchange demutualizations may be an optimal response to technological change in the presence of agent heterogeneity (e.g., Karmel, 2002; Aggarwal and Dahiya, 2006). It is therefore sensible to confirm the robustness of our main results when agents are heterogeneous. Second, there have been several hundred trade-practice violation cases on U.S. futures exchanges alone in the past decade. By showing that agent heterogeneity may lead an SRO to let some agents misreport in equilibrium, our analysis helps capture an important aspect of economic behavior at financial exchanges. Furthermore, to the extent that the propensity to allow misreporting varies across ownership structures, this extension yields another lens through which we can assess the impact of demutualization on the rigor of exchanges' enforcement activities.

### ***A. Modeling Agent Heterogeneity***

The previous section showed that, when agents are homogeneous, customer welfare is the same with mutual and for-profit exchanges. Under either form of exchange ownership, customers only receive their reservation values from trading, and there is no misreporting in equilibrium. In this Section, we present a highly stylized extension of the basic model in which agent heterogeneity can result in misreporting. We then compare the equilibrium amounts of misreporting at mutual versus for-profit SROs.

In practice, a key source of heterogeneity across agents is their wealth (or, alternatively, their productivity). To capture this fact, we let agents differ with respect to the exogenous wealth parameter,  $\gamma$ , in a way that is not readily observable by customers prior to deciding which agent

to hire. This assumption is equivalent to positing that agents' compensation schedules cannot be conditioned on wealth. In the context of the financial intermediaries we analyze, customer's uncertainty about agent wealth could result from agents' owning portfolios whose structures are not easily understood by outsiders and whose values may be subject to considerable variation. We assume that customers do know the distribution of agents' wealth. We make the simplest possible representation of such a distribution, by assuming that a fraction  $s$  of all agents have wealth  $\gamma_H$ , while the rest  $(1-s)$  have wealth  $\gamma_L < \gamma_H$ . Finally, we posit that an agent's wealth is costlessly verifiable by the exchange during investigations of possible wrongdoings.

### ***B. The customer's decision***

Knowing  $\gamma_L$ ,  $\gamma_H$  and  $s$ , customers take as given the exchange's enforcement policy  $\{P, X\}$  and transaction fee  $t$  and choose a fee schedule to maximize their expected income from trading. We abstract from the possibility of offering a menu of schedules that would lead to separation of agents by wealth. Instead, we posit that a single fee schedule must be offered to all agents.

It turns out that this one-size-fits-all schedule will guarantee participation by both agent types. In such an environment, then, the key change introduced by heterogeneity is that a fee schedule which induces an agent with wealth  $\gamma_L$  to honestly report the true state may not induce an agent with wealth  $\gamma_H$  to honestly report, and vice-versa. It is easy to see that the customer optimally chooses  $z^*(w_1) = w_1$  as in Section IV. We now show that the customer always chooses  $z^*(w_2)$  to extract as much surplus as possible from the high-wealth agents.

Our next set of results establishes that the customer should set  $z(w_2)$  so that the AIC is binding for at least one type of agent. Intuitively, the trade-off when choosing  $z(w_2)$  is that "high" values of  $z(w_2)$  (i.e., values high enough that incentive-compatibility is binding for high-wealth agents) maximize the payment from high-wealth agents but lead to misreporting by low-wealth agents – who, given limited liability, are undeterred by high penalties. As in Section IV, agents report honestly if the AIC is met. With heterogeneous agents, there are two relevant AICs:

$$\text{AIC-k} \quad (5) \quad w_2 - z(w_2) \geq p_1 \max\{w_2 - w_1 - x, -\gamma_k\} + (1-p_1)(w_2 - w_1) \quad (\text{k=L,H})$$

**Lemma 4:** AIC-H is met whenever AIC-L is met. The reverse is not true.

*Proof:* See Appendix.

Since customers will always choose the compensation schedule so that at least one AIC is met in equilibrium, Lemma 4 implies that the AIC-H constraint is always met. Given this result, we can now characterize the optimal contract:

**Lemma 5:** When agents are heterogeneous, customers set  $z^*(w_1) = w_1$  and  $z^*(w_2) = w_1 + p_1 x_{12}$

**Corollary 1:** If  $x_{12} > x_L$ , then the optimal payment schedule  $Z^*(W)$  leads to misreporting by low-wealth agents only. If  $x_{12} = x_L$ , then  $Z^*(W)$  induces truth telling by all agents.

*Proofs:* See Appendix.

Lemma 5 and its Corollary show that the SRO's choice of penalty,  $X$ , is key to whether all agents report honestly. The Lemma and its Corollary also show that, if misreporting were to occur, it is the less-wealthy agents who would misreport.

In the next three subsections, we derive and then compare enforcement policies at mutual and for-profit SROs. Throughout, we assume that the values of the model's parameters  $\Pi$ ,  $W$ ,  $\gamma_k$ ,  $a$  and  $c$  are such that it is profitable to operate the exchange, i.e., that the participation constraints of the customers and of their agents can all be met. A sufficient condition to ensure that this assumption is met is that:

$$(6) \quad w_1 + (w_2 - w_1)\pi_2 - \alpha - \pi_1 \left( \frac{w_2 - w_1}{x_L} \right) c > 0$$

### ***C. The Mutual SRO's Decision***

As in Section IV, the MSRO selects  $P$  and  $X$  to maximize the incomes of its member agents, subject to ensuring customer participation. However, whereas in Section IV agents were homogenous and member income was unequivocally defined, in this Section the heterogeneity of members raises the question of whose income the MSRO should maximize – e.g., is the median agent a low-wealth or a high-wealth individual? As emphasized in previous literature (Hart and Moore, 1998; Pirrong, 2000), agent heterogeneity may impact MSRO policy. In the case at hand, if different agents prefer different choices of enforcement variables, then the values of  $P$  and  $X$  chosen by the mutual exchange could depend on whose wealth is being maximized.

We first note that, although they may ultimately prefer different levels of the penalty  $x_{12}$  for misreporting, both types of agents at least agree on the range of possible values for  $x_{12}$ . First, all agents want the exchange to set  $x_{12} \geq x_L \equiv w_2 - w_1 + \gamma_L$ . The intuition here is similar to that in Section IV. As long as  $x_{12} < x_L$ , an increase in  $x_{12}$  accompanied by a reduction in  $p_I$  that leaves the customer's net income unchanged will raise the incomes of all agents, since it allows for a reduction in enforcement costs without inducing misreporting. Second, all agents want to set  $x_{12} \leq x_H \equiv w_2 - w_1 + \gamma_H$ . Logically, given limited liability, increasing  $x$  beyond  $x_H$  has no effect on any agent's incentives to misreport, since the most they can lose if their misreporting is detected is  $x_H$ . Lemma 6 summarizes these results and their implication for the MSRO:

**Lemma 6:** Both high- and low-wealth agents prefer (and, hence, the MSRO sets)  $x_L \leq x_{12} \leq x_H$ .

The next Lemma makes explicit that the SRO's choice of  $p_I$  depends on its choice of  $x$ .

**Lemma 7:** If  $x_{12} = x_L$ , then the mutual SRO chooses  $P_{M,1} = \frac{\alpha - w_1}{\pi_2(w_2 - w_1 + \gamma_L) - \pi_1 c}$

If  $x_{12} = x_H$ , then  $P_{M,2} = \frac{\alpha - w_1}{\pi_2(w_2 - w_1 + \gamma_L) + \pi_2 s(\gamma_H - \gamma_L) - [\pi_1 + \pi_2(1 - s)]c}$

*Proof:* See Appendix.

We know from Corollary 1 that, if the SRO sets  $x > x_L$ , then low-wealth agents misreport. In this case, as the proportion  $s$  of low-wealth individuals rises, the MSRO must adjust  $p_I$  upward in order to ensure customer participation. That is:

**Corollary 2:** As long as it is optimal for the MSRO to set  $x_{12} = x_H$ , then  $\partial P_M / \partial s < 0$ .

Setting  $x > x_L$  increases low-wealth agents' expected incomes but can reduce high-wealth agents' incomes. This observation suggests that whether misreporting is allowed to take place in equilibrium could depend on which type of agent's income the MSRO maximizes. Proposition 2 formalizes this intuition.



**Proposition 2:** The MSRO sets  $x_{I2} = x_L$  or  $x_{I2} = x_H$ . The choice between  $x_L$  and  $x_H$  is a function of the proportion  $s$  of high-wealth agents, and may also depend on whose expected income the MSRO is maximizing. In particular:

- (i) For high values of  $s$  ( $\hat{s} = \frac{\pi_2 x_L (c + \gamma_H - \gamma_L) - \pi_1 (\gamma_H - \gamma_L) c}{\pi_2 x_L (c + \gamma_H - \gamma_L)} < s < 1$ ), the MSRO sets  $x_{I2} = x_H$ ;
- (ii) For intermediate values of  $s$ , the MSRO's enforcement policy depends on the type of agent whose income it maximizes:
- $x_{I2} = x_L$  if the exchange acts on behalf of high-wealth agents;
  - $x_{I2} = x_H$  if the exchange acts on behalf of low-wealth agents;
- (iii) For low values of  $s$  (when  $s < \bar{s} = \max \left\{ \frac{c}{c + \gamma_H - \gamma_L}, \frac{(\alpha - w_1) x_H - (w_2 - w_1) (\pi_2 x_L - c)}{\pi_2 (w_2 - w_1) (\gamma_H - \gamma_L + c)} \right\}$ ), the MSRO sets  $x_{I2} = x_L$ .

**Proof:** See Appendix.

Proposition 2 implies that the general proposition that an enforcement agency will choose maximal fines (which allow for minimal enforcement expenditures for a given expected penalty) need not always apply to MSROs with heterogeneous traders. Intuitively, the higher penalty  $x_H$  allows for a reduction in the probability of inspection  $p$  but also brings about misreporting. As long as there are few low-wealth agents, this misreporting is not much of an issue, and the cost savings from reducing the frequency of inspections are relatively more important. As a result, when  $\hat{s} < s < 1$ , the usual result obtains, i.e., both types of agents prefer (and the MSRO sets)  $x = x_H$ .

Increasing  $x$  and reducing  $p$ , however, has a redistributive effect to the detriment of high-wealth agents. Conditional on  $x > x_L$  (so that misreporting occurs), expected customer income falls (and expected income for low-wealth agents is higher) as  $p$  declines (holding  $px$  fixed). This is because higher  $x$  does not result in higher payments from low-wealth agents once  $x$  exceeds  $x_L$ . This consequence of limited liability means that  $px$  must rise as  $x$  increases, in order to leave customers with incomes of at least  $\alpha$ . That is, customers must be compensated for lower expected penalties, so  $px$  must increase. This effect in turn means that high-wealth agents are effectively transferring income to low-wealth ones. For intermediate values of  $s$ , then, high-wealth agents prefer that the MSRO set  $x_{I2} = x_L$  whereas low-wealth agents prefer  $x_{I2} = x_H$ .

Finally, if low-wealth agents are too numerous ( $s < \bar{s}$ ), then *all* agents prefer the MSRO to set  $x = x_L$ . First, when  $s$  is sufficiently small, expected enforcement costs can actually be higher with  $x = x_H$  (which leads to expected costs of  $(\pi_1 + (1-s)\pi_2) c P_{M,2}$  because low-wealth agents

always report  $W = w_l$ ) than with  $x=x_L$  (where the costs are  $\pi_{IC}P_{M,l}$ ). In such a situation, low-wealth agents may actually prefer  $x=x_L$ . Alternatively, when  $s$  is small, the transfer from high-wealth to low-wealth agents due to misrepresentation may be so large that it is impossible for customers' and high-wealth agents' participation constraints to be met if there is misreporting. Choosing an  $x$  that fails to meet high-wealth agents' participation constraint is never optimal for low-wealth agents. In this case, condition (6) implies the exchange can be operated profitably – as long as  $x=x_L$ .

Proposition 2 complements the extant literature that shows how agent heterogeneity can lead to important disagreements with respect to policy at a mutual exchange. In our setting, the potential disagreement becomes less important as agent heterogeneity declines. Specifically, as  $s$  goes to 1, all agents prefer some misreporting in equilibrium, while as  $s$  goes to 0, all agents prefer low penalties but vigorous monitoring.

#### ***D. The For-Profit SRO's Decision***

As in Section IV, the for-profit SRO maximizes its profits (transaction fee minus investigation costs), subject to participation by all agents and customers. In addition, the assumption that  $w_l < \alpha$  implies that the PSRO must satisfy at least one type of agents' truth-telling constraint. That is, the PSRO seeks to

$$\text{Max}_{\{t, p, x\}} \quad t - \Pi P c$$

subject to

$$\text{AIR} \quad (7) \quad \pi_1 (w_1 - Z^*(w_1)) + \pi_2 (w_2 - Z^*(w_2)) \geq 0$$

$$\text{AIC}^*\text{-k} \quad (8) \quad w_2 - Z^*(w_2) \geq p_1 [\max \{w_2 - w_1 - x, -\gamma_k\}] + (1-p_1) (w_2 - w_1) \quad (k=L,H)$$

$$\text{CIR} \quad (9a) \quad w_1 + \pi_2 p [sx + (1-s) (w_2 - w_1 + \gamma_L)] - t \geq \alpha$$

(if 14 only holds for high-wealth agents)

$$(9b) \quad w_1 + \pi_2 px - t \geq \alpha$$

(if 14 holds for all agents)

where the fee schedule  $Z^*(W)$  reflects the customer's optimizing behavior derived in Lemma 5, and constraint (9a) reflects Lemma 4, i.e., the fact that AIC-L need not hold in equilibrium.

The PSRO's determination of  $P$ ,  $X$  and  $t$  here is similar to its approach when agents are homogeneous. In particular, the PSRO sets  $p_I$  to extract surplus from the agents (in the heterogeneous case, only high-wealth agents are necessarily driven to their reservation utility), and then sets  $t$  to extract surplus from customers. The following Lemma specifies how the PSRO's choices of investigation probability  $p_I$  and of the fee  $t$  depend on its choice of the penalty  $x$ .

**Lemma 8:** If  $x_{I2} = x_L$ , then the for-profit SRO chooses  $P_{F,1} = \frac{w_2 - w_1}{w_2 - w_1 + \gamma_L}$  and

$$\text{sets } t = w_1 + \pi_2 P_{F,1} (w_2 - w_1 + \gamma_L) - \alpha = w_1 + \pi_2 (w_2 - w_1) - \alpha$$

If  $x_{I2} = x_H$ , then the for-profit SRO chooses  $P_{F,2} = \frac{w_2 - w_1}{w_2 - w_1 + \gamma_H}$  and

$$\text{sets } t = w_1 + P_{F,2} \pi_2 (w_2 - w_1 + \gamma_L + s(\gamma_H - \gamma_L)) - \alpha.$$

*Proof:* See Appendix.

The next Proposition identifies the for-profit exchange's optimal choice of penalties.

**Proposition 3:** The PSRO either sets  $x = x_L$  or  $x = x_H$ . It sets  $x = x_H$  when the proportion  $s$  of high-wealth agents is high ( $\hat{s} < s < 1$ ), and sets  $x = x_L$  otherwise ( $0 < s < \hat{s}$ ).

*Proof:* See Appendix.  $\square$

### ***E. Comparison of Ownership Structures***

A comparison of Propositions 2 and 3 shows that the for-profit exchange's enforcement policy is stricter than its mutual counterpart's in two ways. First, there are parameter values ( $\bar{s} < s < \hat{s}$ ) for which the PSRO prevents misreporting while the MSRO may allow misreporting. The opposite is not true. That is, for any set of parameter values for which the mutual SRO prevents all misreporting, the for-profit SRO also prevents all misreporting ( $0 < s < \hat{s}$ ).

At first blush, it might appear that the extent of misreporting in our model is of little consequence, in that customers get the same *ex ante* utility whether or not misreporting occurs. However, if some misreporting does occur in equilibrium, it means that customers are *ex post* randomly made better- or worse-off by this very misreporting. Such randomness is precisely the kind of outcome that trade-practice rules are designed to eliminate.

Second, whenever the PSRO and the MSRO agree on whether to allow some ( $\hat{s} < s < 1$ ) or to prevent all ( $0 < s < \hat{s}$ ) misreporting, the investigation probability is always higher at the PSRO. That is, it is readily shown that  $P_{F,2} > P_{M,2}$  when  $\hat{s} < s < 1$  and that  $P_{F,1} > P_{M,1}$  when ( $0 < s < \hat{s}$ ).

The reason why the PSRO enforces customer protection rules more rigorously for any given fine is the same as in the homogeneous agent case. For any given  $x$ , the PSRO wants to choose a detection probability at which the AIR is just binding. That is necessarily at least as high as the detection probability that maximizes agent profits (which is the goal of the MSRO).

When the exchange sets  $x$  such that misreporting occurs, the percentage of low-wealth agents ( $s$ ) affects outcomes in a different way for the mutual and for-profit SRO. As indicated in Lemma 7, the MSRO has to increase  $p$  as  $s$  falls, in order to induce customer participation. In contrast, the PSRO reduces  $t$  when  $s$  falls in order to induce customer participation, but  $p$  is unaffected. In the former case, high-wealth agents are worse off as  $s$  rises, while in the latter case, the SRO's owners are made worse off.

Figure 6 provides a numerical example of the relationship between exchange ownership structure, detection probability  $P$ , and proportion  $s$  of high-wealth agents. When agents are overwhelmingly high-wealth types ( $s > 0.88$ ; case (ii) in Proposition 2), both the MSRO and the PSRO set  $x = x_H$ . The mutual's investigation probability  $P_{M,2} = 0.374$  at  $s = 0.88$  and is decreasing in  $s$  (Corollary 2), while the for-profit's  $P_{F,2}$  is larger ( $P_{F,2} = 0.375$ ) and independent of  $s$ . In that case, low-wealth agents misreport the good state in equilibrium. If the proportion  $s$  of high-wealth agents is "low" ( $s < 0.83$ ; case (i) in Proposition 2), then the MSRO and PSRO both set  $x = x_L$  and deter misreporting by raising the investigation probability to, respectively,  $P_{M,1} = 0.48$  or  $P_{F,1} = .49 > P_{M,1}$ . Finally, for intermediate values of  $s$  ( $0.83 < s < 0.88$ ; case (iii) in Proposition 2), the PSRO sets  $p_1 = P_{F,1} = 0.49$  whereas the MSRO either sets  $x = x_L$  and  $p = P_{M,1} = 0.48$  to benefit the high-wealth agents or, alternatively, or sets  $x = x_H$  and  $p = P_{M,2} \leq 0.374$  to benefit low-wealth agents.

## **VI. Other Extensions**

Section V established that the key result of Proposition 1 is robust to the introduction of agent heterogeneity and, by showing that misreporting is more likely when the SRO is a mutual exchange, suggested an additional dimension in which demutualization might lead to a more,

rather than less, rigorous trading environment. In this Section, we investigate two other extensions of the basic model of Section IV. In Section VI.A, we establish the robustness of our main results to the assumption that agent income cannot be negative, i.e., that  $z(w_i)$  is restricted to being less than or equal to  $w_i$  ( $i=1,2$ ). Finally, in Section VI.B, we examine the assumption that the exchange can credibly precommit to all the aspects of its enforcement policy, and discuss the potential for *ex-post* opportunistic SRO behavior with respect to enforcement activities.

#### ***A. Equilibrium with the potential for non-negative returns to truth-telling agents***

The analysis in Section IV imposes the restriction that customers choose fee schedules  $Z(W)$  such that the agent's earnings in both states is at least 0. This restriction seems consistent with actual practice. That is, in reality, it does not appear that agents are required to give customers a payment in excess of the agent's actual receipts when the agent honestly reports that receipts were "low." Still, it may be that in practice agents lose money when they report a low receipt, in that agents face positive trading costs which they do not recover in the low state.

In this subsection, we consider how the equilibrium changes when the no-loss constraint in (3) is replaced by the (weaker) no-bankruptcy constraint that the payment to the customer in any state can never be greater than the actual receipt of the agent plus the agent's wealth  $\gamma$ . With this change, the constraint in equation (2) is no longer redundant. That is, agents must earn non-negative profits in expectation, but they can lose money in any state – even when they give an honest report. We now show that allowing for this possibility allows the agents/SRO to earn more than in the case evaluated in Section IV, but that our main conclusions remain unchanged.

To analyze this case, we return to the assumption that agents are homogeneous with respect to  $\gamma$  and then revisit the behavior of customers and SROs, replacing the agent's NLC condition (3) with the agent bankruptcy constraint:

$$(ABC) \quad (3') \quad w_i - z(w_i) - x_{ij} \geq -\gamma, \quad i = 1,2$$

The agent's wealth,  $\gamma$ , plays a somewhat different role in this model. If  $\gamma > 0$ , customers can offer a contract to the agent that yields the agent negative revenues when he announces state 1. Specifically, analogously to Section IV, the customer's profit-maximizing choice of  $z(w_1)$  is the highest  $z(w_1)$  consistent with (3'), which is  $z^*(w_1) = w_1 + \gamma$  in this model. The agent's

income is thus  $-\gamma$  when he announces the true state is state 1. As above, this not only maximizes the customer's payment when state 1 occurs, but also allows for higher  $z(w_2)$ . Given  $z(w_1) = w_1 + \gamma$ , the customer will then choose  $z(w_2)$  so that the incentive compatibility constraint is binding, which means that  $w_2 - z(w_2) = p[\max\{w_2 - z(w_1) - x, -\gamma\}] + (1-p)(w_2 - z(w_1))$ . Hence, for  $x \leq w_2 - w_1$ , the binding AIC is  $z(w_2) = z(w_1) + px$ .

Figure 7 depicts the customer's decision in this alternative model. The main difference between the decision here and that portrayed in Figure 2 is that, rather than indirectly shifting the AIC curve upward as in Section IV, changes in  $\gamma$  shift the constraint in equation (3'). That is, in the present variation of the model, increasing  $\gamma$  shifts the vertical ABC line to the right, allowing higher  $z(w_1)$  for fixed  $p_1$  and  $x_{12}$ . Hence, the customer's optimum either occurs at the intersection of the AIC and the NBC (as portrayed in Figure 7), or at the intersection of the AIC and the AIR – depending on the value of  $\gamma$ . In particular, as discussed below, if  $\gamma$  is sufficiently large, then the latter intersection will be the customer's optimum in equilibrium.

Given this behavior by customers, the mutual SRO once again chooses values for  $p$  and  $x$  to maximize the agent's income. As our earlier analysis implies, the SRO chooses a maximal value for  $x$ , which in this case is  $x = w_2 - w_1$ . The logic here is again that, if  $x$  were less than  $w_2 - w_1$ , then  $x$  could be increased and  $p$  could be decreased without changing the agent's or the customer's incomes, but saving the customer enforcement costs. The reason why  $x$  is lower in this model than in the model in Section IV is that agents pay  $\gamma$  whenever they announce state 1 has occurred, which reduces the maximum penalty that the SRO can assess when the agent is discovered misreporting the true state. This means that the choice of  $z(w_2)$  that results in the incentive compatibility constraint's binding is

$$z(w_2) = p(w_1 + x) + (1-p)(w_1) = w_1 + \gamma + p(w_2 - w_1).$$

As long as  $\gamma \leq \alpha - w_1$  then the SRO will choose  $p$  so that the CIR is binding, or

$$\pi_2 z(w_2) + \pi_1 z(w_1) - t = \alpha.$$

Solving these two equations for  $p$ , and recalling that  $t = \pi_1 p_1 c$  for the mutual SRO, yields

$$p_1 = p'_M = \frac{\alpha - w_1 - \gamma}{\pi_2(w_2 - w_1) - \pi_1 c}$$

Note that, as long as  $\gamma \leq \alpha - w_1$ ,  $P'_M$  is decreasing in  $\gamma$ , as was the case for model presented in Section IV.<sup>21</sup> Here, higher agent wealth allows the customer to create greater incentives for truth-telling via the fee schedule, which in turn allows the MSRO to achieve the same degree of deterrence with a lower probability of detection.

For  $\gamma \geq \alpha - w_1$ ,  $P'_M$  will equal 0. In contrast to the MSRO's decision in the earlier model, in this model the CIR can be satisfied even when  $p = 0$ . Customers receive  $z(w_1) = \gamma + w_1$  when the agent claims the state is state 1. When  $p = 0$ , the AIC implies  $z(w_2) = z(w_1) = \gamma + w_1$ . Hence if  $\gamma + w_1 \geq \alpha$ , this payment is greater than the customer's reservation value, and the MSRO will indeed set  $P'_M = 0$ . Once  $\gamma$  is greater than  $\alpha - w_1$ , higher  $\gamma$  transfers income from the agent to the customer. However, for  $\gamma \geq \pi_2 (w_2 - w_1) \geq \alpha - w_1$ , the AIR is binding, and hence higher  $\gamma$  has no additional effect on fees.

For the for-profit SRO, the optimization problem is once again to set  $p$  and  $x$  to extract all agent surplus, and then set  $t$  to extract all of the customer's surplus. For the reasons described above, the PSRO will set  $x$  at its maximum consist with  $x \leq w_2 - z(w_1) + \gamma$ , or  $x = w_2 - w_1$ . As in Section IV, increasing  $p$  allows the PSRO to increase  $t$ . Increasing  $p$  is profitable as long as  $\partial t / \partial p > \pi_1 c$  (i.e, the change in  $t$  resulting from the higher  $p$  exceeds the marginal cost of raising  $p$ ). When  $z(w_1) = w_1 + \gamma$  and  $x = w_2 - w_1$  so that  $z(w_2) = w_1 + \gamma + p (w_2 - w_1)$ , the CIR becomes

$$(8') \quad p(w_2 - w_1) = (\alpha + t - \gamma - w_1 / \pi_2)$$

This implies that  $\partial t / \partial p = \pi_2 (w_2 - w_1)$ , and since  $\pi_2 (w_2 - w_1) > \pi_1 c$  (which is the range of values for which  $P'_M > 0$ ), the PSRO will continue to increase  $p$  until condition (2) is binding.

Once  $p$  is sufficiently high that (2) is binding, any further increase in  $p$  will require  $z^*(w_1)$  to fall with  $p$  in order to induce the agent to participate. At that point, higher  $p$  only transfers income between states for both the agent and the customer; it does not allow the PSRO to increase revenue. As such, since  $c > 0$ , increases in  $p$  are no longer profitable. At the point where the CIR is binding,  $\pi_2 (w_1 + \gamma + p(w_2 - w_1)) + \pi_1 (w_1 + \gamma) = \pi_2 w_2 + \pi_1 w_1$ , or

$$p_1 = p_F = \frac{\pi_2 (w_2 - w_1) - \gamma}{\pi_2 (w_2 - w_1)} \quad \text{and} \quad t = w_1 + \pi_2 (w_2 - w_1) - \alpha$$

---

<sup>21</sup> Note that in order for the AIR to be satisfied,  $\pi_2 (w_2 - z(w_2)) + \pi_1 (w_1 - z(w_1))$  must be positive, or equivalently,  $\pi_2 [(w_2 - w_1) - (w_2 - w_1) (\alpha - w_1 - \gamma) / (\pi_2 (w_2 - w_1) - \pi_1 c)] \geq \gamma$ .

As with the mutual SRO, the probability of detection is a function of  $\gamma$ . The probability of detection upon a report of a poor outcome by the agent is  $P'_F = 1$  if  $\gamma = 0$ , while  $P'_F$  is between 0 and 1 for  $\gamma \in (0, \pi_2(w_2 - w_1))$ . As was the case for the mutual SRO, we have  $\partial P'_F / \partial \gamma < 0$ . The intuition for  $\partial P'_F / \partial \gamma < 0$  is again that higher agent wealth increases the payment the customer receives and reduces the payment to the agent in state 1. This allows the PSRO to induce honest reporting with a lower frequency of inspection. Analogously with  $P'_M$ ,  $P'_F$  goes to 0 for  $\gamma$  sufficiently large. And, as was the case for the mutual SRO, beyond some point (here  $\gamma = \pi_2[w_2 - w_1]$ ) higher agent wealth transfers income from the SRO owners to the customers.

As in the analysis in Section IV,  $P'_F \geq P'_M$ , if  $\pi_2(w_2 - w_1) - \pi_1 c > \alpha - w_1$  (recalling that, if the inequality fails, there is no gain to trading when  $\gamma = 0$ ).

**Proposition 4:** If conditions are such that the AIR can be satisfied, then  $P'_F \geq P'_M$ , and  $P'_F > P'_M$  for  $\gamma < \alpha - w_1$ ,

*Proof:* See Appendix.

A second and related difference between mutual and for-profit SRO is that, if the SRO could choose its  $\gamma$ , the  $\gamma$  chosen by the for-profit SRO would be larger. In other words, *ceteris paribus*, a for-profit SR exchange strictly prefers wealthier agents (Futures Commission Merchants or FCMs in futures markets) than does a mutual SR exchange. There are two possible interpretations of this finding. (i) First, as noted above, there has been a shift from mutual to for-profit SROs over the past decade. One potential reason for that shift is that exogenous changes have occurred in technology or competition that have increased  $\gamma$  or  $\alpha$  (increased competition can be thought of as an increase in  $\alpha$ ). (ii) Second, if one treats  $\gamma$  as a choice variable for the SRO, then the implication is that the for-profit SRO will choose a larger  $\gamma$ . Whether this is interpreted in terms of increased agent wealth or increased agent productivity is something we turn to below.

### ***B. Observability of enforcement parameters and the role of government***

Like extant CSV models in the financial agency literature, our analysis assumes that the enforcement parameters are observable by all agents. This is an important feature of the model, in that the choice of these parameters by the SRO affects the fee schedule set by the customers.



One might be concerned that this assumption could deviate from reality in a significant way if (i) it were plausible that  $P$  and  $X$  may not be observable in reality and (ii) SROs had incentives to deviate from the announced  $P$  and  $X$  once customers have set the schedule  $Z(W)$ . In practice, while the fines chosen by an SRO are arguably transparent, it seems plausible that customers may not be able to observe the true likelihood of detection chosen by the SRO.

In the case of a mutual SRO, reducing the detection probability  $p_I$  below  $P_M$  can both increase agent income and reduce costs. That is, once the fee schedule is set and customers start trading, the MSRO could conceivably communicate its intention to lower  $p_I$  to the agents, with the results that these agents will all misreport (and earn more) and that the MSRO will save on enforcement expenditures. By contrast, once the fee schedule is set, the incentive for the for-profit SRO to reduce  $p$  only comes about through the potential to save enforcement costs. Unlike the MSRO, the PSRO does not increase its revenues by reducing  $p_I$ . If anything, this difference in the net benefit from reducing  $p_I$  at mutual versus for-profit SROs reinforces our earlier conclusion that a PSRO will have stricter enforcement policies.

Of course, several factors also mitigate an SRO's *ex-post* incentives to reduce  $p_I$ . First, for mutual SROs, saving costs by reducing  $p$  may not be to the agents' benefit, even in the short run (although the benefit of higher expected agent revenue from lower  $p$  is unambiguous). This is because the mutual, non-profit organization form can be viewed as a commitment to a certain level of enforcement expenditures. That is, with the mutual form, transaction fees are set equal to enforcement expenditures, so that members may receive no direct benefit from a reduction in enforcement costs.<sup>22</sup> And, in the case of for-profit SROs, the enforcement budget is *ex-post* observable (e.g., by reading the company's annual report) and thus deviations will harm the PSRO's reputation.<sup>23</sup>

Second, for both kinds of SROs, customers may rationally anticipate such opportunistic *ex-post* behavior. Hence, if it is indeed profitable to cut  $p_I$  once fee schedules are set and customers decide to trade, then the fee schedules  $Z(W)$  will reflect that potential. In turn, this response by customers will reduce agent and/or shareholder income. Thus, SROs will attempt to establish a reputation, whereby it is costly to them to deviate from the probabilities derived in the static environment of Sections IV, V and VI.A.

---

<sup>22</sup> Perhaps the concern expressed by various regulatory agencies that for-profit SROs will cut enforcement spending reflects the differential effect of cost savings from reduced enforcement expenditures for the two kinds of SROs.

<sup>23</sup> DeMarzo, Fishman & Hagerty (2007) formally model the effect of reputation on agents' incentives to misreport.

A formal consideration of the potential for opportunistic behavior on the part of SROs is beyond the scope of this paper. These considerations do have some implications for the role of regulatory authorities towards enforcement of rules against misreporting (such as trade-practice rules in the futures context), however. Specifically, DFH suggest that the threat of duplicative government enforcement can induce mutual SROs to increase the probability of detection. This can increase customer welfare (and in a model in which  $\alpha$  is heterogeneous, increase social welfare). In the case of for-profit SROs, that policy is less likely to be welfare-enhancing, because increases in  $p$  beyond  $P_F$  do not increase customer welfare. The potential for opportunism suggests that government policy might better be aimed at insuring that SROs provide the promised level of enforcement. As a practical matter, this intervention might consist of reviewing an SRO's budgeted resources for enforcement, and monitoring the level of actual expenditure, to insure against opportunistic behavior. In addition, since profit maximization (absent opportunism) increases enforcement expenditures, policies that lead for-profit SROs closer to the objective of maximizing shareholder income should result in greater enforcement.

## **VII. Conclusion**

This paper analyzes how a self-regulatory financial exchange (SRO) might make use of penalties and incentives to influence contracts between investors and their agents. Of particular interest is how the SRO's ownership structure (for-profit *vs.* mutual) influences its enforcement policy. Broadly speaking, we find that a for-profit SRO uses more "sticks" and fewer "carrots" to provide incentives for agents to report honestly to their clients. That is, a for-profit SRO spends more on enforcement than a mutual SRO. We also find that trade-practice violations are more likely (that is, occur for additional parameter values) when the SRO is a mutual exchange.

We have tried to capture the essence of alternative ownership structures in a tractable environment by limiting the strategy space. Nevertheless, we think that the basic effects we find should carry over to richer environments. For example, because we model a static environment, we limit the analysis to monetary fines. In a dynamic environment, the SRO would have the ability to suspend the trading privileges of an agent that violates a trade-practice rule. In that case, the agent's wealth parameter  $\gamma$  would reflect future trading profits, which could differ between a for-profit and a mutual SRO. In particular, since agents earn less when the SRO is

for-profit (and, hence, have less to lose from a trading ban), a for-profit SRO – which we find to be a stricter enforcer than its mutual counterpart – will have to rely on an *even* greater likelihood of inspection to insure against misreporting.

The role of agent heterogeneity is simplified in our model, in that we abstract from the possibility that it may affect the decision to demutualize in the first place. Still, previous work (e.g., Pirrong, 2000; Hart and Moore, 1998) has emphasized the role of heterogeneity in the choice of organizational structure. To the extent that agent heterogeneity does influence that choice, an interesting extension of our analysis of the enforcement of customer protection rules would be to integrate the endogeneity of the form of ownership structure. Another interesting venue for further work would be to analyze how demutualization affects the enforcement of other rules governing the behavior of agents at financial exchanges, such as rules against price manipulation or illegal insider trading.

We show that the mutual SRO selects the weakest intensity of monitoring that is consistent with customer participation, whereas the for-profit SRO selects the highest intensity of monitoring that is consistent with agent participation. In that sense, government mandates of stricter enforcement are unlikely to improve customer welfare at PSROs. In our stylized model, however, all transactions take place at a single exchange. Yet, transactions involving two or more exchanges have become numerous in recent years amidst a sharp increase in the volume of complex financial transactions. Another interesting avenue for further research would therefore be to analyze the role of the government in monitoring trader behavior when transactions involve two or more exchanges, and to investigate how such a role could be affected by demutualization.

Finally, it would be interesting to test our model's predictions. For example, controlling for the evolution of technology (such as monitoring improvements made possible by electronic trading) and the product market (in particular, changes in the competitive landscape), have enforcement budgets and staff grown after demutualization? Has the investigation frequency gone up? In a similar vein, could the improvement in liquidity and the drop in spreads that have been found to follow demutualization (Krishnamurti et al., 2003; Treptow, 2006) be linked to changes in enforcement incentives and practices? We leave these questions for further research.

## References.

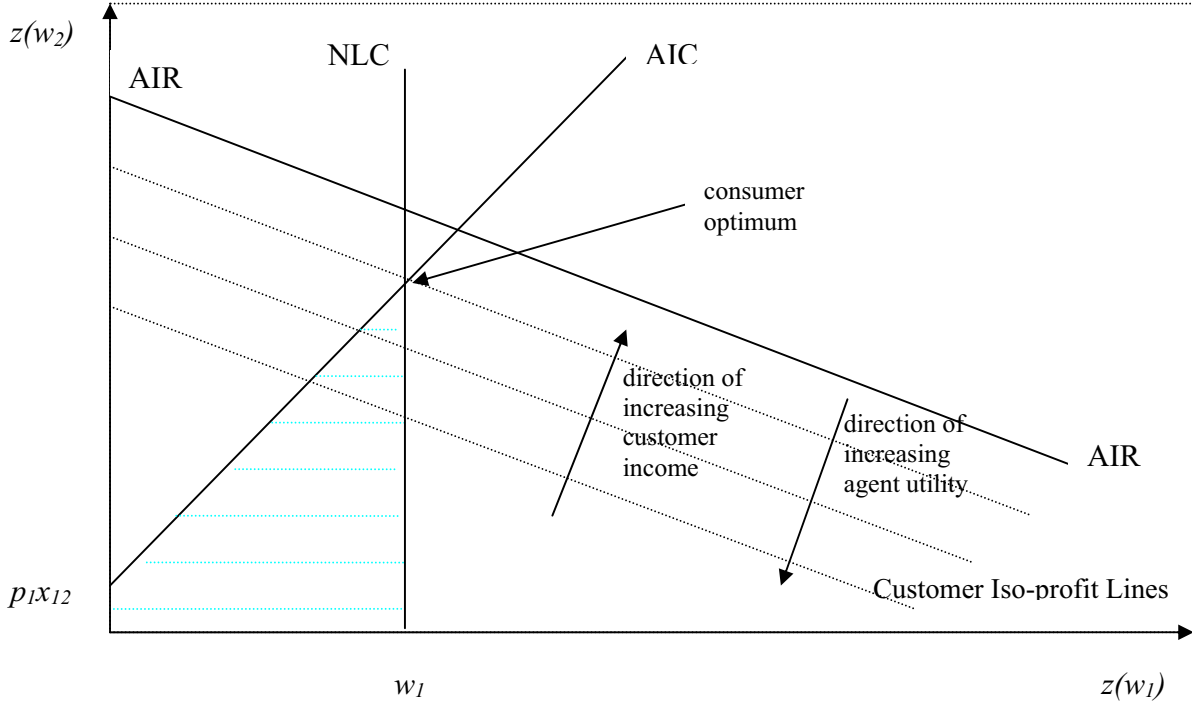
- Aggarwal, Reena and Sandeep Dahiya, 2006. "Demutualization and Public Offerings of Financial Exchanges," *Journal of Applied Corporate Finance*, 18 (3), pp. 96-106.
- Akhtar, Shamshad (Editor), 2002. *Demutualization of Stock Exchanges: Problems, Solutions and Case Studies*. Manila: Asian Development Bank.
- Boyd, John H. and Bruce D. Smith, 1994. "How Good Are Standard Debt Contracts: Stochastic versus Nonstochastic Monitoring in a Costly State Verification Environment," *Journal of Business*, 67 (4), pp. 539-61.
- CFTC, 2005. *Self-Regulation and Self-Regulatory Organizations in the Futures Industry* (Federal Register Notice, 11/25/2005), U.S. Commodity Futures Trading Commission.
- Carson, John W., 2003. "Conflicts of Interest in Self-regulation: Can Demutualized Exchanges Successfully Manage Them?" World Bank Policy Research Working Paper No. 3183, December.
- Chaddad, Fabio R. and Michael L. Cook, 2004. "The Economics of Organization Structure Changes: A U.S. Perspective on Demutualization," *Annals of Public and Cooperative Economics*, 75 (4), pp. 575-94.
- Chemmanur, Thomas J. and Paolo Fulghieri, 2006. "Competition and Cooperation Among Exchanges: A Theory of Cross-Listing and Endogenous Listing Standards," *Journal of Financial Economics*, 82 (2), pp. 455-89.
- Claessens, Stijn, Ruben Lee and Josef Zechner, 2003. *The Future of Stock Exchanges in European Union Accession Countries*, CEPR report commissioned by The Corporation of London.
- Cole, Rebel A. and Hamid Mehran, 1998. "The Effect of Changes in Ownership Structure on Performance: Evidence from the Thrift Industry," *Journal of Financial Economics*, 50 (3), pp. 291-317.
- Cordell, Lawrence R., Gregor D. MacDonald, and Mark E. Wohar, 1993. "Corporate Ownership and the Thrift Crisis," *Journal of Law and Economics*, 36 (2), pp. 719-56.
- Cumming, Douglas and Sofia Johan, 2006. "Market Surveillance Around the World," Working Paper, Rensselaer Polytechnic Institute and Tilburg University, June.
- DeMarzo, Peter M., Michael J. Fishman, and Kathleen M. Hagerty, 2005. "Self-Regulation and Government Oversight," *Review of Economic Studies*, 72 (3), pp. 687-706.
- DeMarzo, Peter M., Michael J. Fishman, and Kathleen M. Hagerty, 2007. "Reputations, Investigations, and Self-Regulation," *Mimeo*, Stanford and Northwestern Universities, April.
- Doherty, Neil A., 1991. "The Design of Insurance Contracts When Liability Rules Are Unstable," *Journal of Risk and Insurance*, 58 (2), pp. 227-46.
- Doherty, Neil A. and Georges Dionne, 1993. "Insurance with Undiversifiable Risk: Contract Structure and Organizational Form of Insurance Firms," *Journal of Risk and Uncertainty*, 6 (2), pp. 187-203.
- Elliott, Jennifer, 2002. "Demutualization of Securities Exchanges: A Regulatory Perspective," *IMF Working Paper 02/119*, International Monetary Fund, Washington, DC.
- Esty, Benjamin C., 1997a. "Organizational Form and Risk Taking in the Savings and Loan Industry," *Journal of Financial Economics*, 44 (1), pp. 25-55.
- Esty, Benjamin C., 1997b. "A Case Study of Organizational Form and Risk Shifting in the Savings and Loan Industry," *Journal of Financial Economics*, 44 (1), pp. 57-76.
- Fama, Eugene F. and Michael C. Jensen, 1983a. "Separation of Ownership and Control," *Journal of Law and Economics*, 26 (2), pp. 301-25.

- Fama, Eugene F. and Michael C. Jensen, 1983b. "Agency Problems and Residual Claims," *Journal of Law and Economics*, 26 (2), pp. 327-49.
- Fama, Eugene F. and Michael C. Jensen, 1985. "Organizational Forms and Investment Decisions," *Journal of Financial Economics*, 14 (1), pp. 101-19.
- Fitzgerald, John F., Jr., 1973. "Demutualization of Mutual Property and Liability Insurers," *Journal of Risk and Insurance*, 40 (4), pp. 575-84.
- Fleckner, Andreas, 2006. "Stock Exchanges at the Crossroads," *Fordham Law Review*, 74, pp. 2541-620.
- Gale, Douglas and Martin Hellwig, 1985. "Incentive-Compatible Debt Contracts: The One-Period Problem," *Review of Economic Studies*, 52 (4), pp. 647-63.
- Hart, Oliver and John Moore, 1996. "The Governance of Exchanges: Members' Cooperatives versus Outside Ownership," *Oxford Review of Economic Policy*, 12 (4), pp. 53-69.
- Hart, Oliver and John Moore, 1998. "Cooperatives vs. Outside Ownership," NBER Working Paper 6421, National Bureau of Economic Research, February.
- Hasan, Iftekhhar, Markku Malkamäki and Heiko Schmiedel, 2003. "Technology, Automation, and Productivity of Stock Exchanges: International Evidence," *Journal of Banking and Finance*, 27, pp. 1743-73.
- Hazarika, Sonali, 2005. "Governance Change in Stock Exchanges," Working Paper, Baruch College, City University of New York, January.
- IMF, 2005. *Financial Sector Assessment: A Handbook*. International Monetary Fund, Washington, DC.
- IOSCO Technical Committee, 2006. *Regulatory Issues Arising from Exchange Evolution*. International Organization of Securities Commissions, Madrid, Spain.
- Johnson, Philip M. and Thomas L. Hazen, 2004. *Derivatives Regulation*, Aspen Law and Business, New York, NY.
- Karmel, Roberta S., 2002. "Turning Seats Into Shares: Implications of Demutualization for the Regulation of Stock Exchanges," *Hastings Law Journal*, 53 (2), pp. 367-430.
- Kerr, Dana A., 2005. "The Effect of Ownership Structure on Insurance Company Litigation Strategy," *Journal of Legal Studies*, 34 (1), pp. 273-94.
- Krishnamurti, Chandrasekhar John M. Sequeira, and Fu Fangjian, 2003. "Stock Exchange Governance and Market Quality," *Journal of Banking and Finance*, 27 (9), pp. 1859-78.
- Lamm-Tennant, Joan and Laura T. Starks, 1993. "Stock versus Mutual Ownership Structures: The Risk Implications," *Journal of Business*, 66 (1), pp. 29-46.
- Lee, Ruben, 2003. "Changing Market Structures, Demutualization, and the Future of Securities Trading," Working Paper, Oxford Finance Group and World Bank Institute, April.
- Ligon, James A. and Paul D. Thistle, 2005. "The Formation of Mutual Insurers in Markets with Adverse Selection," *Journal of Business*, 78 (2), pp. 529-55.
- Macey, J.R. and M. O'Hara, 2005. "From Markets to Venues: Securities Regulation in an Evolving World," *Stanford Law Review*, 58, pp. 563-99.
- Mahoney, Paul G., 1997. "The Exchange as Regulator," *Virginia Law Review*, 83 (7), pp. 1453-1500.
- Masulis, Ronald W., 1987. "Changes in Ownership Structure: Conversions of Mutual Savings and Loans to Stock Charter," *Journal of Financial Economics*, 18 (1), pp. 29-59.

- Mayers, David and Clifford W. Smith, Jr., 1981. "Contractual Provisions, Organizational Structure, and Conflict Control in Insurance Markets," *Journal of Business*, 54 (3), pp. 407-34.
- Mayers, David and Clifford W. Smith, Jr., 1986. "Ownership Structure and Control: The Mutualization of Stock Life Insurance Companies," *Journal of Financial Economics*, 16 (1), pp. 73-98.
- Mayers, David and Clifford W. Smith, Jr., 1988. "Ownership Structure across Lines of Property-Casualty Insurance," *Journal of Law and Economics*, 31 (2), pp. 351-78.
- Mayers, David and Clifford W. Smith, Jr., 1994. "Managerial Discretion, Regulation, and Stock Insurer Ownership Structure," *Journal of Risk and Insurance*, 61 (4), pp. 638-55.
- Mayers, David and Clifford W. Smith, Jr., 2005. "Agency Problems and the Corporate Charter," *Journal of Law, Economics, and Organization*, 21 (2), pp. 417-40.
- McNamara, Michael J. and S. Ghon Rhee, 1992. "Ownership Structure and Performance: The Demutualization of Life Insurers," *Journal of Risk and Insurance*, 59 (2), pp. 221-38.
- Mester, Loretta J., 1991. "Agency Costs among Savings and Loans," *Journal of Financial Intermediation*, 1 (3), pp. 257-78.
- O'Hara, Maureen and Alfredo M. Mendiola, 2003. "Taking Stock in Stock Markets: The Changing Governance of Exchanges," Cornell University Working Paper presented at the 2004 NBER Market Microstructure Meeting, November.
- Pirrong, Craig, 1995. "The Self-Regulation of Commodity Exchanges: The Case of Market Manipulation," *Journal of Law and Economics*, 38 (1), pp. 141-206.
- Pirrong, Craig, 1999. "The Organization of Financial Exchange Markets: Theory and Evidence," *Journal of Financial Markets*, 2 (4), pp. 329-57.
- Pirrong, Craig, 2000. "A Theory of Financial Exchange Organization," *Journal of Law and Economics*, 43 (2), pp. 437-71.
- SEC, 2004. U.S. Securities and Exchange Commission Concept Release Concerning Self-Regulation, *Federal Register* v69 n235, 71256-82 (17 CFR PART 240, Release No. 34-50700; File No. S7-40-04, RIN 3235-AJ36). Available online: <http://www.sec.gov/rules/concept/34-50700.pdf>
- Santos, Tano and Jose A. Scheinkman, 2001. "Competition among Exchanges," *Quarterly Journal of Economics*, 116 (3), pp. 1027-61.
- Serifsoy, Baris, 2007. "Demutualization, Outside Ownership, and Stock Exchange Performance – Empirical Evidence," *Economics of Governance*, Forthcoming.
- Serifsoy, Baris and Mark Tyrell, 2006. "Investment Behavior of Stock Exchanges and the Rationale for Demutualization – Theory and Empirical Evidence," Wharton Financial Institutions Center Working Paper No. 06-17, September.
- Smith, Bruce D. and Michael Stutzer, 1990a. "Adverse Selection, Aggregate Uncertainty, and the Role for Mutual Insurance Contracts," *Journal of Business*, 63 (4), pp. 493-510.
- Smith, Bruce D. and Michael Stutzer, 1990b. "Adverse Selection and Mutuality: The Case of the Farm Credit System," *Journal of Financial Intermediation*, 1 (2), pp. 125-49.
- Smith, Bruce D. and Michael Stutzer, 1995. "A Theory of Mutual Formation and Moral Hazard with Evidence from the History of the Insurance Industry," *Review of Financial Studies*, 8(2), pp. 545-77.
- Steil, Benn, 2002. "Changes in the Ownership and Governance of Securities Exchanges: Causes and Consequences," *Brookings-Wharton Papers on Financial Services*, 2002, pp. 61-82.

- Stoll, Hans R., 2002. "Comments on 'Changes in the Ownership and Governance of Securities Exchanges: Causes and Consequences', by Benn Steil," *Brookings-Wharton Papers on Financial Services*, 2002, pp. 83-8.
- Townsend, Robert, 1979. "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory*, 21 (2), pp. 265-93.
- Treptow, Felix, 2006. *The Economics of Demutualization: An Empirical Analysis of the Securities Exchange Industry*. Wiesbaden: Deutscher Universitaets-Verlag, xiv+111 pages.
- Viswanathan, Krupa S., 2006. "The Pricing of Insurer Demutualization Initial Public Offerings," *Journal of Risk and Insurance*, 73 (3), pp. 439-68.
- Viswanathan, Krupa S., and J. David Cummins, 2003. "Ownership Structure Changes in the Insurance Industry: An Analysis of Demutualization," *Journal of Risk and Insurance*, 70 (3), pp. 401-37.

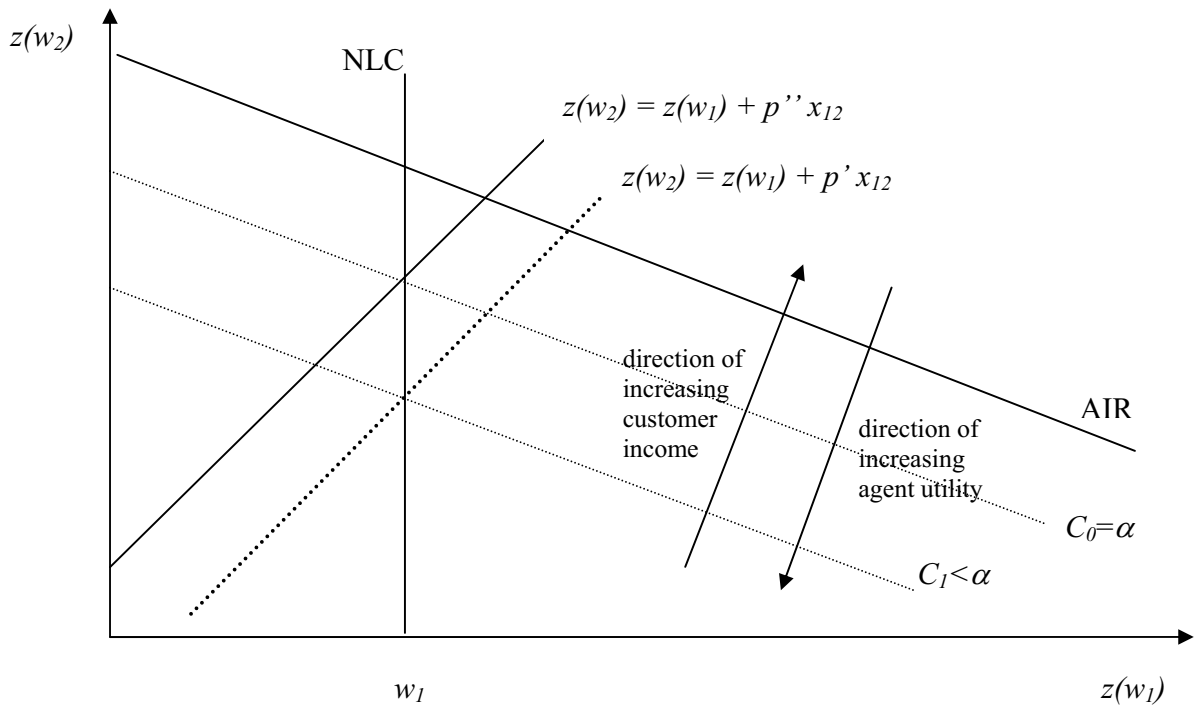
**Figure 2: The Customer's Problem**



**Notes:** Figure 2 provides a graphical illustration of the constraints that the customer faces when choosing the agent's compensation schedule  $\{w_1 - z(w_1), w_2 - z(w_2)\}$ . The agent's incentive compatibility constraint (1) is depicted by the 45-degree AIC line  $z(w_2) = z(w_1) + p_1 x_{12}$ . To ensure truth-telling,  $z(w_2)$  must lie on or below this line. The agent's individual rationality constraint (2) is depicted by the downward-sloping AIR line  $z(w_2) = w_2 + \pi_1 w_1 / \pi_2 - \pi_1 z(w_1) / \pi_2$ . To get participation,  $z(w_2)$  must lie on or below this line. Finally, the no-loss condition (3) is depicted for  $i = 1$  by the vertical line NLC at  $z(w_2) = w_1$ . To respect limited liability,  $z(w_2)$  must lie to the left of this line. The striped five-sided area depicts the combinations of  $z(w_1)$  and  $z(w_2)$  that simultaneously meet all three constraints (AIC, AIR and NLC).

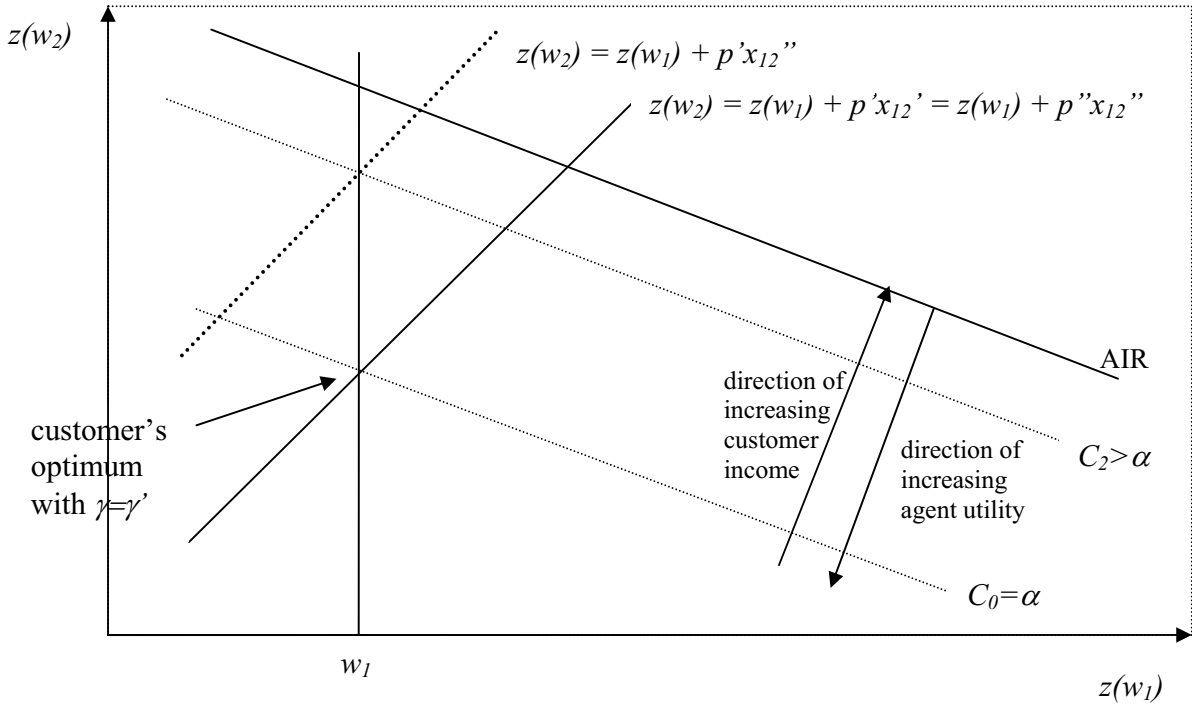


**Figure 3: The Mutual SRO's Choice of Investigation Probability**



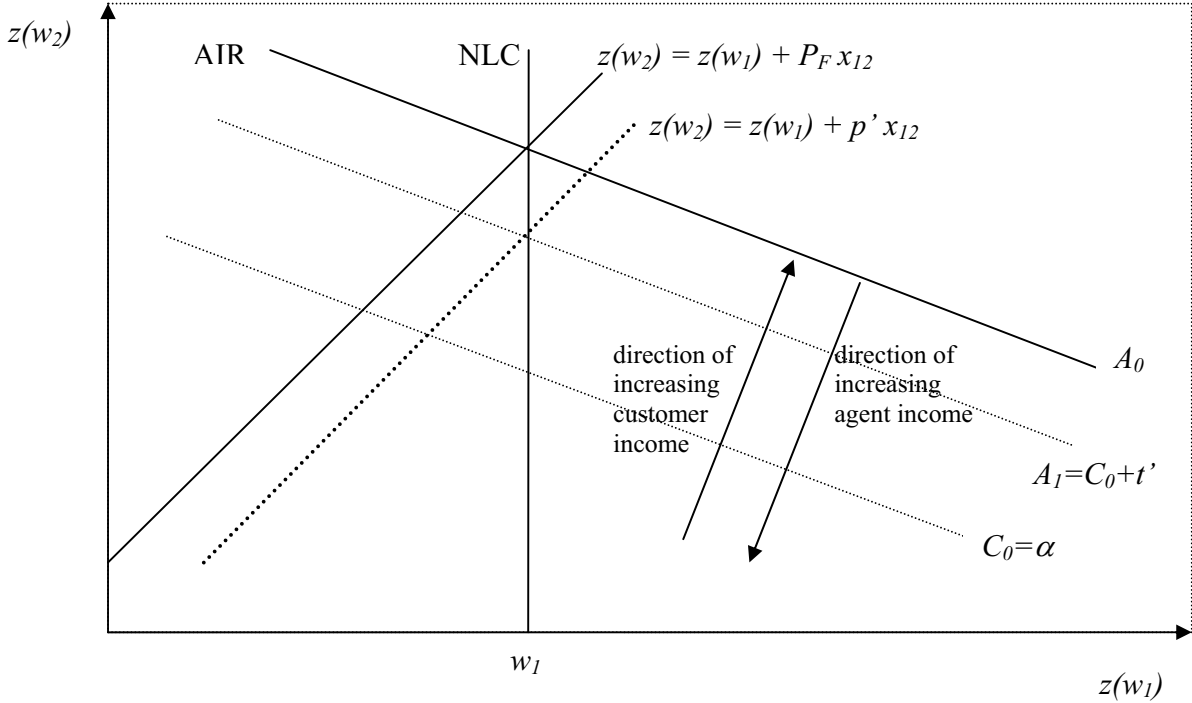
**Notes:** Figure 3 illustrates how the mutual SRO's choice of investigation probability,  $p_1$ , affects the agent's and the customer's respective expected utilities. *Ceteris paribus*, by increasing  $p_1$  from  $p'$  to  $p''$ , the SRO shifts the AIC upward and increases the customer's expected income from  $C_1 < \alpha$  to  $C_0 = \alpha$ .

**Figure 4: Effect of Agent Wealth on Mutual SRO Policy**



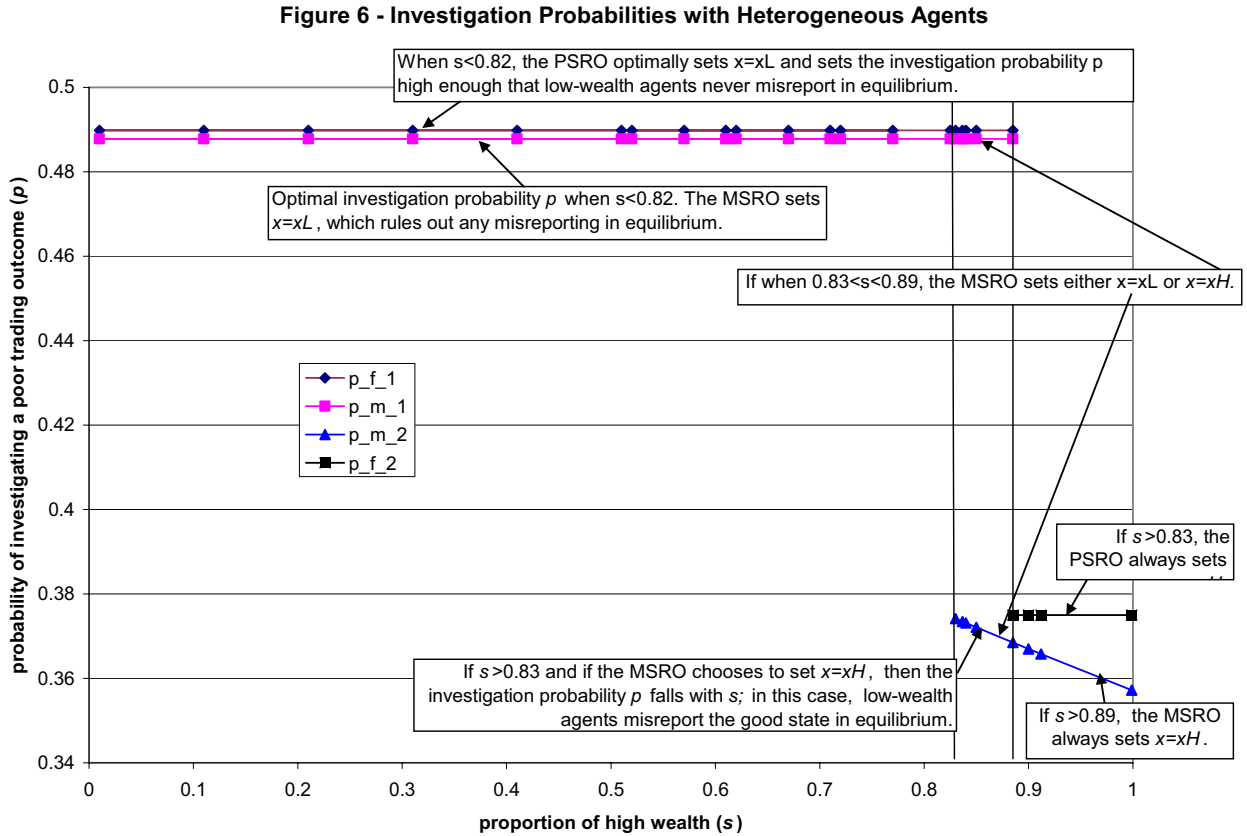
**Notes:** Figure 4 illustrates the relation between customer income and the exogenous agent-wealth parameter,  $\gamma$ . If  $\gamma$  increases from  $\gamma'$  to  $\gamma''$ , then the optimal penalty  $x = w_2 - w_1 + \gamma$  increases and the customer, *ceteris paribus*, could attain a higher expected income (moving from  $C_0$  to  $C_2$ ) and the agent would move to a lower income (holding  $p_1$  constant at  $p'$ ). Accordingly, as long as  $w_1 < \alpha$ , the SRO lowers  $p_1$  as  $\gamma$  increases (from  $p'$  to  $p''$  in this depiction), which reduces the customer's expected income back to  $\alpha$ . By reducing the expected investigation costs  $p_1\pi_2c$ , this reduction in  $p_1$  increases the combined agent-customer income net of the fees  $t = p_1\pi_2c$  and, thus, raises the agent's expected income.

**Figure 5: The For-Profit SRO's Problem**



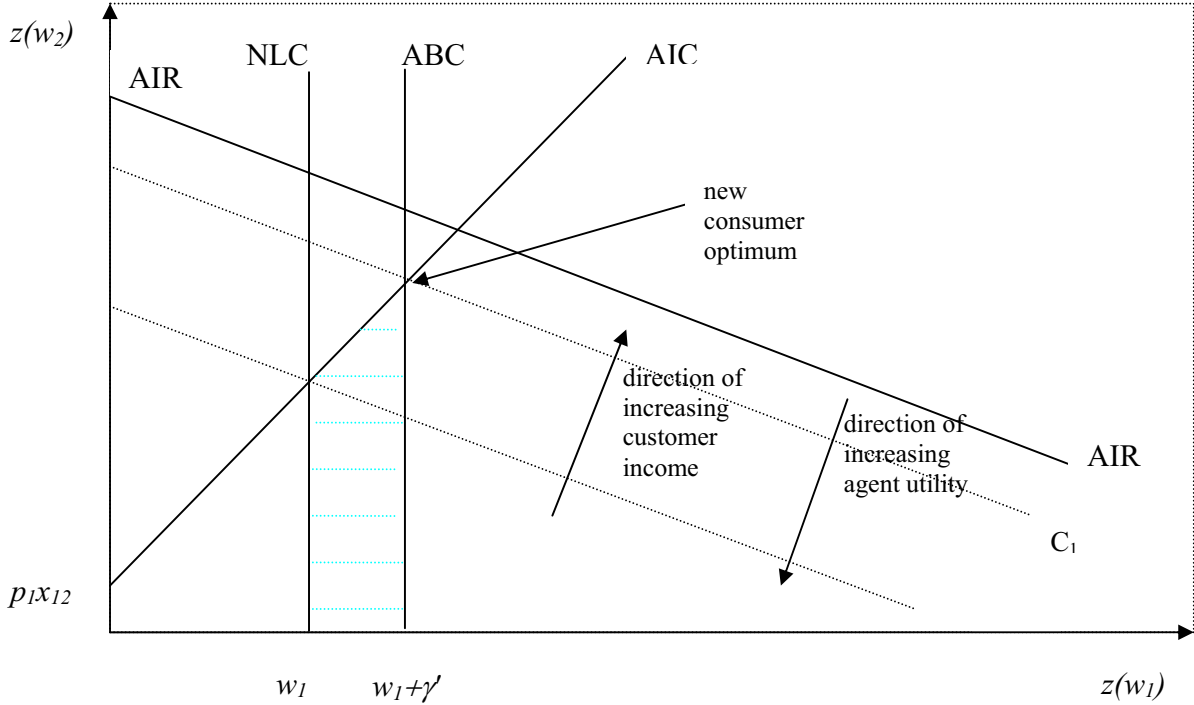
**Notes:** Figure 5 illustrates the for-profit SRO's optimization.  $C_0$  is the income level associated with the customer's participation constraint, i.e., with the CIR condition (4).  $A_0$  is the income level associated with the agent's participation constraint, i.e., with the AIR condition (2). As in Figure 4,  $z^*(w_2) = z^*(w_1) + p_1 (w_2 - w_1)$  is the agent's incentive compatibility constraint (1), and the vertical NLC line at  $w_1$  reflects the no-loss constraint (3). For some arbitrary value of  $p_1 < P_F$ , say  $p'$ , the agent would expect income  $A_1 > A_0$ , and  $t$  would be set equal to  $A_1 - C_0$ . As  $p_1$  is increased towards  $P_F$ , the agent's income falls towards  $A_0$ , and  $t$  can be increased without violating condition (4). Thus, the equilibrium fee  $t$  is the vertical distance between  $A_0$  and  $C_0$ , with  $p_1 = P_F$ .

**Figure 6 – Detection Probability with Heterogeneous Agents**



**Notes:** Figure 6 provides a numerical example of the relationship between exchange ownership structure, detection probability  $P$ , and proportion  $s$  of high-wealth agents. Parameter values are set to  $\alpha = 1.5$ ,  $\gamma_H = 2 > \gamma_L = 1.25$ ,  $\pi_I = 0.5$ ,  $c = 0.25$ ,  $w_2 = 2.2$ , and  $w_1 = 1$ . When agents are overwhelmingly high-wealth types ( $s > 0.88$ ; case (ii) in Proposition 2), both the MSRO and the PSRO set  $x = x_H$ . The mutual's investigation probability  $P_{M,2} = 0.374$  at  $s = 0.88$  and is decreasing in  $s$  (Corollary 2), while the for-profit's  $P_{F,2}$  is larger ( $P_{F,2} = 0.375$ ) and independent of  $s$ . In that case, low-wealth agents misreport the good state in equilibrium. If the proportion  $s$  of high-wealth agents is "low" ( $s < 0.83$ ; case (i) in Proposition 2), then the MSRO and the PSRO both set  $x = x_L$  and deter misreporting by raising the investigation probability to, respectively,  $P_{M,1} = 0.48$  or  $P_{F,1} = .49 > P_{M,1}$ . Finally, for intermediate values of  $s$  ( $0.83 < s < 0.88$ ; case (iii) in Proposition 2), the PSRO sets  $p_I = P_{F,1} = 0.49$  whereas either the MSRO sets  $x = x_L$  and  $p = P_{M,1} = 0.48$  to benefit the high-wealth agents or, alternatively, or the MSRO sets  $x = x_H$  and  $p = P_{M,2} \leq 0.374$  to benefit low-wealth agents.

**Figure 7: Importance of Agent-Liability Limits**



**Notes:** Figure 7 illustrates the relation between customer income and the exogenous agent-wealth parameter,  $\gamma$ , in the event that agent income can be negative in some states of the world. In Figure 2, the agent's income is restricted to non-negative values in all states of the world, i.e., the no-loss condition (3) required that  $z(w_1) \leq w_1$ . This constraint is represented by the vertical NLC line above. In Figure 7, this no-loss constraint is replaced by the weaker requirement that the agent can lose money in some states of the world, but only to the extent that his wealth  $\gamma$  is sufficient to cover the loss:  $z(w_1) \leq w_1 + \gamma$ . This weaker no-bankruptcy condition is represented by the vertical ABC line above. *Ceteris paribus*, relaxing the NLC allows the customer to attain a higher expected income (moving from  $C_0$  to  $C_1$ ) and forces the agent to a lower income (holding  $p_1$  constant).

**Appendix: Proofs.**

**Lemma 2:** The mutual SRO sets  $p_1 = p_M = \frac{\alpha - w_1}{\pi_2(w_2 - w_1 + \gamma) - \pi_1 c}$ ;  $p_2 = 0$ ,  $x_{12} = w_2 - w_1 + \gamma$ ;  
and  $x_{ij} = 0$  for all other  $i, j$ .

**Proof:** To see that  $p_2 = 0$  and  $x_{2j} = 0$  ( $j=1,2$ ), note that condition (3) requires that  $z(w_1) < \alpha$ , so condition (4) requires  $z(w_2) > \alpha > z(w_1)$ . Because  $z(w_2) > z(w_1)$ , the revelation principle implies that we can restrict our attention to contracts that induce truthful reporting by agents. As such, there is no reason for an SRO to set  $p_2 > 0$ , since the agent will never report that state 2 occurred if state 1 actually occurred. Consequently,  $x_{2j}$  is irrelevant, and we therefore set it equal to 0.

To show that  $x_{12} \equiv x = w_2 - w_1 + \gamma$ , first note that if  $x \leq w_2 - w_1 + \gamma$  then the agent's income is  $\pi_2 (w_2 - w_1 - px)$ . Suppose  $x = x' < w_2 - w_1 + \gamma$ , and let  $p'$  be the value of  $p$  such that the CIR is satisfied when  $x = x'$ . Let  $p'x' = k$ . It follows that agent income is the same for any  $p$  as long as  $px = k$ . Customer income is  $\pi_2 (w_1 + px) + \pi_1 w_1 - \pi_1 pc$ , and therefore decreasing in  $p$  for  $x = k/p$ . Hence, setting  $x = w_2 - w_1 + \gamma$  (rather than  $x'$ ) has no direct effect on agent's income, but relaxes the CIR, allowing the agent to profitably increase his income without violating the CIR.

To determine  $p$ , note that for given  $z^*(w_1)$  and  $z^*(w_2, \gamma)$  functions, the agent's wealth is decreasing in  $px$ . Therefore, the MSRO will choose the minimum  $px$  that is consistent with condition (4'). Since we know that  $x = w_2 - w_1 + \gamma$ , the MSRO will choose the smallest  $p$  that satisfies

$$(A.1) \quad z^*(w_2) \geq (\alpha - \pi_1 (w_1 - c p))/\pi_2$$

or, given the customer's optimal fee schedule,

$$(A.1') \quad w_1 + p_1(w_2 - w_1 + \gamma) \geq (\alpha - \pi_1 (w_1 - c p_1))/\pi_2$$

This implies that, if  $\pi_2 (w_2 - w_1 + \gamma) - \pi_1 c \leq \alpha - w_1$ , then there is no  $p \leq 1$  that allows the MSRO to provide customers with an expected income  $\geq \alpha$ . In a similar vein, unless  $(w_2 - w_1 + \gamma)(\pi_1 w_1 + \pi_2 w_2 - \alpha) > \pi_1 c (w_2 - w_1)$ , the AIR cannot be met. Both the CIR and the AIR are met as long as there are gains from trade – which is guaranteed by our assumption that  $\pi_1 w_1 + \pi_2 w_2 > \alpha + \pi_1 c$ . Solving (A.1') yields:

$$p = p_M = \frac{\alpha - w_1}{\pi_2(w_2 - w_1 + \gamma) - \pi_1 c} \quad \square$$

**Lemma 3:** The for-profit SRO sets  $p_1 = p_F = \frac{w_2 - w_1}{w_2 - w_1 + \gamma}$ ;  $p_2 = 0$ ,  $x_{12} = w_2 - w_1 + \gamma$ ;

and  $x_{ij} = 0$  for all other  $i, j$ ; and,  $t = w_1 + \pi_2 P_F x - \alpha = w_1 + \pi_2 (w_2 - w_1) - \alpha$

**Proof:** By the same logic as for the MSRO, the PSRO sets  $x = x_{12} = w_2 - w_1 + \gamma$ . To find  $P_F$ , note that the CIR condition (4) must be binding. That is, (4) can be written:

$$(A.2) \quad \pi_2 p (w_2 - w_1 + \gamma) = \alpha + t - w_1$$

Because  $z^*(w_1) = w_1$ , the NLC condition (3) is binding in state 1 for any  $p$ . Condition (3) will be binding in state 2 if  $z^*(w_2) = w_2$ , i.e., if  $p = (w_2 - w_1)/x$ .

If condition (3) is not binding in state 2, then the optimal transaction fee solves the PSRO's first order condition with respect to  $t$ , so that

$$(A.3) \quad 1 - \pi_1 c \partial p / \partial t = 0,$$

where,  $\partial p / \partial t$  is the slope of the CIR constraint (8), holding customer income fixed at  $\alpha$ . This slope is equal to  $1 / ((w_2 - w_1 + \gamma) \pi_2) > 0$ , and therefore (9) can be written

$$(A.3') \quad 1 - \pi_1 c / ((w_2 - w_1 + \gamma) \pi_2) = 0.$$

Given that  $c < \pi_2 (w_2 - w_1 + \gamma) / \pi_1$ , it follows from (A.3') that, as long as equation (3) is not binding in state 2, the PSRO should increase the transactions fee  $t$  and the detection probability  $p$ , in order to keep customers at income level  $\alpha$ . As long as (3) is not binding, the SRO can raise  $p$ , resulting in a higher  $z^*(w_2)$ , without violating any other constraint.

Once  $t$  is sufficiently large that the resulting  $z^*(w_2)$  is high enough that the NLC (3) is binding, then further increases in  $t$  are not profitable. That is, once (3) is binding in state 2, the PSRO can only increase  $p$  (without inducing agent exit) by reducing  $x$ . Since the optimal penalty is  $x = w_2 - w_1 + \gamma$ , it will not be profitable to increase  $p$  beyond

$$P_F = \frac{w_2 - w_1}{w_2 - w_1 + \gamma}.$$

Given this choice of  $p$ , the PSRO chooses  $t$  to set customer income equal to  $\alpha$  (so that equation (4') binds), or  $t = w_1 + \pi_2 P_F (w_2 - w_1 + \gamma) - \alpha = w_1 + \pi_2 (w_2 - w_1) - \alpha$ . Finally, the PSRO will not operate unless profits are positive, i.e., unless:

$$\pi_2 (w_2 - w_1) - \pi_1 c (w_2 - w_1) / (w_2 - w_1 + \gamma) \geq \alpha - w_1$$

This condition, which also ensures customer participation, is satisfied given our assumption that there are gains from trade, i.e., that  $\pi_1 w_1 + \pi_2 w_2 > \alpha + \pi_1 c$ .  $\square$

**Proposition 1:** The for-profit SRO spends at least as much on enforcement as does the mutual SRO.

**Proof:** From Lemmas 2 and 3, we have  $P_M = \frac{\alpha - w_1}{\pi_2 (w_2 - w_1 + \gamma) - \pi_1 c}$  and  $P_F = \frac{w_2 - w_1}{w_2 - w_1 + \gamma}$ .

$P_F > P_M$  if  $(w_2 - w_1)[\pi_2 (w_2 - w_1) - (\alpha - w_1) - \pi_1 c] + \gamma[\pi_2 (w_2 - w_1) - (\alpha - w_1)] > 0$  – which is the case given  $\pi_1 w_1 + \pi_2 w_2 > \alpha + \pi_1 c$  and, hence,  $\pi_2 (w_2 - w_1) - \pi_1 c > \alpha - w_1$ .  $\square$

**Lemma 4:** AIC-H is met whenever AIC-L is met. The reverse is not true.

**Proof:** There are two cases to consider, depending on the SRO's choice for the penalty  $x$ .

- (i) If the SRO has set  $x \leq w_2 - w_1 + \gamma_L$ , then  $\gamma_L$  does not enter the AIC. In this first case, the Lemma holds trivially.
- (ii) If the SRO has instead set  $x > w_2 - w_1 + \gamma_L$ , then the right-hand side of AIC-L equals  $-p\gamma_L + (1-p)(w_2 - w_1)$ . In this second case, it follows from (5) and from  $x \geq w_2 - w_1 + \gamma_L$  that AIC-H is met whenever AIC-L is met – whereas the reverse is not true.  $\square$

**Lemma 5:** When agents are heterogeneous, customers set  $z^*(w_1) = w_1$  and  $z^*(w_2) = w_1 + p_1 x_{12}$

**Corollary 1:** If  $x_{12} > x_L$ , then the optimal payment schedule  $Z^*(W)$  leads to misreporting by low-wealth agents. If  $x_{12} = x_L$ , then  $Z^*(W)$  induces truth telling by all agents.

**Proof:** Using the same logic as in Lemma 1, it is immediate that  $z^*(w_1) = w_1$ . In regard to the customer's choice of  $z^*(w_2)$ , there are two cases to consider, depending on the SRO's choice for the penalty  $x$ .

- (i) If the SRO has set  $x_{12} \leq x_L \equiv w_2 - w_1 + \gamma_L$ , then  $\gamma_L$  does not enter the AIC. This case is similar to the situation in Section IV, with the customer choosing  $z^*(w_2) = w_1 + p_1 x_{12}$  and no misreporting taking place in equilibrium.
- (ii) If the SRO has set  $x_{12} > x_L$ , the right-hand side of AIC-L equals  $-p\gamma_L + (1-p)(w_2 - w_1)$ . In this case, the customer's choice of  $z^*(w_2)$  determines whether AIC-L is binding.

If the customer chooses  $z(w_2) = w_1 + p_{12} x_L$ , then even low-wealth agents have no incentive to misreport the state, since the AIC binds for them. This in turn implies that the customer will never set  $z(w_2) < w_1 + p_{12} x_L$ . Indeed, setting  $z(w_2)$  below  $w_1 + p_{12} x_L$  would reduce customer's revenue without changing the agent's incentive to misreport, and hence is strictly dominated by  $z(w_2) = w_1 + p_{12} x_L$ .

An alternative choice for the customer is to set  $z(w_2) = w_1 + p_{12} x > w_1 + p_{12} x_L$ . By doing so, the customer can get  $p_{12}(x - x_L)$  more from high-wealth agents when the good state occurs. However, if the SRO has set  $x > x_L$  and the customer has chosen  $z(w_2) = w_1 + px$ , then the low-wealth agent's AIC becomes  $w_2 - w_1 - px \geq -p\gamma_L + (1-p)(w_2 - w_1)$ , or  $w_2 - w_1 + \gamma_L > x$ . Since  $x > w_2 - w_1 + \gamma_L$  in this case, it follows that the low-wealth agent's AIC does not hold, and that he will misreport when state 2 occurs. This means that the revenue a customer can expect to receive from a low-wealth agent when she chooses  $z(w_2) = w_1 + px$  is  $w_1 + px_L < w_1 + px$ .

By assumption, the customer takes the exchange's enforcement policy as exogenous and focuses solely on the marginal impact, on her expected income from trading, of her tolerating misreporting by low-wealth agents. Hence, given that  $z^*(w_1) = w_1$  in all cases, the customer will prefer to set  $z(w_2) = w_1 + px$  and to face a positive probability of misreporting – rather than to lower  $z(w_2)$  to  $w_1 + px_L$  so as to discourage all misreporting – as long as

$$(A.4) \quad s(w_1 + px) + (1-s)[w_1 + px_L] > w_1 + px_L$$

The left-hand of (A.4) represents the customer's expected revenue in state 2 when some misreporting takes place, and the right-hand side is the expected revenue when  $x$



$= x_L$  and no misreporting occurs. Since this inequality holds for all  $x > x_L$ , it is optimal for the customer to set the fee schedule so as to maximize the revenue received from high-wealth types. Put differently, when  $x > x_L$ , the customer chooses  $z^*(w_2) = w_1 + px$  (thereby allowing some misreporting) rather than setting  $z(w_2)$  sufficiently low to discourage all misreporting.  $\square$

**Lemma 7:** If  $x_{I2} = x_L$ , then the mutual SRO chooses  $P_{M,1} = \frac{\alpha - w_1}{\pi_2(w_2 - w_1 + \gamma_L) - \pi_1 c}$

If  $x_{I2} = x_H$ , then  $P_{M,2} = \frac{\alpha - w_1}{\pi_2(w_2 - w_1 + \gamma_L) + \pi_2 s(\gamma_H - \gamma_L) - [\pi_1 + \pi_2(1-s)]c}$

**Proof:** Since agents' income is decreasing in  $p$  (holding  $x$  constant), the mutual SRO wants to choose the minimal  $p$  consistent with the customer's participation constraint (CIR).

(i) When  $x = x_L$ , the CIR for all agents is  $\pi_1 w_1 + \pi_2 [w_1 + p x_L] - \pi_1 c p \geq \alpha$ , or

$$w_1 + \pi_2 p x_L - \pi_1 c p \geq \alpha,$$

Solving for  $p$  yields  $P_{M,1} = \frac{\alpha - w_1}{\pi_2 x_L - \pi_1 c}$ , i.e.,  $P_{M,1} = \frac{\alpha - w_1}{\pi_2(w_2 - w_1 + \gamma_L) - \pi_1 c}$ .

The last expression is the same as that found in Lemma 2 when agents are homogenous.

(ii) When  $x > x_L$ ,  $x_L$  is the maximum fine that can be imposed on a low-wealth agent. Hence, as stated in Corollary 1, AIC-L will be violated and low-wealth agents will misreport. This violation implies that, for  $x > x_L$ , the CIR becomes

$$w_1 + \pi_2 [spx + (1-s)px_L] - [\pi_1 + (1-s)\pi_2] pc \geq \alpha$$

where, the probability  $\pi_1 + (1-s)\pi_2$  of having state 1 reported reflects misreporting of state 2 by low-wealth agents.

In particular, when  $x = x_H$ , the CIR simplifies to

$$w_1 + \pi_2 p[x_L + s(\gamma_H - \gamma_L)] - [\pi_1 + (1-s)\pi_2] pc \geq \alpha$$

Solving for  $p$  yields  $P_{M,2} = \frac{\alpha - w_1}{\pi_2(w_2 - w_1 + \gamma_L) + \pi_2 s(\gamma_H - \gamma_L) - [\pi_1 + \pi_2(1-s)]c}$   $\square$

**Proposition 2:** The MSRO sets  $x_{I2} = x_L$  or  $x_{I2} = x_H$ . The choice between  $x_L$  and  $x_H$  is a function of the proportion  $s$  of high-wealth agents, and may also depend on whose expected income the MSRO is maximizing. In particular:

- (a) For sufficiently low values of  $s$  ( $s \leq \bar{s}$ ), the MSRO sets  $x_{I2} = x_L$
- (b) For sufficiently high values of  $s$  ( $s > \hat{s}$ ) the MSRO sets  $x_{I2} = x_H$
- (c) Otherwise (for intermediate values of  $s$ ), the MSRO's enforcement policy depends on the type of agent whose income it maximizes:
  - $x_{I2} = x_L$  if the exchange acts on behalf of high-wealth agents;
  - $x_{I2} = x_H$  if the exchange acts on behalf of low-wealth agents.

Where 
$$\hat{s} = \frac{\pi_2 x_L (c + \gamma_H - \gamma_L) - \pi_1 (\gamma_H - \gamma_L) c}{\pi_2 x_L (c + \gamma_H - \gamma_L)} \text{ and}$$

$$\bar{s} = \max \left\{ \frac{c}{c + \gamma_H - \gamma_L}, \frac{(\alpha - w_1) x_H - (w_2 - w_1) (\pi_2 x_L - c)}{\pi_2 (w_2 - w_1) (\gamma_H - \gamma_L + c)} \right\}$$

**Proof:** We first establish that the MSRO always sets either  $x = x_L$  or  $x = x_H$ . Since Lemma 6 implies that  $x \geq w_2 - w_1 + \gamma_L \equiv x_L$ , proving this first claim is equivalent to showing that every agent either prefers  $x_L$  to any  $x \in (x_L, x_H)$  or prefers  $x_H$  to any  $x \in [x_L, x_H)$ .

*Low-wealth agents:* whenever  $x > x_L$ , low-wealth agents misreport (see Corollary 1). Hence, as long as high-wealth agents participate, each low-wealth agent will earn an expected income of

$$\pi_2 [(1-p)(w_2 - w_1) - p \gamma_L]$$

which is decreasing in  $p$  and independent of  $x$ . Since, once  $x$  is strictly greater than  $x_L$ , an increase in  $x$  allows the MSRO to decrease  $p$  without inducing customer exit, low-wealth agents always prefer higher  $x$ , up to the point that  $x$  is no longer binding on high-wealth individuals, i.e.,  $x = x_H$ . Hence, low-wealth agents strictly prefer  $x_H$  to  $x \in (x_L, x_H)$  as long as high-wealth agents participate.

*High-wealth agents:* for a high-wealth agent (who does not misreport), when  $x \in [x_L, x_H]$  expected income is

$$\pi_1 (w_1 - z^*(w_1)) + \pi_2 (w_2 - z^*(w_2)) = \pi_2 (w_2 - w_1 - px)$$

Hence, for  $x > x_L$ , increasing  $x$  will only be profitable for high-wealth individuals if it allows  $px$  to fall (i.e., if it relaxes the CIR). The CIR in this case is

$$\text{CIR} \quad (4'') \quad w_1 + p \pi_2 [s x + (1-s) (w_2 - w_1 + \gamma_L)] - (\pi_1 + (1-s) \pi_2) cp \geq \alpha$$

where the transaction fee  $t = (\pi_1 + (1-s) \pi_2) cp$  because state 2 is always misreported by low-wealth agents. Letting  $px = k$ , the change in the CIR from increasing  $p$  is

$$(A.5) \quad (1-s) \pi_2 x_L - [\pi_1 + (1-s) \pi_2] c$$

If expression (A.5) is positive, then a reduction in  $x$  (along with an increase in  $p$  so that  $px$  is unchanged) will relax the CIR, which means that the MSRO can reduce  $px$  (thereby increasing the high-wealth's agent's income) without violating the CIR. Hence, if (A.5) is positive, then high-wealth agents' profits will be maximized at  $x = x_L$  (since profits at  $x = x_L$  are greater than profits at  $x = x_L + \delta$ , which in turn are greater than profits at  $x = x_H$ ). Conversely, if this derivative is negative, then a high-wealth agent will earn more at  $x = x_H$  than at  $x \in (x_L, x_H)$ .

Together, the above results establish that any agent either prefers  $x_{I2} = x_L$  or  $x_{I2} = x_H$  to any other possible choice of the penalty  $x$ . The remainder of the proof identifies under what circumstances the MSRO sets  $x_{I2} = x_L$  or, alternatively,  $x_{I2} = x_H$ .

(a) For  $s$  sufficiently close to 0, both kinds of agents will prefer  $x = x_L$  to  $x = x_H$ .

Proof: When the penalty is  $x$ , high-wealth agents' incomes are  $w_2 - w_1 - px$ , so that their incomes are higher at  $x = x_L$  than at  $x = x_H$  if  $x_H P_{M,2} > x_L P_{M,1}$ , or

$$(\alpha - w_1) \left[ \frac{x_L}{\pi_2 x_L - \pi_1 c} - \frac{x_L + (\gamma_H - \gamma_L)}{\pi_2 x_L + \pi_2 s (\gamma_H - \gamma_L) - (\pi_1 + (1-s)\pi_2)c} \right] < 0$$

which is true if

$$(A.6) \quad c[(\gamma_H - \gamma_L)\pi_1 - (1-s)x_L\pi_2] - (1-s)(\gamma_H - \gamma_L)\pi_2 x_L$$

is negative. As  $s$  goes to zero, (A.6) goes to  $c[(\gamma_H - \gamma_L)\pi_1 - x_L\pi_2] - (\gamma_H - \gamma_L)\pi_2 x_L$ , which is negative since we have assumed that condition (6) holds and, hence, that  $\pi_1 c < \pi_2 x_L$ .

Low-wealth agent's income at  $x = x_L$  is higher than at  $x = x_H$  (which induces misreporting) if  $\pi_2(w_2 - w_1 - p_{M,1}x_L) > \pi_2[(1 - p_{M,2})(w_2 - w_1) - p_{M,2}\gamma_L]$ , or simply  $P_{M,2} > P_{M,1}$ .

If this inequality is met, then since  $x_H > x_L$ , all agents will prefer  $x = x_L$ .  $P_{M,2} - P_{M,1}$

takes the sign of  $(1-s)c - s(\gamma_H - \gamma_L)$ , which is positive for  $s \leq s_1 \equiv \frac{c}{c + \gamma_H - \gamma_L}$ .

Finally, note that condition (6) implies that all of the relevant constraints can be met when the MSRO sets  $x = x_L$ . In order for any  $x$  and  $p$  to constitute an equilibrium, it must be that both the CIR and AIR can simultaneously be satisfied. The highest  $p$  that can satisfy the high-wealth agent's AIR in the equilibrium with no misreporting by low-wealth agents solves

$$\pi_1 w_1 + \pi_2 [w_1 + p x_L] = \pi_1 w_1 + \pi_2 w_2 \quad \text{or}$$

$$p = (w_2 - w_1)/x_L$$

and this  $p$  can satisfy the CIR if

$$w_1 + \pi_2 p x_L - p \pi_1 c - \alpha \geq 0, \text{ or } w_1 + \left(\frac{w_2 - w_1}{x_L}\right)\pi_2 x_L - \alpha - \pi_1 \left(\frac{w_2 - w_1}{x_L}\right)c \geq 0$$

This last inequality is condition (6). It follows that an equilibrium exists in which  $x = x_L$  if  $s \leq s_1 \equiv \frac{c}{c + \gamma_H - \gamma_L}$ .

To complete the proof of statement (a), note that an equilibrium in which  $x = x_H$  need not exist even if one exists when the MSRO sets  $x = x_L$ . That is, if high-wealth agents do not participate for an  $x > x_L$ , then setting  $x$  greater than  $x_L$  can have no additional effect on expected penalties, and hence  $x$  is effectively limited to  $x \equiv x_L$ . This situation happens if high-wealth agents' participation constraint is not met at  $x = x_H$ , i.e., if  $w_2 - w_1 - P_{M,2} x_H < 0$  or

$$s < s_1 \equiv \frac{(\alpha - w_1)x_H - (w_2 - w_1)(\pi_2 x_L - c)}{\pi_2 (w_2 - w_1)(\gamma_H - \gamma_L + c)}$$

Hence, for  $s \leq \max\{s_1, s_2\}$ , both kinds of agents prefer  $x = x_L$ .

(b) For  $s$  sufficiently close to 1, both kinds of agents will prefer  $x = x_H$ .

Proof: A high-wealth agent's incomes is higher at  $x = x_H$  than at  $x = x_L$  if  $x_H P_{M,2} < x_L P_{M,1}$ , or equivalently, if equation (A.6) is negative. As  $s$  goes to 1, equation (A.6) will necessarily be negative. In addition, since the CIR and AIR-H constraints can be satisfied at  $x = x_L$  for all  $s$ , it follows that they both can be satisfied at  $x = x_H$  if equation (A.6) is negative.

A low-wealth agents' incomes at  $x = x_H$  is higher than at  $x = x_L$  if  $P_{M,2} < P_{M,1}$ . Since  $x_H > x_L$ , it follows that  $x_H P_{M,2} < x_L P_{M,1}$ , is a sufficient condition for  $P_{M,2} < P_{M,1}$ . Hence, low-wealth agents prefer  $x = x_H$  whenever high-wealth agents do.

(c) We know that (A.6) must be negative for  $s$  sufficiently close to 1, positive for  $s$  sufficiently close to 0 and continuously differentiable. Hence, there must exist an  $s$ , which we denote  $\hat{s}$ , such that (A.6) is equal to zero; i.e.,

$$\hat{s} = \frac{\pi_2 x_L (c + \gamma_H - \gamma_L) - \pi_1 (\gamma_H - \gamma_L) c}{\pi_2 x_L (c + \gamma_H - \gamma_L)}.$$

That is, at  $s = \hat{s}$ , high-wealth agents earn the same amount whether  $x = x_L$  or  $x = x_H$ .

Their incomes at  $x = x_L$  are positive, given expression (6). By continuity, high-wealth agents' incomes are positive at  $x = x_H$  for some  $s < \hat{s}$ , but they earn higher income if  $x = x_L$ . In contrast, low-wealth agents strictly prefer  $x = x_H$  at  $s = \hat{s}$ . To see this, recall that low-wealth agents prefer  $x = x_H$  if  $s(\gamma_H - \gamma_L) > (1 - s)c$ . At  $s = \hat{s}$ ,

$$s(\gamma_H - \gamma_L) - (1 - s)c = \frac{\pi_1 (\gamma_H - \gamma_L) c}{\pi_2 (1 - s) x_L} > 0. \text{ By continuity of low-wealth agent incomes, it}$$

follows that for some values of  $s$  less than  $\hat{s}$ , they will prefer  $x = x_H$ , while high-wealth agents strictly prefer  $x = x_L$ .  $\square$

**Lemma 8:** If  $x_{I2} = x_L$ , then the for-profit SRO chooses  $P_{F,1} = \frac{w_2 - w_1}{w_2 - w_1 + \gamma_L}$  and sets  $t = w_1 + \pi_2 P_{F,1} (w_2 - w_1 + \gamma_L) - \alpha = w_1 + \pi_2 (w_2 - w_1) - \alpha$

If  $x_{I2} = x_H$ , then the for-profit SRO chooses  $P_{F,2} = \frac{w_2 - w_1}{w_2 - w_1 + \gamma_H}$  and sets  $t = w_1 + P_{F,2} \pi_2 (w_2 - w_1 + \gamma_L + s(\gamma_H - \gamma_L)) - \alpha$ .

**Proof:** Similar to the case when agents are homogeneous, the PSRO sets  $p$  to extract surplus from agents, and then sets  $t$  to extract surplus from customers. By the same logic as in Lemma 3, if the PSRO sets  $x = x_H$ , then it will choose  $p$  to just satisfy the high-wealth agents' AIR, or

$$P_{F,2} = \frac{w_2 - w_1}{w_2 - w_1 + \gamma_H}$$

and sets  $t$  to satisfy the CIR, or  $t = w_1 + P_{F,2} \pi_2 (w_2 - w_1 + \gamma_L + s(\gamma_H - \gamma_L)) - \alpha$ .

If the PSRO sets  $x = x_L$ , then it sets  $P_F$  to just satisfy both types of agents' AIR, or

$$P_{F,1} = \frac{w_2 - w_1}{w_2 - w_1 + \gamma_L}$$

and the fee  $t$  to satisfy the CIR:  $t = w_1 + \pi_2 P_{F,1} (w_2 - w_1 + \gamma_L) - \alpha = w_1 + \pi_2 (w_2 - w_1) - \alpha$ .  $\square$

**Proposition 3:** The PSRO either sets  $x = x_L$  or  $x = x_H$ . It sets  $x = x_H$  when the proportion  $s$  of high-wealth agents is high ( $\hat{s} < s < 1$ ), and sets  $x = x_L$  otherwise ( $0 < s < \hat{s}$ ).

**Proof:** Misreporting occurs in equilibrium if  $x > x_L$ . First note that  $x$  will be at least  $x_L$ . To see this, note that when  $x < x_L$ , the logic of Lemma 6 holds; it will be profitable to raise  $x$  and lower  $p$ . That is, the AIC is the same for high and low-wealth agents for  $x < x_L$ , and can be written as

$$z(w_2) \leq w_1 + px.$$

Thus, if the PSRO sets  $p$  and  $x$  in order that equation (8) just binds and  $x < x_L$ , then only the product  $p$  times  $x$  matters. Similarly, the customer's income is

$$\pi_2 (w_1 + px) + \pi_1 w_1 - t.$$

so that only the product  $px$  affects customer income.

Hence, if  $x < x_L$ , and the PSRO increases  $x$  and lowers  $p$  to keep  $px$  fixed (so that the constraints continue to bind), it will not affect SRO income, but it will reduce expected enforcement costs. Thus, it will never be optimal to set  $x < x_L$ .

Next we show that  $x$  will either be  $x_L$  or  $x_H$ . Consider some  $x > x_L$ . Suppose that  $x$  and  $p$  are chosen so that the AIR is met for high-wealth agents (and hence by Lemma 4, necessarily is met for low-wealth agents). Holding  $xp$  constant at  $k$ , the AIR will be met at all  $x \in (x_L, x_H]$ . As in Lemma 8, for any  $p$  and  $x$ , the PSRO will set  $t$  such that (9) is binding on customers. The CIR when  $x > x_L$  is

$$\text{CIR} \quad (9a) \quad [\pi_2 (s px + (1-s) x_L) + w_1] - t \geq \alpha$$

So that the change in the CIR from increasing  $p$  (decreasing  $x$ ) is

$$\pi_2 (1-s) x_L$$

This represents the amount by which the SRO can increase  $t$  as it lowers  $x$  and raises  $p$  (to keep  $px = k$ ). If this expression is more than the cost of increasing  $p$  (which is  $[\pi_1 + (1-s) \pi_2] c$ ), then the PSRO will increase its profits by lowering  $x$  (until  $x$  is arbitrarily close to  $x_L$ ). Conversely, if

$$(A.5) \quad \pi_2(1-s)x_L - [\pi_1 + (1-s) \pi_2] c$$

is negative, then the PSRO profits are higher at  $x = x_H$  than at any  $x \in (x_L, x_H)$ .

Finally, note that (A.5)  $< 0$  is a necessary condition for the PSRO to choose  $x = x_H$ , but not sufficient. Profits are higher at  $x = x_H$  than at  $x = x_L$  if the increase in the transactions fee from avoiding misreporting is less than the increased enforcement costs of avoiding misreporting. (Misreporting may either increase or decrease enforcement costs. If misreporting increases enforcement costs, then the for-profit exchange will always choose  $x = x_L$ , so that there is no misreporting.) The increase in transaction fees is less than the enforcement costs saving if

$$t_1 - t_2 = \pi_2 P_{F,1}(x_L) - \pi_2 P_{F,2}[(1-s) x_L + s x_H] < \pi_1 c (P_{F,1} - P_{F,2}) - \pi_2 c (1-s) P_{F,2}, \text{ or}$$

$$(A.7) \quad c[(\gamma_H - \gamma_L)\pi_1 - (1-s)x_L\pi_2] > (1-s)(\gamma_H - \gamma_L)\pi_2 x_L$$

If (A.7) is satisfied, then (A.5) will necessarily be negative. Hence, a necessary and sufficient condition for the PSRO to set  $x = x_H$  and allow misreporting is that (A.7) hold,

$$\text{or } s < \hat{s} = \frac{\pi_2 x_L (c + \gamma_H - \gamma_L) - \pi_1 (\gamma_H - \gamma_L) c}{\pi_2 x_L (c + \gamma_H - \gamma_L)}. \quad \square$$

**Proposition 4:** If conditions are such that the AIR can be satisfied, then  $P'_F \geq P'_M$ , and  $P'_F > P'_M$  for  $\gamma < \alpha - w_1$ ,

**Proof:** If  $\pi_2 (w_2 - w_1) - \pi_1 c > \alpha - w_1 > \gamma$  then

$$P'_M = \frac{\alpha - w_1 - \gamma}{\pi_2 (w_2 - w_1) - \pi_1 c}$$

and

$$P'_F = \frac{\pi_2 (w_2 - w_1) - \gamma}{\pi_2 (w_2 - w_1)}$$

so that  $P'_F \geq P'_M$  if  $\pi_2 (w_2 - w_1) [\pi_2 (w_2 - w_1) - (\alpha - w_1 - \gamma) - \pi_1 c] + \gamma \pi_1 c \geq 0$ , and this is non-negative whenever conditions are such that the AIR holds. When  $\pi_2 (w_2 - w_1) \geq \gamma \geq \alpha - w_1$ ,  $P'_M = 0$  and  $P'_F \geq 0$ .  $\square$

## CFS Working Paper Series:

No.	Author(s)	Title
2008/43	<b>Craig Pirrong</b>	The Industrial Organization of Execution, Clearing and Settlement in Financial Markets
2008/42	<b>Thorsten V. Koepl</b> <b>Cyril Monnet</b>	Central Counterparties
2008/41	<b>Terrence Hendershott</b> <b>Charles M. Jones</b> <b>Albert J. Menkveld</b>	Does Algorithmic Trading Improve Liquidity?
2008/40	<b>Jonathan Field</b> <b>Jeremy Large</b>	Pro-Rata Matching and One-Tick Futures Markets
2008/39	<b>Giovanni Cespa</b> <b>Thierry Foucault</b>	Insiders-Outsiders, Transparency and the Value of the Ticker
2008/38	<b>Rachel J. Huang</b> <b>Alexander Muermann</b> <b>Larry Y. Tzeng</b>	Hidden Regret in Insurance Markets: Adverse and Advantageous Selection
2008/37	<b>Alexander Muermann</b> <b>Stephen H. Shore</b>	Monopoly Power Limits Hedging
2008/36	<b>Roman Inderst</b>	Retail Finance: Bringing in the Supply Side
2008/35	<b>Dimitris Christelis</b> <b>Tullio Jappelli</b> <b>Mario Padula</b>	Cognitive Abilities and Portfolio Choice
2008/34	<b>Tullio Jappelli</b> <b>Marco Pagano</b>	Information Sharing and Credit: Firm-Level Evidence from Transition Countries

Copies of working papers can be downloaded at <http://www.ifk-cfs.de>