



IMF Working Paper

Competition among Exchanges and Enforcement Policy

Cecilia Caglio and Andrea Pescatori

IMF Working Paper

Research Department

Competition among Exchanges and Enforcement Policy

Prepared by Cecilia Caglio and Andrea Pescatori¹

Authorized for distribution by Thomas Helbling

February 2013

Abstract

This Working Paper should not be reported as representing the views of the IMF.

The views expressed in this Working Paper are those of the author(s) and do not necessarily represent those of the IMF or IMF policy. Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate.

In this paper, we explore how competition among stock exchanges, operated as self-regulatory organizations (SROs), affects the design of their members' surveillance. We develop a model where two for-profit SROs compete for trading volume, while brokers execute transactions on behalf of the investors and may misreport the true cash flow. The SROs can deter a fraud by announcing an investigation and imposing a monetary penalty. The success of the investigation depends upon both the amount of resources devoted to monitoring and the efficiency of monitoring technologies. We show that when contracts are incomplete and investors do not have perfect information about the monitoring efficiency, competition among exchanges induces a race to the bottom in enforcement policy and a reduction in total welfare, compared to the case of a monopolist SRO.

JEL Classification Numbers: G12; G14; G18; K22

Authors E-Mail Address: apescatori@imf.org; cecilia.r.caglio@frb.gov

Keywords: Stock exchanges; self-regulatory organizations; financial regulation

¹ Andrea Pescatori is an economist in the research department of the IMF; Cecilia Caglio is an economist at the Board of Governors at the Federal Reserve System. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System, any other person associated with it. We also would like to thank for helpful comments seminar participants at the Workshop on Money, Banking, and Payments (FRB Cleveland), FRB San Francisco, American University, GW University and the Securities and Exchange Commission and the participants of the FIRS 2010.

Contents

I.	The Model	5
	A. Investors and Brokers	5
	B. The SROs	6
	C. The Timing	7
	D. The Broker-Investor Contract: General Results	7
II.	The Race to the Bottom	9
	A. The Investor Problem	10
	B. The Broker Problem	11
	C. The SRO Problem and the Equilibrium	12
	1. The Monopolist Solution	12
III.	Fraud and Market Participation	14
	A. The Investor Problem	14
	B. The Broker's misconduct	16
	C. The SRO Problem and Equilibrium	17
	1. The Monopolist Solution under Incomplete Contracts	18
IV.	Extension: Sophisticated Investor	20
	A. Homogenous Sophisticated investors	20
V.	Conclusions	22
	A. Appendix 1	24
	B. Appendix 2	26
	1. The SROs' Problem	27

Tables

1.	Monopoly vs. Competition	19
2.	<i>Naive</i> vs <i>Sophisticated</i>	28

Figures

The recent international economic turmoil has forced to reconsider the effectiveness of financial regulatory systems and has highlighted the importance of the enforcement of regulation for financial stability. In this context, a prominent issue is how to design an institutional framework that helps safeguard the integrity and stability of financial markets. This paper takes a narrower focus and studies how the institutional characteristics of the security industry—that have varied across time and regions—can affect the enforcement of regulation and, in turn, investors’ market participation and welfare.

In the last few decades, the securities industry has experienced dramatic institutional and technological changes that have spurred competition among stock exchanges and induced the proliferation of new trading venues and platforms. While competition has probably lowered transaction fees and costs, the evolution of market centers from mutually-owned to profit-driven competitors¹ has raised concerns that the conflict of interest between the their regulatory function and their business operations could trigger a race to the bottom in market surveillance in order to attract trading activity and minimize regulatory costs.² These concerns are exacerbated by the increased fragmentation of the order flow across multiple markets which makes more difficult to detect and deter frauds and generate questions about the responsibility for trading activity across market centers. In the paper we address this issue by analyzing the effects of competition among stock exchanges on their enforcement policy and, in turn, on the functioning of securities markets as proxied by overall investor participation in the market and the likelihood that brokers may be induced in illicit conduct.³

To analyze the effect of competition on enforcement policy we extend and adapt the framework of DeMarzo et al. (2005) to a competitive setting. Specifically, we consider a model where two for-profit Self-Regulatory Organizations (SROs) compete to attract trading volume by setting the transaction fees and enforcement policy.⁴ While trading volume is ultimately related to investors’ willingness to trade, transactions are actually executed by brokers on the behalf of investors. Since investors do not observe the transaction cash flow, however, the broker-investor relation is distorted by asymmetric

¹Since the Stockholm Stock Exchange demutualized in 1993, changing its organizational form to a for-profit publicly listed organization, a considerable number of stock exchanges have took similar steps. For example, in the U.S. the Chicago Mercantile Exchange (CME) and Nasdaq went public in 2002, while NYSE demutualized in 2006, approximately 214 years after its founding.

²These concerns are supported by some evidence. For example, the Securities and Exchange Commission (SEC) imposed sanctions on a SRO that failed to pursue the wrongdoing by a member firm, alleging that the failure to address the misconduct was maybe related to the firm been a critical component of the SRO’s business model to generate revenue for the SRO (see Securities and Exchange Commission, Securities Exchange Act of 1934, Release No. 34-51163, February 9, 2005).

³For example, in a recent article, by studying the differences across 42 exchanges worldwide, Cummings et al. (2011) show market that having detailed rules can improve investors’ confidence and trading activity.

⁴The term self regulation in securities markets may be used to refer to formal SROs (i.e., securities exchanges with self-regulatory responsibilities), to common standards established by financial industry associations, and to the internal supervision and compliance functions within financial firms. This paper primarily examines self regulation performed by organizations with the power to regulate and oversee the activity of their members. The financial exchange industry around the world usually relies on SROs, such as the New York Stock Exchange (NYSE) and the Financial Industry Regulatory Authority (FINRA), to substantiate a regulatory system and guarantee its enforcement. In fact, SROs usually establish trading rules and listing standards, conduct surveillance of market operations, and oversee the activity of their members, i.e., broker-dealers.

information.⁵ The SRO where the transaction is executed performs the investigation to verify the accuracy of the broker's report to the investor. A successful probability of investigation is related to the amount of resources allocated by the SRO to the monitoring activity. Hence, in our model the role of the SRO's enforcement policy is to devote monitoring resources to investigate brokers' conduct, charge penalties in case of misconduct, and, thus, dissuade brokers from deceiving investors.⁶

We analyze the equilibrium under different hypotheses concerning investors' information sets. In the first part of the paper we assume that investors can perfectly observe the monitoring technology. When contracts are complete, SRO competition, compared to monopoly, enhances welfare because it reduces both transaction fees and monitoring resource, while still preserving maximum market participation and preventing brokers' misconduct. In other words, a monopolist SRO wastes resources by over-monitoring brokers to reduce their information rent to its own advantage.

The previous result is overturned when we relax the assumption of complete contracts. Specifically, in the second part of the paper we postulate that whereas brokers and SROs have perfect knowledge of the monitoring technology, investors are uncertain (i.e., hold different beliefs) about the effectiveness of SRO monitoring in detecting a fraud. By introducing non contractible uncertainty, we show that competition among exchanges induces a reduction in total welfare, compared to monopoly. The result is driven by the relaxation of enforcement policies and, consequently, by the increased number of frauds in equilibrium. This has two effects: first, the higher probability of fraud reduces investors' participation, and, by construction, total trading volume; second, there is a waste of resources due to the higher number of investigations and to the legal and opportunity costs associated with them.

Finally, we show that our results are robust to different assumptions on the investor ability to influence the broker's trading venue choice. While in the main setup we assume that naive investors do not internalize the effect of their actions on the broker's choice of the trading venue, we extend the model to include sophisticated investors. They can be thought of as institutional investors who might be in the position of inducing a broker to trade in their preferred trading venue. In this case, fees are the means of competition used by SROs to gain market share and attract trading volume. However, the reduction in the SROs revenues due to lower fees forces SROs to relax their enforcement policies making this equilibrium observationally equivalent to the one where SROs compete through enforcement policies.

Our paper makes two main contributions to the literature. First, we contribute to the

⁵To give a few examples of conflicts between brokers and customers, brokers may engage in excessive trading in a customer's account in an attempt to generate commissions (churning); a brokerage firm or a broker may fail to properly inform the investor or misrepresent facts regarding an investment (omission or misrepresentation); also, if a broker fails to make recommendations that are appropriate and suitable to his client's circumstances, the broker may be liable to that client (unsuitability). According to the Dispute Resolution Statistics compiled by FINRA, from 2006 to September 2011, 25% and 11% of the number of controversies served by FINRA involved misrepresentation or omission of the broker and unsuitability cases, respectively.

⁶As in De Marzo et al. (2005), enforcement is modeled in a two-tier framework. Technically, this is considered a case of *delegated auditing* that falls into the broader category of hierarchical contracting (see Faure-Grimaud et al. 1999, among others).

debate about competition and effectiveness of self-regulation in the financial industry. In particular, we show that when for-profit SROs compete for trading volume, the quality of enforcement policy is penalized. In a paper closely related to ours, Santos and Scheinkman (2001) conclude that competition does not necessarily imply a race to the bottom.⁷ In their model, traders may default and may differ in their credit quality. They show that when credit quality is observable, competitive intermediaries require the optimal amount of guarantees, whereas a monopolist demands fewer. If there is private information about credit quality, the result depends on the cost of default. Our analysis shows, on the contrary, that competition is likely to produce a less rigorous enforcement, reduce market participation, and increase the number of frauds, compared to a monopoly situation.

A standard result in the literature of self-regulating profession (Shacked and Sutton, 1981; Gehrig and Jost, 1995; DeMarzo et al., 2005) is that the cost of self-regulation is associated to some degree of monopoly power conferred to the members of the profession. Our second contribution is to show that this result applies also to a setting where SROs are independent for-profit competitors.⁸

The paper is organized as follows. In section 2, we present the model. In section 3, we analyze how SRO competition drives down enforcement policies compared to the monopoly case. In section 4 we introduce non-contractible uncertainty over SROs enforcement ability to analyze the impact of competition on frauds and investor participation. In section 5 we extend the model by considering the case of sophisticated investors. Section 6 concludes. All proofs are in the appendix.

I. The Model

We consider an economy with three classes of agents: investors, brokers, and two competitive and identical SROs. Each SRO owns and manages a trading venue where brokers can execute transactions on behalf of the investors. As in De Marzo et al (2005), we rely on the costly-state-verification framework of Townsend (1979) and Mookherjee and Png (1989) to model the broker-investor relationship by assuming that brokers have private information on the cash flow of the transaction and may misreport this information to deceive investors.

A. Investors and Brokers

There is a continuum of risk-neutral brokers and investors, both of measure 1. After being randomly matched with a broker, each investor offers a take-it-or-leave-it contract to the

⁷Other related papers are Foucault and Parlour (2004), who look at competition in listing fees, and Chemmanur and Fulghieri (2007), who focus on listing standards as a tool in competing with other exchanges for listings.

⁸DeMarzo et al (2005) prove that "*the control of the enforcement policy governing contracts confers substantial market power to a group of otherwise competitive agents (brokers)*". However, Reiffen and Robe (2008) show that this result is not robust to the ownership structure of the SRO (for-profit vs. mutual), since a single for-profit SRO does not operate to the benefit of its member (see also Hart and Moore 1996 and Pirrong 2000 for a discussion on the governance of exchanges).

broker to execute a trade that generates a random cash flow W (i.e., the value of the transaction). The offer is actually extended only if the investor's expected pay-off is at least equal to the investor's outside option α . Because we assume that the value of the broker's outside option is 0, the broker accepts any feasible contract.

The transaction's cash flow, W , is a binary random variable that has support $\Theta = \{\underline{w}, \bar{w}\} \subset \mathbb{R}_+$ with $\underline{w} < \alpha < \bar{w}$, and is independent of the SRO where the transaction takes place.⁹ Its realization w is costlessly observable only by the broker and takes the value \underline{w} with probability π .

For any given contract, the broker optimally chooses whether to execute the transaction on SRO $i = 1$ or $i = 2$ —SROs are indexed by $i \in \{1, 2\}$ or $i^* \in \{1, 2\} \setminus \{y|y = i\}$. Ex post, this choice is perfectly observable, thus, contractible, and captured by a binary random variable χ with support $\{1, 2\}$ and probability distribution $\{q_i\}_{i=1}^2$. Technically, χ is a trading venue indicator and the broker chooses the probability distribution of χ . Hence, in equilibrium q_1 represents the fraction of transactions executed on SRO 1, i.e., its market share.

After the transaction takes place, the investor receives a report $r(w)$ from the broker. Since the support of W is common knowledge, it is reasonable to assume that $r : \Theta \rightarrow \Theta$. In other words, any report $r \notin \Theta$ is trivially false.¹⁰

The menu of contracts available to the investor is represented by a class of functions \mathbb{Z} , where each function $z(r, i) \in \mathbb{Z}$ defines the contractual amount of cash that the investor receives from the broker after the transaction is executed, contingent on the broker's report $r(w)$ and the actual trading venue where the transaction took place. In other words, $z(r, i)$ is the investor's state-contingent net trading profit.

Brokers may have an incentive to misreport the actual realization of W ; however, each SRO devotes an amount m of resources to verify the veracity of the report and is able to enforce a penalty $x \geq 0$ to brokers, according to some pre-specified rules.

B. The SROs

Each SRO announces and commits to an enforcement policy (m_i, x_i) and a trading fee t_i before the investor and the broker negotiate the contract z . The probability of detecting a fraud, $p(m_i)$, depends on the amount of resources spent in investigating brokers' reports, $m_i \geq 0$. A convenient formulation for $p(m_i)$ is¹¹

$$p(m_i) = \max\{1 - \psi/m_i, 0\},$$

⁹Prima facie, a positive W seems to suggest that investors are always on the sell side. However, this is not the only interpretation: It is always possible to think of W as the investor's cash equivalent of the value of a buy-transaction.

¹⁰As it will be clear later on, in our framework the revelation principle does not necessarily hold, which may rehabilitate some alternative reporting strategies. Our assumption rules out this possibility that, in any case, does not seem appealing.

¹¹This is different from De Marzo *et al* (2005) which assume that the probability of an investigation is directly chosen by the SRO, while a fixed cost is paid by the SRO only in case of actual investigation. For a similar formulation see Greenwood *et al.* (2010).

where $1/\psi \geq 0$ represents the efficiency level of the monitoring technology. The cost function $C(m)$ associated with monitoring is increasing in the amount of resources devoted to monitoring, for simplicity we use the linear form

$$C_i(m) = cm,$$

where the constant c is the cost of a unit monitoring.

If a fraud is detected, the SRO i charges to the broker a monetary penalty, $x_i(r, w)$, which is a function of the cash flow reported by the broker $r(w)$ and the actually observed cash flow w . The penalty is returned to the investor net of a fraction γ lost in legal and opportunity costs.

Finally, contrary to De Marzo et al (2005), we assume that SROs are profit maximizing entities that are owned by outside owners and compete with each other to attract trading volume.

C. The Timing

The timing of events in the model can be summarized as follows:

1. Each SRO simultaneously chooses and commits to an enforcement policy (m_i, x_i) and a trading fee t_i .
2. Based on the announced enforcement policies and transaction fees $(m_i, x_i, t_i)_{i=1,2}$, the investor can either offer a contract z to a broker or take the reservation utility α .
3. If the investor offers the contract and the broker rejects it, the investor gets α , while the broker gets 0. If the contract is accepted, the broker chooses where to execute the transaction (by choosing the probability q_i).
4. The broker executes the transaction on the SRO i and privately observes the realization of W . He chooses the cash flow to report, r , and returns to the investor $z(r, i)$.
5. Given the report r , the SRO where the trade took place devotes $m(r)$ resources to investigating it. A fraud has a probability $p(m)$ of being detected. In case of detection, the SRO collects the penalty x from the broker and returns it to the investor net of a fraction γ lost in legal and opportunity costs.

D. The Broker-Investor Contract: General Results

As in DeMarzo et al. (2005), we initially introduce some general results for the contract between the broker and the investor that will simplify the derivation of the equilibrium conditions.

The investor maximizes his expected payoff choosing a contract $z \in \mathbb{Z}$, such that $z : \Theta \times \{1, 2\} \rightarrow \mathbb{R}$. The function $z(r, i)$ specifies the cash transfer from the broker to the investor for each trading venue i and broker's cash-flow report r . Implicitly, it defines the compensation paid by the investor to the broker for his intermediation service.

Before choosing a reporting strategy, the broker observes the enforcement policies (m_i, x_i) , the realization of the trading venue indicator χ and cash flow W , and transfers the transaction fee t_i to SRO i .¹²

The broker's feasibility constraint (BF) rules out the possibility of a strictly negative broker's payoff in any state of the world:

$$(BF) : z(r, i) \leq w - t_i, \quad \forall w, r \in \Theta, i \in \{1, 2\}. \quad (1)$$

Since any feasible contract implies that the broker's payoff $W - t - z$ is a non-negative random variable, the broker participation constraint (BIR) is always satisfied.

To simplify the notation we set $\underline{z}_i \equiv z(\underline{w}, i)$ and $\bar{z}_i \equiv z(\bar{w}, i)$, for $i = 1, 2$. In the rest of the paper, we assume that the broker's transfer to the investor is always higher in the high cash-flow state than in the low cash-flow state, i.e., $\underline{z}_i < \bar{z}_i$.¹³ It follows that it is never optimal for the broker to misreport the occurrence of the low cash flow—i.e., $r(\underline{w}) = \underline{w}$. Hence, the broker's feasibility constraint can be written as $\underline{z}_i \leq \underline{w} - t_i$, $i = 1, 2$, when the low cash flow occurs. We can now state the following lemma:

Lemma 1 *Given $\alpha > \underline{w}$, for each investor-broker pair, the broker's compensation for trading on SRO i and reporting the low cash flow, \underline{w} , is zero:*

$$\underline{z}_i = \underline{w} - t_i.$$

Because both SROs anticipate that $r = \bar{w}$ is always a truthful report and investigation is costly, they devote zero monitoring resources to investigate high cash-flow reports, $\bar{m} = 0$ and $p(r = \bar{w}) = 0$.

Given that $\bar{z}_i > \underline{z}_i$, the broker may have an incentive to report the low state when the high cash flow occurs. In fact, if no fraud is detected, he obtains a strictly positive payoff from misreporting: $\bar{w} - t_i - \underline{z}_i = \bar{w} - \underline{w} > 0$. However, there are two mechanisms that can prevent the broker from misreporting: the “carrot” of a compensation and the “stick” of a penalty. For instance, the investor can offer a positive compensation to the broker if the high state is reported; at the same time, the SROs can impose a penalty for misreporting and threaten to verify brokers' reports with some probability.¹⁴

¹²As we will see, the transaction fee is, ultimately, passed to the investor.

¹³This assumption is not strictly necessary and is made for tractability. Here we sketch the outline of a possible proof. Since $\alpha > \underline{w}$ and the broker cannot pay back the investor more than the realized cash flow, in the case of a single SRO it must be that $\bar{z} > \underline{z}$. By the same logic, it would be easy to prove that it is not possible to have $\bar{z}_i \leq \underline{z}_i \forall i \in \{1, 2\}$. Finally, by contradiction, if $\exists i \in \{1, 2\}$ such that $\bar{z}_i \leq \underline{z}_i$ the broker would choose to trade on SRO i with probability one (we will see this later on), violating the investor's participation constraint in equilibrium.

¹⁴Interestingly, contrary to standard CSV models (Townsend 1979 or Williamson 1987) those two actions are no longer taken by the same agent.

To formalize the argument of the carrot and the stick, we explicitly write the two relevant broker's incentive compatibility constraints:¹⁵

$$(BIC) : \quad \bar{w} - t_i - \bar{z}_i \geq p_i(\underline{w}) \max\{\bar{w} - \underline{w} - x_i, 0\} + \dots \\ \dots + [1 - p_i(\underline{w})](\bar{w} - \underline{w}), \quad i = 1, 2. \quad (2)$$

The LHS of equation (2) is the broker's compensation when the high state realization is reported truthfully (on SRO i)—which we denote by $\rho_i \equiv \bar{w} - t_i - \bar{z}_i$. The RHS represents the expected return from misreporting: This is a weighted average of the payoff obtained in case of detection, $\max\{\bar{w} - \underline{w} - x_i, 0\}$, and no detection of the fraud, $(\bar{w} - \underline{w})$.¹⁶

At the same time, to satisfy the investor's participation constraint both SROs must provide at least some minimal enforcement policy that prevents brokers from cheating. For this purpose, SROs can increase either the penalty or the monitoring (see equation 2). Unlike monitoring, however, increasing the penalty is costless. Hence, SROs always find optimal to set the penalty as high as possible. The limited liability assumption implies that the penalty x_i cannot exceed what the broker owns, i.e., $\max\{\bar{w} - \underline{w} - x_i, 0\} \geq 0$. It is useful to write the following lemma:

Lemma 2 *Given $\alpha > \underline{w}$, both SROs optimally set the penalty to its maximum: $x_i = \bar{w} - \underline{w}$.*

Using the previously stated lemmas and noting that the probability of *not* detecting a fraud is ψ/\underline{m} , we can rewrite equation (2) as

$$(BIC) : \quad v_i(\bar{z}_i) \equiv \frac{\bar{w} - \bar{z}_i - t_i}{\bar{w} - \underline{w}} \underline{m}_i \geq \psi \quad (3)$$

The function v_i characterizes the contract's ability to induce a broker to truthfully report a transaction that has taken place on SRO i . In other words, it captures the *incentive power* of a contract. It is worth noting that v_i is decreasing in \bar{z} , increasing in the broker's compensation, and increasing in \underline{m} . Finally, the knowledge of the monitoring technology, ψ , is crucial for designing a contract that is incentive compatible.

II. The Race to the Bottom

In this section, we study how competition can induce SROs to relax enforcement policies to attract trading volume in a basic setup with homogenous investors and complete contracts. One implication of these hypotheses is that in equilibrium investors always participate in the market and there is no fraud. The first result stems from the investor homogeneity assumption,¹⁷ the second from the fact that the revelation principle holds under the complete contracts hypothesis.¹⁸

¹⁵We do not allow negative penalties (prizes), and there is no penalty for truthful reports, i.e., $x_i(w, w) = 0$.

¹⁶In other words, if the BIC is satisfied, we have $r(\bar{w}) = \bar{w}$; otherwise, $r(\bar{w}) = \underline{w}$.

¹⁷Since investors are homogenous either all or none participate in the market. We will focus on the equilibrium where there is participation.

¹⁸In the basic setup we impose no restriction to the contract space (ψ is perfectly observable). Hence, the revelation principle applies and it guarantees that it is never optimal to offer a contract that would

A. The Investor Problem

Before offering a contract to the broker, the investor forms expectations about where the transaction will be executed. We define \tilde{q}_i to be the investor's subjective probability that $\chi = i$, to distinguish it from the objective probability q_i defined in section A.. We assume that the investor follows a *simple rule* that assigns probability 1 to the trade to be executed on the SRO with the lowest monitoring resources. More precisely, $\tilde{q}_i = 1$ for $\underline{m}_i < \underline{m}_{i^*}$ and $\tilde{q}_i = 1/2$ for $\underline{m}_i = \underline{m}_{i^*}$. As shown in the next section, this rule is consistent with the broker's optimal choice and in equilibrium $\tilde{q}_i = q_i$.

Implicitly, we are assuming that the investor does not design the contract strategically to induce brokers to trade on one of the two exchanges. In other words, an investor does not internalize his ability to induce the broker to trade on the SRO with higher monitoring resources.¹⁹ This simplifies the investor problem: Wherever the transaction takes place, the investor minimizes the broker's compensation subject to the broker's incentive compatibility constraint (see equation 3).²⁰ The solution of the investor problem is summarized by the following proposition.

Proposition 3 *Given $\underline{m}_i \geq \psi$, and $\tilde{q}_i = \tilde{q}(\underline{m}_i, \underline{m}_{i^*})$, the optimal contract is*

$$\bar{z}_i = \bar{w} - \rho_i - t_i, \quad i = 1, 2, \quad (4)$$

where the high-state optimal broker compensation ρ_i is given by

$$\rho_i = \frac{\psi}{\underline{m}_i}(\bar{w} - \underline{w}), \quad i = 1, 2. \quad (5)$$

It is worth noting that, the transaction fee is ultimately paid by the investor, while the broker compensation is inversely related to the effectiveness of the monitoring ($1 - \psi/\underline{m}_i$), i.e., the probability of detecting a fraud.

To derive the investor participation constraint we first introduce the investor's expected profit. This is a weighted average of four possible outcomes, i.e., low or high cash flow executed on SRO 1 or 2, with the probabilities \tilde{q} and π reflecting the expectations relative to the trading venue χ and cash-flow realization, respectively:²¹

$$\sum_{i=1,2} \tilde{q}_i [\pi(\underline{w} - t_i) + (1 - \pi)(\bar{w} - \frac{\psi}{\underline{m}_i}(\bar{w} - \underline{w}) - t_i)].$$

induce a fraud, i.e., where $v_i < \psi$ (see Mookherjee and Png 1989 for the revelation principle with feasibility constraints).

¹⁹In the extension of the model (see section 6), we relax this assumption and analyze the case where investors internalize the effect of their contracts on the broker's choice of the trading venue. However, as we will see, the main conclusions are unaltered.

²⁰More precisely, for all $\tilde{q}_i > 0$ the problem is isomorphic to maximize \bar{z}_i subject to the BIC constraint (3), for $i = 1, 2$. If $\tilde{q}_i = 0$ variables related to SRO i are irrelevant for the investor. However, without loss of generality, we can still take the BIC constraint as binding $v_i(\bar{z}_i) = \psi$.

²¹Notice that we use the results of lemma (1) and proposition (3) which stated that the low- and high-cash-flow broker compensation is zero and $\frac{\psi}{\underline{m}_i}(\bar{w} - \underline{w})$, respectively.

On the one hand, the investor's expected profit is increasing in \underline{m}_i because the more likely the investigations the lower the broker's compensation. On the other hand, because the broker's compensation is unaffected by transaction fees, the investor's profit is decreasing in t_i . Hence, rearranging the previous equation, we can write the investor (customer) individual rationality constraint (CIR) as

$$(CIR) : E[W] - \sum_{i=1,2} \tilde{q}_i [t_i + (1 - \pi)(\bar{w} - \underline{w}) \frac{\psi}{\underline{m}_i}] \geq \alpha, \quad (6)$$

where $E[W] = \pi \underline{w} + (1 - \pi) \bar{w}$ is the expected value of the transaction.

B. The Broker Problem

To choose the trading venue, brokers compare the expected utility U_i from trading on one SRO relative to the other. To derive U_i it is useful to analyze the high and low cash-flow realization separately. When the low cash flow realizes (with a probability π), there is no gain from misreporting and the broker pay-off is $\underline{w} - \underline{z}_i - t_i$. When the high cash flow realizes, the broker may expect to profit from misreporting. However, an investigation is successful with a probability p , in which case the broker's payoff is $\max\{\bar{w} - z_i(r(\bar{w})) - x_i(r(\bar{w}), \bar{w}) - t_i, 0\}$. With a probability $1 - p$ the investigation is not successful, which implies a payoff equal to $\bar{w} - z_i(r(\bar{w})) - t_i$. Hence, in its general form, the broker's expected utility derived from executing a trade on SRO i is

$$U_i \equiv \pi(\underline{w} - \underline{z}_i - t_i) + (1 - \pi)[p_i(\max\{\bar{w} - z_i(r(\bar{w})) - x_i(r(\bar{w}), \bar{w}) - t_i, 0\}) + \dots \\ \dots + (1 - p_i)(\bar{w} - z_i(r(\bar{w})) - t_i)]. \quad (7)$$

Because the revelation principle applies, we can replace $r(\bar{w})$ with \bar{w} in equation (7). Furthermore, independently of the trading venue, the low cash-flow compensation is zero while by assumption there are no prizes for reporting truthfully, i.e., $x_i(\bar{w}, \bar{w}) = 0$, while $x(\underline{w}, \bar{w}) = \bar{w} - \underline{w}$ (see Lemmas 1 and 2). Hence, we can simplify the expression for U_i to

$$U_i = (1 - \pi)(\bar{w} - \underline{w})\psi/\underline{m}_i. \quad (8)$$

We observe from equation (8) that a broker would strictly prefer to trade on SRO i if and only if the monitoring resources invested by the SRO i are less than the ones of its competitor. Hence, we can write the probability of executing a transaction on SRO i as (if $\underline{m}_i = \underline{m}_{i^*}$ we set $q_i = 1/2$)²²

$$q_i = \begin{cases} 1 & , \underline{m}_i < \underline{m}_{i^*} \\ 1/2 & , \underline{m}_i = \underline{m}_{i^*} \\ 0 & , \underline{m}_i > \underline{m}_{i^*}. \end{cases} \quad (9)$$

The broker's choice of the trading venue is only affected by the relative amount of monitoring resources spent by each SRO to investigate his activity; in fact, the lower the monitoring resources, the higher the compensation offered by investors to brokers. Accordingly, an SRO with lax enforcement policies is preferred by the broker because it allows her to best exploit the information rent.

²²If $\underline{m}_1 = \underline{m}_2$ the broker is indifferent about trading on one of the two SROs and any mixed strategy would be optimal. By the law of large numbers we should still expect that the market share will be half. Here, without loss of generality we have simply assumed that the best strategy is to pick $q = 1/2$.

C. The SRO Problem and the Equilibrium

SRO i chooses an enforcement policy and a transaction fee $(\underline{m}_i, x_i, t_i)$ anticipating investors' and brokers' optimal behavior, while taking the choice of SRO i^* $(\underline{m}_{i^*}, x_{i^*}, t_{i^*})$ as given. The objective of each SRO is to maximize total expected profits, which are given by the profit per transaction multiplied by the trading volume—i.e., the fraction of brokers trading on the SRO. The difference between the trading fee t and the expected cost of monitoring, $E[mc] = \pi \underline{m}c$, determines the profit per transaction.²³

The SRO i problem can be cast in the following form (notice that $\tilde{q}_i = q_i$)

$$\begin{aligned} & \max_{t_i, \underline{m}_i} \quad q_i(\underline{m}_i; \underline{m}_{i^*})[t_i - \pi \underline{m}_i c] \\ \text{s.t. } \quad (CIR) : & \quad q_1 t_1 + q_2 t_2 + (1 - \pi)(\bar{w} - \underline{w})\psi \left(\frac{q_1}{\underline{m}_1} + \frac{q_2}{\underline{m}_2} \right) \leq \tilde{\alpha}, \end{aligned}$$

where $\tilde{\alpha} \equiv E[W] - \alpha > 0$ by construction, is the total expected investor's surplus that stems from having the transaction executed. In other words, investors will not participate if the sum of the expected transaction costs and the compensation paid to the broker is greater than $\tilde{\alpha}$.

We can now determine the equilibrium described by the following proposition:

Proposition 4 *Given $\alpha > \underline{w}$ and $\tilde{\alpha} > \max\{\sqrt{(\bar{w} - \underline{w})c\psi}, 2\pi c\psi\}$, if a Nash equilibrium where the SROs get zero profits and $\tilde{q}_i = q_i$ exists, then it is a symmetric equilibrium—i.e., $q_i = 1/2$, $t_1 = t_2 = t$ and $\underline{m}_1 = \underline{m}_2 = \underline{m}$ —and it must satisfy:*

$$t = \pi c \underline{m} \tag{10}$$

$$\underline{m} = \frac{\tilde{\alpha}}{2\pi c} - \sqrt{\left(\frac{\tilde{\alpha}}{2\pi c}\right)^2 - \frac{(1 - \pi)\psi}{\pi c}(\bar{w} - \underline{w})}. \tag{11}$$

According to Proposition 4, competition for trading volume drives SROs' profits down to zero. When the investor is naive, competition operates through monitoring resources, which SROs reduce to gain or maintain their market share. The reduction in \underline{m} increases the brokers' compensation at the cost of investors' expected profits. However, trading volume ultimately depends on investors' participation; hence, SROs must also reduce the transaction fees to keep investors participating in the market.

1. The Monopolist Solution

To provide a benchmark for the solution with competitive SROs we consider the problem of a monopolist SRO.²⁴ The logic of the investor problem is identical to the competition case. Instead, for a monopolist SRO the problem can be written as

²³Recall that $\bar{m} = 0$ and the revelation principles implies no misreports.

²⁴Interestingly, the broker's compensation and the amount of monitoring resources found under monopoly replicate the non-hierarchical case where a representative investor/auditor (the principal) offers a take-it-or-leave-it contract to a representative broker (the agent).

$$\begin{aligned} \max_{m,t} \quad & t - \pi \underline{m} c \\ \text{s.t. (CIR)} \quad & t + (1 - \pi)(\bar{w} - \underline{w})\psi/\underline{m} \leq \tilde{\alpha}. \end{aligned} \tag{12}$$

The following proposition describes the enforcement policy and the transaction fee in equilibrium:²⁵

Proposition 5 *Under monopoly, given $\psi c \leq \frac{\tilde{\alpha}^2}{4(1-\pi)\pi(\bar{w}-\underline{w})}$, the optimal enforcement policy and transaction fee chosen by the SRO are*

$$\begin{aligned} m &= \sqrt{(\bar{w} - \underline{w}) \frac{(1 - \pi)\psi}{\pi c}} \\ t &= \tilde{\alpha} - \pi c m. \end{aligned} \tag{13}$$

Panel A and B in Figure 1 compare the monopoly and competition cases in the (m, t) and (p, t) space, respectively. In both panels, the region below the dark solid line satisfies the CIR constraint, while the dotted lines represent SRO iso-profits—the lowest iso-profit line is the zero profit locus. While the monopolist SRO maximizes its profits in M —where the tangent iso-profit line represents the highest feasible profit—competition drives SRO profits to zero. Hence, a symmetric equilibrium must lie at the intersection between the zero-profit line and the CIR, represented by the points B and N . However, B is clearly not a Nash equilibrium: SRO i can increase both unit profits and market share, by reducing monitoring resources. This strategy is feasible and profitable because (in B) the slope of the CIR is lower than the slope of iso-profits. Hence, competition pushes the equilibrium toward the point N where monitoring resources and fees are at the lowest. Once in N , there is no incentive to deviate by reducing the monitoring, because to induce investor participation unit profits would become negative.

Since the broker's compensation, represented by the solid grey line, is decreasing in the detecting probability, the broker receives a higher compensation under competition than under monopoly. In other words, because of competition the monopolist rent is transferred to brokers rather than to investors.

A simple measure of welfare \mathbb{W} is obtained by summing up the surpluses/profits of all classes of agents (investors, brokers, and SROs). Using the equilibrium conditions, we have that $\mathbb{W} = E[W] - \alpha - \pi \underline{m} c$ in both competition and monopolist case. Since \underline{m} is relatively high in the monopolist case, the welfare measure is higher under competition $\mathbb{W}^c - \mathbb{W}^m = \pi c(\underline{m}^m - \underline{m}^c) > 0$. This means that a monopolist SRO over-monitors brokers reducing their information rent at its advantage. This result is quite intuitive, since in equilibrium no frauds occur and full participation is obtained under both monopoly and competition.

²⁵The condition on parameters guarantees positive profits and a positive detecting probability.

III. Fraud and Market Participation

In this section, we show how the introduction of (even small) non-contractible investor uncertainty may induce broker frauds and reduce investors' participation, changing the welfare ranking described in the previous section.

Specifically, we relax the assumption that investors can perfectly observe the monitoring technology parameter, ψ , which is private information of SROs and brokers. Investor j , instead, sees ψ as a random variable, denoted Ψ , with a probability distribution $F(x; j)$. The average unconditional expectation of Ψ across investors, however, is still unbiased:²⁶

$$\int_0^1 \int_{-\infty}^{+\infty} x f(x; j) dx dj = \psi.$$

As shown previously, a contract's ability to satisfy the BIC constraint can be written as $\psi \leq v_i(\bar{z}_{ij})$ (see equation 3). Since under the new assumption ψ is no longer perfectly observable, investor j expects to satisfy the BIC constraint only with some probability. This probability is given by $F(v; j)$, which is the investor's belief of getting a truthful report for a given contract's incentive power v (see equation 3).

To simplify the notation, let's denote with $\tilde{\psi}_j$ the expectation over Ψ of investor j , that is $\tilde{\psi}_j \equiv \int_{-\infty}^{+\infty} x F(x; j) dx$. We sort investors such that investor θ is the most "optimistic" about the SROs monitoring efficiency, while investor 1 is the most "pessimistic", i.e., $\tilde{\psi}_{j'} < \tilde{\psi}_{j''}$, $\forall j' < j'' \in [0, 1]$. Finally, we assume that the median investor is unbiased $\tilde{\psi}_{j=1/2} = \psi$.

A convenient formulation for $F(x; j)$ is a uniform distribution $\Psi \sim U(\psi - 2a(1 - j), \psi + 2aj)$, such that

$$F(y; j) = \frac{y - \psi + 2a(1 - j)}{2a}, \quad (14)$$

for $y \in [\psi - 2a(1 - j), \psi + 2aj]$ and $\forall j \in [0, 1]$.

This formulation satisfies the conditions imposed on F and guarantees that the investor j 's expected value of the (inverse of the) enforcement technology parameter, $\tilde{\psi}_j = \psi + a(2j - 1)$, is increasing in j .

A. The Investor Problem

The uncertainty over ψ modifies the investor problem. In fact, for some investors it is now too costly to offer a contract that is incentive compatible, but not costly enough to prevent them from participating in the market.²⁷ In other words, in the new setup there is partial market participation and brokers may cheat investors in equilibrium.

²⁶While brokers have no incentive in truthfully revealing ψ , SROs may have. In fact, the distributions $F(x; j)$ can be thought of as the posterior distribution derived by investors with uninformative priors that receive a noisy signal over ψ sent by the SROs from some known distribution.

²⁷Technically, the contract offered is only constrained optimal. The underlying assumption is that it is too costly to overcome the informational friction through some contractual arrangements.

The investor's low cash-flow return is zero as in section 2.4. However, when the high cash flow realizes, the expected return incorporates the possibility of W being reported as a low cash flow. Conditional on the broker's misreport, the investor j 's expected return (on SRO i) is

$$[(1 - \gamma)x + \underline{z}_i]\tilde{p}_{ji} + \underline{z}_i(1 - \tilde{p}_{ji}), \quad i = 1, 2,$$

where $\tilde{p}_{ji} \equiv E_j[\max\{1 - \Psi/\underline{m}_i, 0\}|\Psi > v_i(\bar{z})]$ is the probability of a fraud being detected by SRO i conditional on investor j 's information set.²⁸ The first term, $[(1 - \gamma)x + \underline{z}_i]$, is the payoff when the fraud is actually detected, where $(1 - \gamma)$ is the fraction of the penalty returned to the investor, net of legal and opportunity costs. The second term is simply \underline{z}_i . The sum of the two terms simplifies to $(1 - \gamma)x\tilde{p}_{ji} + \underline{z}_i$.

The investor j 's problem reads as:

$$\begin{aligned} \max_{\bar{z}_i} \quad & F_j(v_i(\bar{z}_i))\bar{z}_i + [1 - F_j(v_i(\bar{z}_i))]\{(1 - \gamma)x\tilde{p}_{ji} + \underline{z}_i\} \\ \text{s.t.} \quad & \tilde{p}_{ji} = E_j[\max\{1 - \Psi/\underline{m}_i, 0\}|\Psi > v_i(\bar{z})] \\ & v_i(\bar{z}_i) = \frac{\bar{w} - \bar{z}_i - t_i}{\bar{w} - \underline{w}}\underline{m}_i. \end{aligned} \quad (15)$$

The two terms in the objective function capture the trade-off that the investor faces between reducing the broker's compensation (first term) and increasing the probability of misreport (second term).

The following proposition summarizes the general solution to the investor problem.

Proposition 6 $\forall \gamma \in (0, 1)$ and given $\underline{m}_i \geq \psi + 2a$, the investor j 's optimal high-cash-flow repayment is

$$\bar{z}_{ji} = \bar{w} - \rho_j(\underline{m}_i) - t_i. \quad (16)$$

The broker high cash-flow compensation from trading on SRO i is

$$\rho_{ji} = (\bar{w} - \underline{w}) \min \left\{ \frac{\gamma}{1 + \gamma} + \frac{\tilde{\psi}_j - a}{(1 + \gamma)\underline{m}_i}, \frac{\tilde{\psi}_j + a}{\underline{m}_i} \right\}. \quad (17)$$

As in the case of perfect knowledge of ψ , which is recovered when $a \rightarrow 0$, the broker's compensations is inversely related to monitoring resources.²⁹ However, when the uncertainty over ψ is sufficiently high some investors find too costly to offer a contract that satisfies the broker's incentive compatibility constraint with probability 1, i.e., $F_j(v_i) < 1$. These investors optimally choose to bear the risk of a fraud.³⁰ On the other

²⁸Conditional and unconditional probabilities over Ψ are not the same because the action chosen by the broker may reveal some information on ψ to the investor. For example, the investor knows that the broker misreports only when $1/\psi$ is lower than what he initially expected. Hence, conditional to a misreport, the expectation over Ψ is higher than the unconditional expectation.

²⁹It is also worth noting that the presence of legal and opportunity costs, γ , increases the broker compensation.

³⁰All investors are willing to take some risk only if $\underline{m}_i < \psi + 2a/\gamma, \forall j \in [0, 1]$. Hence, if $\gamma = 0$ no investor will be sure of offering an incentive compatible contract.

hand, overoptimistic investors (i.e., $j < 1/2$) underestimate the risk of a fraud and might offer a contract that is *actually* not incentive compatible. In fact, the investor j 's belief of offering an incentive compatible contract is decreasing with respect to j , while the incentive power of the contract actually increases. This means that a more optimistic view over Ψ makes the investor unawarely less cautious about the broker behavior and, possibly, exposed to fraud.

The following corollary summarizes the discussion above.

Corollary 7 *Under uncertainty over ψ , the broker compensation in the high cash-flow state determined in Proposition 6 can be expressed as*

$$\begin{aligned} \rho_{ji} &= (\bar{w} - \underline{w}) \left[\frac{\gamma}{1 + \gamma} + \frac{\tilde{\psi}_j - a}{(1 + \gamma)\underline{m}_i} \right]; \text{ iff } F_j(v_i) < 1 \\ \rho_{ji} &= (\bar{w} - \underline{w}) \frac{\tilde{\psi}_j + a}{\underline{m}_i}; \text{ iff } F_j(v_i) = 1 \end{aligned}$$

The presence of various degrees of optimism over Ψ has implications also for the investor participation constraint. Intuitively, the mass of participating investors is no longer 1 or 0, as in the previous section, but it is increasing in the monitoring resources \underline{m}_i and decreasing in the transaction fees t_i , for $i = 1, 2$. In fact, as monitoring falls and/or transaction fees rise, the most pessimistic investors no longer find optimal to participate in the market. The next proposition formalizes the result.

Proposition 8 *Under uncertainty over ψ , the mass of investors participating in the market is given by*

$$n = \min\left\{\max\left\{1/2 + \frac{\bar{\psi}^n - \psi}{2a}, 0\right\}, 1\right\}. \quad (18)$$

with $\bar{\psi}^n = B - \sqrt{B^2 - C}$, while coefficients B and C are functions of \underline{m}_i and t_i (see appendix).

The variable $\bar{\psi}^n$ is a cut-off value for the expectation over the monitoring efficiency. When $\tilde{\psi}_j > \bar{\psi}^n$ the investor j is pessimistic enough to expect his profits to be below the outside option α . This means that he will not offer a contract to the broker and will not participate in the market.

B. The Broker's misconduct

We now determine the mass of contracts that imply a fraud when the high cash flow realizes. If $F(v_i; j) = 1$, investors offer a contract that is incentive compatible by construction.³¹ Hence, we can focus on those investors for which $F(v_i; j) < 1$. Within this group, the investors that are over-optimistic with respect to ψ may offer a compensation

³¹Inspecting corollary 7, if $F(v_j) = 1$ even investor 0 would offer a $\rho_j = (\bar{w} - \underline{w})\psi/\underline{m}$, which, as we have seen, satisfies the BIC.

to the broker that is actually not incentive compatible. The mass of those contracts is given by:³²

$$\phi_i = \min\{\max\{1 - \gamma \frac{m_i - \psi}{2a}, 0\}, n\}, \quad i = 1, 2. \quad (19)$$

The variable ϕ_i is decreasing in monitoring resources and increasing in the uncertainty investors face concerning the monitoring efficiency.³³

Figure 2 summarizes the discussion of the section above. Panel A shows the high cash-flow trading profit for investor j 's repayment \bar{z}_j , while Panel B shows the corresponding expected earnings as a decreasing function of the signal $\tilde{\psi}$. The cut-off investor that is willing to participate in the market is determined by the intersection of the investor expected earning line with the reservation utility α . Panel D shows the contracts that may actually imply a misreport: A contract is incentive compatible only if the solid line, which represents v_j , is higher than ψ —i.e., when the broker's return from not misreporting is higher than the expected return from misreporting. Hence, the light and dark thick solid lines on the x-axis represent the mass of contracts that may and will not imply a fraud, respectively. On the same x-axis, the dark thick dashed line represents the mass of investors that do not participate in the market.

C. The SRO Problem and Equilibrium

Compared to the SRO problem in section C., the objective function of each SRO is now modified by the fact that full participation is not guaranteed and some contracts may actually imply a misreport with the consequence of increasing total investigation costs. In fact, in the equilibrium without frauds the mass of investigations is equal to the low cash-flow probability, π . However, if some high cash-flow state are reported as low the total mass of investigation will increase. By contrast, since brokers can perfectly observe ψ and their compensation is not a function of transaction fees, the brokers' choice of the trading venue has the same form as in equation (9). This choice is independent of the cash-flow realization and, thus, of the broker's reporting strategy.³⁴

The SRO i problem reads (symmetrically for SRO i^*) as

$$\max_{\underline{m}_i, t_i} q[n(t_i - \pi c \underline{m}_i) - (1 - \pi)\phi_i c \underline{m}_i], \quad (20)$$

where it is understood that q , n , and ϕ_i are functions of \underline{m}_i and t_i , and parameterized with respect \underline{m}_{i^*} and t_{i^*} (see equations 9 and 18-19). The following proposition describes the equilibrium.

³²Using the result of corollary 7 to substitute for ρ_i in equation (3), a contract does not satisfies the BIC constraint when $v_{ji} = \frac{\gamma \underline{m}_i + \psi_j - a}{1 + \gamma} > \psi$. Integrating with respect to j gives equation (19).

³³With full participation, there is a simple relation between the average investor's probability of offering a compatible contract and the actual mass of non-compatible contracts, ϕ , which is $\phi = 1 - F + \gamma(1/2 - F)$. Legal costs disconnect the average investor's belief and the actual broker's behavior by skewing the \bar{z} distribution.

³⁴Comparing the broker's compensation from trading on SRO i , we easily see that $\rho_{j1} > \rho_{j2}$ iff $\underline{m}_1 < \underline{m}_2$. Hence, whenever the optimal report is truth-telling, the broker chooses the SRO with lower monitoring. In case of misreporting, the broker's expected return is $\psi(\bar{w} - \underline{w})/\underline{m}_1 > \rho_{j1} > \rho_{j2}$ iff $\underline{m}_1 < \underline{m}_2$. Hence, regardless of the reporting decision, it is always optimal for the broker to choose the SRO with the lowest monitoring.

Proposition 9 *Given $\gamma \in (0, 1)$, if a Nash equilibrium where $n < 1$, $\phi_i < n$, and $\phi_i > 0$ exists, then it is a symmetric equilibrium—i.e., $q_i = 1/2$, $t_1 = t_2 = t$ and $\underline{m}_1 = \underline{m}_2 = \underline{m}$ —and it must satisfy*

$$n(t - \pi c \underline{m}) = (1 - \pi) \phi c \underline{m}$$

with $\phi = 1 - \gamma \frac{m - \psi}{2a}$ and $n = -n_t(t - \pi c \underline{m})$, where

$$\begin{aligned} n_t &= -\frac{1 + \gamma}{(1 - \pi)(\bar{w} - \underline{w})\gamma^2 \sqrt{B^2 - C}} \underline{m} \\ B &= \underline{m} + a(2\gamma^{-2} - 1) \\ C &= \underline{m}^2 + \frac{4a}{\gamma^2} \left[1 - \gamma^2/2 + (1 - \gamma) \frac{w - \alpha - t}{(1 - \pi)(\bar{w} - \underline{w})} \right] \underline{m} + a^2. \end{aligned}$$

Proposition 9 shows that SROs still earn zero profit in equilibrium, however, differently from section C., each SRO incur in higher costs because of frauds. The term $(1 - \pi) \phi c \underline{m}$ represents the expected investigation costs associated with frauds. Moreover, investor participation is below 1, which implies a reduction in total market volume and revenues. Finally, we also notice that since overoptimistic investors are the last investors to prefer the outside option, as market participation falls the mass of fraud ϕ_i is not affected until $n = \phi_i$ (once $n = \phi_i$ no contract is incentive compatible). Hence, a reduction in investor participation never reduces the ratio of frauds over total transactions.

1. The Monopolist Solution under Incomplete Contracts

We compare the previous result with the monopolist case. A monopolist SRO solves the same problem as in (20), however, the market share is fixed, while frauds and participation, ϕ and n , depend only on choices of \underline{m} and t . The solution of the monopolist problem is described by the following proposition:

Proposition 10 *Given $\frac{\tilde{\alpha}^2}{1 - \pi} > 4(\bar{w} - \underline{w}) \frac{(\psi + 3)^2}{\psi + 2a} \pi c$ and $\gamma > \frac{2a}{-2a - \psi + \sqrt{(1 - \pi)(\bar{w} - \underline{w}) \frac{\psi + 2a}{\pi c}}}$, for a monopolist SRO, the equilibrium is characterized by the following system of equations*

$$\begin{aligned} \underline{m} &= \sqrt{(1 - \pi)(\bar{w} - \underline{w}) \frac{\psi + 2a}{\pi c}} \\ t &= \tilde{\alpha} - \pi c \underline{m} \\ \bar{z}_j &= \bar{w} - \frac{\tilde{\psi}_j + a}{\underline{m}} (\bar{w} - \underline{w}) - t, \end{aligned}$$

which implies $n = 1$ and $\phi = 0$.

The monopolist SRO fully internalizes the effect of its enforcement policy and transaction fee on investor participation. Moreover, the market share no longer plays a role in this case. Hence, under the conditions imposed on the parameters in proposition 10, transaction fees and monitoring resources are set such that the *last* investor (i.e., the most

pessimistic) is still participating in the market. Greater uncertainty over ψ (i.e., a higher a), increases \underline{m} and negatively affects SRO's profits. In fact, a higher a has a negative effect on investor expected profits that the SRO offsets by toughening enforcement policies and reducing transaction fees to preserve the full investor participation in the market. Finally, since fraud generates extra investigation costs, a monopolist SRO optimally chooses monitoring resources to guarantee that the most optimistic investor offers an incentive compatible contract.³⁵

Table 1. Monopoly vs. Competition

Description	Monopoly	Competition	Ratio %
Detecting Probability	0.82	0.75	- 8.5
Transaction Fee over $E[W]$	0.55%	0.41%	-25.5
Total Volume (n)	1	0.955	- 4.5
Total Frauds	0	0.059	-
SROs' Profits	0.124	0	-
SRO 1 Market Share	-	50%	-
Brokers' Average Profits	0.152	0.524	24.5*10
Investors' Actual Average surplus	0.287	0.047	-16.4
Investors' Expected Average surplus	0.287	0.046	-16.4
Total Monitoring Costs	0.438	0.421	- 3.9
Total Welfare	0.563	0.545	- 3.2

Columns 1 and 2 show the results of the calibrated example under monopoly and competition, respectively. The average investor's surplus is defined as the difference between the average investor's expected (actual) profits and the reservation profit. The last column is the ratio of column 2 over 1 in percent.

Table 1 presents a calibrated numerical example to compare results under competition and monopoly when ψ is not perfectly observable.³⁶ Under competition the probability of detecting a fraud drops by about 9% from 0.82 to 0.75, relatively to monopoly, reflecting the lower resources spent in monitoring. This is coupled with a 25% reduction in transaction fees. However, the benefits of lower fees are not enjoyed by investors which sees their profits to drop dramatically, by almost 85% (from 0.287 to 0.047),³⁷ while brokers increase their average profits more than 3 folds. Because from the investor point of view the reduction in enforcement policies outweighs the benefits of lower transaction fees, some of them will not participate in the market entailing a reduction in the total transaction volume (by about 5%). At the same time, under competition about 6% of

³⁵When a is sufficiently high, the conditions of proposition 10 do not hold anymore. This implies that even in the monopolistic case, we may have a participation lower than 1 and frauds. In any case, the effect of competition would be to further reduce participation and increase frauds, leaving the qualitative results unchanged. We focus on the full participation case.

³⁶In the numerical example we take the spread between high and low cash flow to be $\bar{w}/\underline{w} - 1 = 2.5\%$, the reservation utility to be 1% higher than the low state $\alpha/\underline{w} - 1 = 1\%$, and the low state probability π equal to 20%. We cannot identify c and ψ separately, hence we fix $\psi = 10$ and set the cost parameter c such that the unit monitoring cost cm is 0.8% of the average value of a transaction, when the detecting probability is 50%. This implies $c = 0.04$. The legal costs are set at 6% of the penalty, and the noise a/ψ is set equal to 10%. Total transaction surplus is $\tilde{\alpha} = 1$.

³⁷The sign of the difference between actual and expected investor profits is determined by the shape of the distribution of \bar{z}_j over investors.

total transactions involves a fraud and, given a 0.75 detecting probability, only about 4.4% of frauds will be actually detected.

The first best social welfare is simply given by the expected surplus $\tilde{\alpha}$ derived from all of the transactions. Under monopoly the social welfare is equal to $\tilde{\alpha}$ minus total monitoring costs. In the previous section we showed that the over-monitoring of the monopolist would lead to a lower social welfare. However, with incomplete contracts, competition among SROs reduces *total* monitoring costs only slightly, about 4% from 0.44 to 0.42. This is because the reduction in resources spent *per* investigation is offset by the increase in the number of investigations due to the presence of frauds.³⁸ The small reduction in monitoring resources does not compensate for the lower investor participation, $n = .95$, that by definition reduces the total expected surplus. This implies a reduction of social welfare under competition by about 3.2% relatively to monopoly, $\mathbb{W}^m = 0.56 > \mathbb{W}^c = 0.55$.

IV. Extension: Sophisticated Investor

In this section, we extend the previous setup to show that it is robust to the assumption that investors are sufficiently small to internalize the effect of their contract on the broker's trading venue choice. In particular, we study the equilibrium under the assumption that investors are *sophisticated*, that is, they anticipate the broker's strategic choice of the trading venue. In other words, investors maximize their expected payoff strategically inducing the broker to trade on the exchange where transaction fees are the lowest. Consequently, SROs reduce transaction fees to attract trading volume. This generalization is useful because it shows that the results of the previous sections do not depend on whether SRO competition operates through transaction fees or enforcement policies. In fact, because monitoring is costly, a race to the bottom for transaction fees forces SROs to devote less resources to investigating potentially illicit conducts, which, in turn, leads to higher broker profits.

A. Homogenous Sophisticated investors

The assumption of *sophisticated* investors does not alter two important results obtained in the previous setting. First, the broker has no incentive to report a high cash-flow when a low cash-flow is realized, given that $\bar{z}_i > \underline{z}_i$; hence, because investigating is costly, $\bar{p}_i = 0$, i.e., $\bar{m}_i = 0$. Second, the low state contract is still pinned down by the BF constraint, $\underline{z}_i = \underline{w} - t_i$.

As in the case of *naive* investors, the probability of the transaction being executed on SRO i , q_i , depends on the broker's utility derived from trading in the two SROs, as described in equation (7). However, differently from before, q_i is no longer taken as given by the investor.

³⁸Without frauds, only low cash-flow reports are investigated. Under competition, the existence of mis-reports implies that some high cash flows are reported as low, increasing the number low-cash flow reports and, thus, investigations.

For convenience, we formulate the investor's decision problem as a function of the broker's compensation, ρ :

$$\begin{aligned} \max_{\rho_i, q} &= E[W] - q_1[(1 - \pi)\rho_1 + t_1] - (1 - q_1)[(1 - \pi)\rho_2 + t_2] \\ q_i \rho_i &\geq q_i \hat{\rho}_i \\ q_1 &= \begin{cases} 1, & \rho_1 \geq \max\{\rho_2, \hat{\rho}_2\} \\ 1/2, & \rho_1 = \rho_2 \\ 0, & \rho_2 \geq \max\{\rho_1, \hat{\rho}_1\}, \end{cases} \end{aligned} \quad (21)$$

where $\hat{\rho}_i \equiv (\bar{w} - \underline{w})\frac{\psi}{m_i}$ is the broker's expected return from misreporting on the SRO i .³⁹ We can now state the following proposition that describes the investor optimal contract.

Proposition 11 *The optimal contract for a sophisticated investor with no uncertainty regarding the SRO's monitoring technology, $a = 0$, is*

$$[\rho_1, \rho_2] = \begin{cases} [(\bar{w} - \underline{w}) \max_i \{\frac{\psi}{m_i}\}, 0], & t_1 < t_2 \\ [(\bar{w} - \underline{w}) \max_i \{\frac{\psi}{m_i}\}, (\bar{w} - \underline{w}) \max_i \{\frac{\psi}{m_i}\}], & t_1 = t_2 \\ [0, (\bar{w} - \underline{w}) \max_i \{\frac{\psi}{m_i}\}], 0, & t_1 > t_2 \end{cases} \quad (22)$$

and

$$q_1 = \begin{cases} 1, & t_1 < t_2 \\ 1/2, & t_1 = t_2 \\ 0, & t_1 > t_2 \end{cases} \quad (23)$$

The intuition for Proposition 11 is that *sophisticated* investors can affect the broker's trading choice by differentiating the compensation offered to the broker to trade on one SRO relatively to the other. In particular, to induce the broker to trade on the exchange with the highest monitoring, the *sophisticated* investor needs to offer a contract whose compensation is no lower than the maximum profit the broker could obtain by misreporting on each SRO. Hence, the lower bound for the broker's compensation is set equal to the compensation he expects by choosing the SRO with the lowest enforcement.

On the other hand, transaction fees are ultimately paid by the investors and thus affect their expected return. Consequently, it is optimal for the investor to induce the broker to trade on the SRO with the lowest transaction fees. In fact, by comparing how trading volume is determined in presence of sophisticated investors (proposition 11) with *naive* investors (see equation 9), we note that in proposition 11 q is a function of the transaction fees instead of the monitoring resources.

The investor's participation constraint depends on the SRO in which the trade takes place. We can write it as:

$$(CIR) : t_i + (1 - \pi) \max_i \{\hat{\rho}_i\} \leq \tilde{\alpha}, \quad t_i < t_{i^*}. \quad (24)$$

³⁹In writing q_1 , we have assumed that for the same expected utility the broker prefers a truthful report. For convenience, we remind that the unconditional expected cash-flow is $E[W] = \pi\bar{w} + (1 - \pi)\underline{w}$. Finally, the investor objective function already includes the result $\underline{z}_i = \underline{w} - t_i$.

In addition, each SRO's problem is to maximize its profits taking the enforcement policy and transaction fee of the competitor as given. Hence, SRO 1 maximizes $q_1(t_1 - \pi c \underline{m}_1)$, given q_1 as described in proposition 11 and the CIR (equation 24). The next proposition describes the symmetric equilibrium.

Proposition 12 *In presence of sophisticated investors, there exists a unique symmetric equilibrium with $q = 1/2$, $t_1 = t_2 = t$, and $\underline{m}_1 = \underline{m}_2 = m$, where m and t must satisfy*

$$\begin{aligned} \pi c m + (1 - \pi)(\bar{w} - \underline{w}) \frac{\psi}{m} &= \tilde{\alpha} \\ t &= \pi c m. \end{aligned}$$

We have obtained the same equilibrium values as in the basic model with *naive* investors. The intuition is that competition, through monitoring or transaction fees, drives SROs' profits to zero. Moreover, it is never optimal for an SRO to leave slackness in the investors' participation constraint (CIR), i.e., SROs minimize monitoring resources, thus, boosting brokers' compensation and reducing investors' surplus. The two conditions, CIR binding and zero profits, are the same in the case of both *sophisticated* and the *naive* investors.

Introducing uncertainty over ψ leads to results very similar to the one described in Table 1. The intuition is the same as the one just given: the means of competition is irrelevant. However, since there is no closed-form solution to the investor problem we have relegated the description and numerical solution of this further extension to Appendix B.

V. Conclusions

We have modified extended the hierarchical costly state verification framework used by De Marzo et al. (2005) to analyze the effect of competition among Self Regulatory Organizations (SROs) such as stock exchanges. Competition among for-profit SROs implies a reduction in both transaction fees and resources devoted to monitor and investigate brokers' trading activity. While lower fees benefit investors, reduced monitoring resources advantage brokers by increasing their information rent. In this framework competition is welfare enhancing, compared to monopoly, because a for-profit monopolist SRO would waste resources by over-monitoring brokers to reduce their information rent to its own advantage. When we relax the assumption of complete contracts, however, we demonstrate that competition among exchanges reduces total welfare relatively to monopoly. The result is driven by the reduction in enforcement policy and, consequently, by the presence of brokers' fraud in equilibrium: first, this causes an overall decline in investors' participation, and by definition in total trading volume; second, due to the higher number of misreports, there is a waste of resources due to the increased number of investigations and the associated legal and opportunity costs.

Finally, results are robust to different assumptions on the investor's ability to influence the broker's trading venue choice. In the main setup we assume that investors are *naive*, in the sense they do not internalize the effect of their actions on the broker's choice of the trading venue. When we extend the model to include sophisticated investors, fees are the

means of competition used by SROs to gain market share and attract trading volume. However, the reduction in the SROs revenues due to lower fees forces SROs to relax their enforcement policies making this equilibrium observationally equivalent to the one where SROs compete through enforcement policies.

Appendices

A. Appendix 1

Here we present the proofs of the lemmas and propositions stated in the text.

Proof of Lemma 1. First of all we show that $r(\underline{w}) = \underline{w}$. Given $\bar{z}_i \geq \underline{z}_i$ if $r(\underline{w}) = \bar{w}$, BF constraints would imply that $\bar{z}_i \leq \underline{w} - t_i$ which cannot be an equilibrium when $\alpha > \underline{w}$. Given $r(\underline{w}) = \underline{w}$, by the same argument, it must be that, in equilibrium, $\bar{z}_i > \underline{z}_i$, for some i (no pooling).

When $r(\underline{w}) = \underline{w}$, the low cash-flow BF constraint implies $z_i(\underline{w}) \leq \underline{w} - t_i; \forall q_i > 0$, investors' expected returns are strictly increasing in $z_i(\underline{w})$, hence, $z_i(\underline{w}) = \underline{w} - t_i$ must be the unique solution of the investor's problem. When $q_i = 0$, $z_i(\underline{w}) = \underline{w} - t_i$ is clearly still a solution. ■

Proof of Lemma 2. By contradiction assume $(\underline{m}_1, x_1, t_1)$ is such that $x_1 < x_1^{max}$, for some given $(\underline{m}_2, x_2, t_2)$. Take a new triple $(\underline{m}'_1, x'_1, t'_1)$ such that $x_1 < x'_1 \leq x_1^{max}$ and $\underline{m}_1 < \underline{m}'_1$. It is clearly true that $t'_1 = t_1$ is feasible for SRO 1. We want to show that we can find some x and m such that $q' = q$. To show this is sufficient to demonstrate that the (BIC)' is equivalent to the previous (BIC). From equation (2) we realize that this is easily accomplished by setting \underline{m}'_1 such that $(1 - \psi/\underline{m}'_1)x_1 = (1 - \psi/\underline{m}_1)x'_1$. This guarantees that the investor is solving the same problem. Hence, given that $q' = q$, $t'_1 = t_1$ is generally feasible, and $\underline{m}'_1 < \underline{m}_1$, it must be that $(\underline{m}_1, x_1, t_1)$ does not maximize SRO 1 profits. ■

Proof of Proposition 3. Given that the revelation principle holds, the BIC constraint has to be satisfied. Since the broker's compensation for trading on either SROs are independent, the BIC constraint will hold with equality for $\tilde{q}_i \in (0, 1)$ for both SROs. This uniquely determines \bar{z}_i when $\tilde{q}_i > 0$. When $\tilde{q}_i = 0$ the choice of \bar{z}_i becomes irrelevant, however, we can still determine \bar{z}_i using the BIC constraint. ■

Proof of Proposition 4. The CIR is always binding. Say $q \geq 1/2$ and the CIR is not binding. Given SRO 2 actions $\exists t'_1 > t_1$ and $\underline{m}'_1 = \underline{m}_1$ such that the CIR is still satisfied, $q' = q$, and SRO 1 enjoys higher revenues and the same costs.

SROs' profits must be zero. Say SRO 1 enjoys strictly positive profits and $q = 1/2$ (we only prove the most interesting case).⁴⁰ We must have $\underline{m}_1 = \underline{m}_2$, as a matter of exposition let us choose $t_1 > t_2$ —i.e., SRO 1 makes more profits. By continuity, \exists neighborhood I of (m, t) such that unit profits are strictly positive; hence, \exists some $\underline{m}'_2 < \underline{m}_1$ and $t'_1 \leq t_1$ such that the CIR is still satisfied and $(\underline{m}'_2, t'_2) \in I$.

The symmetric equilibrium is unique.⁴¹ The CIR, combined with the zero profit condition, implies $\pi cm - \tilde{\alpha}m + (1 - \pi)(\bar{w} - \underline{w})\psi = 0$. There are two candidate solutions, $m^{(1)}$ and $m^{(2)}$. Say $m^{(1)} < m^{(2)}$ then $m^{(2)}$ is not a solution. It would be optimal to reduce both the monitoring and the transaction fee (see monopolist case and figure 1). ■

Proof of Proposition 5. The proof of proposition (5) is simple and thus omitted. ■

⁴⁰The case $q_i = 1$ is straightforward. SRO i' would be better off choosing $(\underline{m}_{i'}, t_{i'}) = (\underline{m}_i, t_2)$ and getting half the profits of SRO 1.

⁴¹It is always possible to have an equilibrium in which $q = 1$ or $q = 0$, the CIR is satisfied and there are zero profits.

Proof of proposition 6. We focus on the case where $\underline{m}_i \geq \psi + 2a$ $i = 1, 2$, this guarantees that

$$E_j[\max\{1 - \Psi/\underline{m}_i, 0\}|\Psi > v_i(\bar{z})] = 1 - E_j[\Psi|\Psi > v_i(\bar{z})]/\underline{m}_i.$$

We will use the general result that $E[x|x > a] = \int_a^\infty uf(u)du/(1 - F(a))$. We first find the interior solution, $0 < F < 1$, by applying the calculus of variations. We can cast the investor j problem as (we drop subscripts i and j)

$$\max_{\bar{z}} F(v(\bar{z}))\bar{z} + [1 - F(v(\bar{z}))][(1 - \gamma)x + \underline{z}] - (1 - \gamma)x[G(b) - G(v(\bar{z}))]/\underline{m}$$

where $b = \psi + 2aj$. We have defined $g(u) \equiv uf(u)$ to have $\int_v^b uf(u)du = G(b) - G(v)$. Finally, taking derivatives with respect \bar{z} and recalling that $v' = -m/x$, $F' = v'f$ and $G' = v'vf$, we find the necessary optimality condition for an interior solution satisfies $\gamma(\underline{m} - v)f(v) = F(v)$. Substituting the functional forms of the uniform distribution we have

$$\rho = (\bar{w} - \underline{w})\left[\frac{\gamma}{1 + \gamma} + \frac{\tilde{\psi}_j - a}{(1 + \gamma)\underline{m}}\right]$$

For the boundary cases, note that $\rho = (\psi + 2aj)/\underline{m}$ is the solution obtained imposing $F = 1$, while imposing $F = 0$ makes the investor's return independent of \bar{z} . ■

Proof of Proposition 8. Using proposition 6, we can write investor j 's expected profit, given by (15), as a function of the SRO's announced enforcement policies and transaction fees, and the expectation over Ψ . Because investors area ranked according to $\tilde{\psi}_j$, it is convenient to express the expected profit as a function of $\tilde{\psi}_j$: $A\tilde{\psi}_j^2 - 2BA\tilde{\psi}_j + CA$, where A , B , and C are functions of \underline{m}_i and t_i : $A = (1 - \pi)(1/m_2 + 1/m_1)\gamma^2/a/k/(1 + \gamma)/8$; $B = 2m_1m_2/(m_1 + m_2) + a(2\gamma^{-2} - 1)$; $C = 8a(k/(1 - \pi)/\gamma^2(\underline{w}\gamma + \bar{w} - (1 + \gamma)(\alpha + t_1/2 + t_2/2)) - 0.5 - \pi/(1 - \pi)/\gamma^2)m_1m_2/(m_1 + m_2) + m_1m_2 + a^2$; where $k = 1/(\bar{w} - \underline{w})$.

The investor participation constraint requires that $A\tilde{\psi}_j^2 - 2BA\tilde{\psi}_j + CA - \alpha \geq 0$. Hence, equating this last object to zero determines a cut-off value, $\bar{\psi}^n$, above which no investor participates. The cut-off value is the solution of the quadratic equation in $\tilde{\psi}$; i.e., $\bar{\psi}^n = B - \sqrt{B^2 - C}$. Because investors with $\tilde{\psi}_j > \bar{\psi}^n$ do not participate, we can determine total participation by integrating over j up to the cut-off investor gives $n = \min\{\max\{1/2 + \frac{\bar{\psi}^n - \psi}{2a}, 0\}, 1\}$. ■

Proof of proposition 9. The interior solution of the constrained maximization of (20) subject to equations (9), (18) and (19) parameterized with respect \underline{m}_2 and t_2 delivers the stated proposition. ■

Proof of Proposition 10. The optimal investor contract is given by $\bar{z}_j = \bar{w} - \rho_j - t$, where the broker's compensation has been defined as

$$\rho_j = (\bar{w} - \underline{w}) \min\left\{\frac{\gamma}{1 + \gamma} + \frac{\tilde{\psi}_j - a}{(1 + \gamma)\underline{m}}, \frac{\tilde{\psi}_j + a}{\underline{m}}\right\}.$$

The incentive power of the contract is given by $v_j = \underline{m}\rho_j(\bar{w} - \underline{w})^{-1} = \min\left\{\frac{\gamma\underline{m} + \tilde{\psi}_j - a}{1 + \gamma}, \tilde{\psi}_j + a\right\}$

Assume that $\underline{m} > \psi + 2a/\gamma$, then we have $\rho_j = \frac{\tilde{\psi}_j + a}{\underline{m}}(\bar{w} - \underline{w})$ and $v_j = \tilde{\psi}_j + a, \forall j \in [0, 1]$. This also implies that $\tilde{s}_j = 1$ and $\phi = 0$ (no frauds). Integrating over investors, we find the participation $n = \frac{\tilde{\alpha} - t}{\bar{w} - \underline{w}} \frac{(\bar{w} - \underline{w})\underline{m}}{2a(1-\pi) - \psi/2a}$. The SRO's problem reduces to maximize $n(t - \pi c \underline{m})$, with $0 \leq n \leq 1$. When $n = 1$ we have the result of the proposition. Notice that for $n \in [0, 1]$ optimality implies $t = \tilde{\alpha} - \pi c \underline{m}$. We can rewrite the problem as a one-variable maximization problem and the condition $\frac{\tilde{\alpha}^2}{1-\pi} > 4(\bar{w} - \underline{w}) \frac{(\psi+3)^2}{\psi+2a} \pi c$ is determined by requiring the Lagrange multiplier on $n \leq 1$ to be strictly positive. ■

Proof of proposition 11. Set $q = 1$. It must be $\rho_1 \geq \max\{\rho_2, \hat{\rho}_2\}$. Conditioned on $q = 1$, the BIC implies $\rho_1 \geq \hat{\rho}_1$. Because the investor wants to minimize the broker's compensation, it is optimal to choose $\rho_2 < \rho_1$, say $\rho_2 = 0$, and $\rho_1 = \max\{\hat{\rho}_1, \hat{\rho}_2\}$. Disregarding the common constant term, the investor's profits are $\mathcal{E}_{q=1} = -t_1 - (1 - \pi) \max\{\hat{\rho}_1, \hat{\rho}_2\}$. The case for $q = 0$ is symmetric and the investor's profits are $\mathcal{E}_{q=0} = -t_2 - (1 - \pi) \max\{\hat{\rho}_1, \hat{\rho}_2\}$. If $q = 1/2$, it must be that $\rho_1 = \rho_2$, hence $\rho_i = \max\{\hat{\rho}_1, \hat{\rho}_2\}$ and investor's profits are $\mathcal{E}_{q=1/2} = -(t_1 + t_2)/2 - (1 - \pi) \max\{\hat{\rho}_1, \hat{\rho}_2\}$. A simple comparison of the profits for the three cases, $q = \{1, 0, 1/2\}$ delivers the result of the proposition. ■

Proof of proposition 12. The CIR is always binding. Define the SRO's profit as F , say $q \geq 1/2$, and assume that the CIR is not binding. SRO 1 can choose a strategy (t'_1, m'_1) where $m'_1 < \underline{m}_1$ and $t'_1 = t_1$ such that $q' = q$ and $F'_1 > F_1$. (The same logic holds for $q \leq 1/2$.)

$F_i > 0$ is not an equilibrium. Say $q \geq 1$ and $F_1 > 0$ is an equilibrium. SRO 2 can choose $t'_2 = t_1 - \epsilon$, with $\epsilon > 0$ arbitrarily small, such that $q = 0$. By continuity of F_i with respect to t and m , $\exists m'_2 < \underline{m}_1$ such that $F_2 > 0$ and the CIR is still satisfied. ■

B. Appendix 2

In this appendix we consider the case of imperfect knowledge of the monitoring technology, i.e., $a > 0$, when investors are *sophisticated*.

In this case, it is also possible to show that the optimal low-cash flow contract is the same as in the other cases, $z_i = \underline{w} - t_i$, and that the penalty is set at the maximum; however, the investor's objective function cannot be written in a closed form. In fact, given the investor j information set Ω_j , the expected trading choice $E(q|\Omega_j)$ depends on the investor j probability distribution over Ψ , as well; in particular, it may well be that $Cov(q, \Psi|\psi \leq v_j, \Omega_j) \neq 0$, which makes it hard to obtain a simple expression for the investor expected return. Hence, we cast the investor j 's problem in a very general form as maximizing $E(z|\Omega_j)$, subject to the broker's optimal reporting strategy and trading choice, and given the announced enforcement policies and transaction fees.

1. The SROs' Problem

Even if we have no closed form solution for the investors' problem, the SROs' best response problem is analogous to the one described in section A..

The unit (per transaction) profits are the difference between the transaction fee t and the expected monitoring cost in case the high state is reported, $E\mathbf{m}c$.

SRO i chooses an enforcement policy and transaction fee (p_i, x_i, t_i) , anticipating the customer's and broker's response and taking the actions of SRO i^* , $(m_{i^*}, x_{i^*}, t_{i^*})$ as given. Total profits are equal to unit profits times the SRO's trading volume.

It is useful to define the set of broker-investor pairs for which a contract is actually signed, given an enforcement policy and a transaction fee $(m_i, t_i, \underline{m}_{i^*}, t_{i^*})$, as:

$$N = \{j : \mathcal{E}_j(m_i, t_i, \underline{m}_{i^*}, t_{i^*}) \geq \alpha\} \subseteq [0, 1].$$

where \mathcal{E}_j is the investor j expected profit. The variable N summarizes the investors' participation constraint (CIR) and is affected by the enforcement policy and transaction fees chosen by the two SROs.

We define the set of contracts that are not incentive compatible

$$\Phi_i = \{j : v_j(z) \geq \psi\} \subseteq [0, 1].$$

Let \mathbb{I}_S be the characteristic (indicator) function of a generic set S and Ω_o represents the SRO information set. The problem of the SRO i with a linear monitoring cost function $C(m) = cm$ can be written as

$$\max_{\underline{m}, t} E \left[\int_0^1 q_i^j(\underline{m}_i, t_i; \underline{m}_{i^*}, t_{i^*}) [t - (1 + \mathbb{I}_{\Phi_i}(j))c\underline{m}] \mathbb{I}_N(j) dj \mid \Omega_o \right] \quad (25)$$

Because the problem does not have a closed form solution, we only report the result of the simulation exercise. Table 2 compares the result obtained for *sophisticated* investors with the ones obtained for *naive* investors. In our parameterized exercise, results are qualitatively the same. This suggests that, even when heterogenous investor beliefs are introduced, the means of competition, either monitoring or transaction fees, is not relevant for the equilibrium determination—as we have shown in the previous section for homogenous investors.⁴²

In Table 2 we compare the results for sophisticated heterogenous investors with the ones obtained for naive investors in Table 1. Most of the differences are due to approximation errors of the algorithm used for solving the sophisticated investor equilibrium.

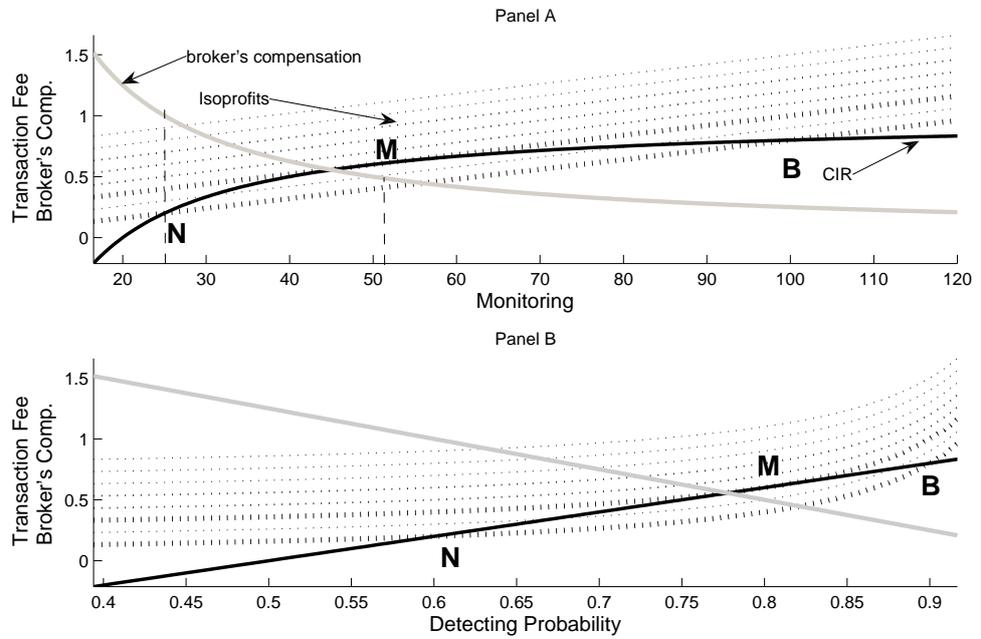
⁴²The numerical algorithm is computationally demanding. The two-dimension best response functions are approximated by Chebyshev polynomials of 12th order. A fixed point is searched in a neighborhood of the *naive* solution.

Table 2. *Naive vs Sophisticated*

Description	Sophisticated	Naive
Detecting Probability	0.75	0.75
Transaction Fee over Average Transaction Value	0.39%	0.41%
Total Volume (n)	.960	0.955
Total Frauds	0.055	0.059
SROs' Profits	0	0
SRO 1 Market Share	50%	50%
Brokers' Average Profits	0.523	0.524
Investors' Actual Average surplus	n.a.	0.047
Investors' Expected Average surplus	0.048	0.046
Total Monitoring Costs	0.424	0.421
Welfare	0.548	0.545

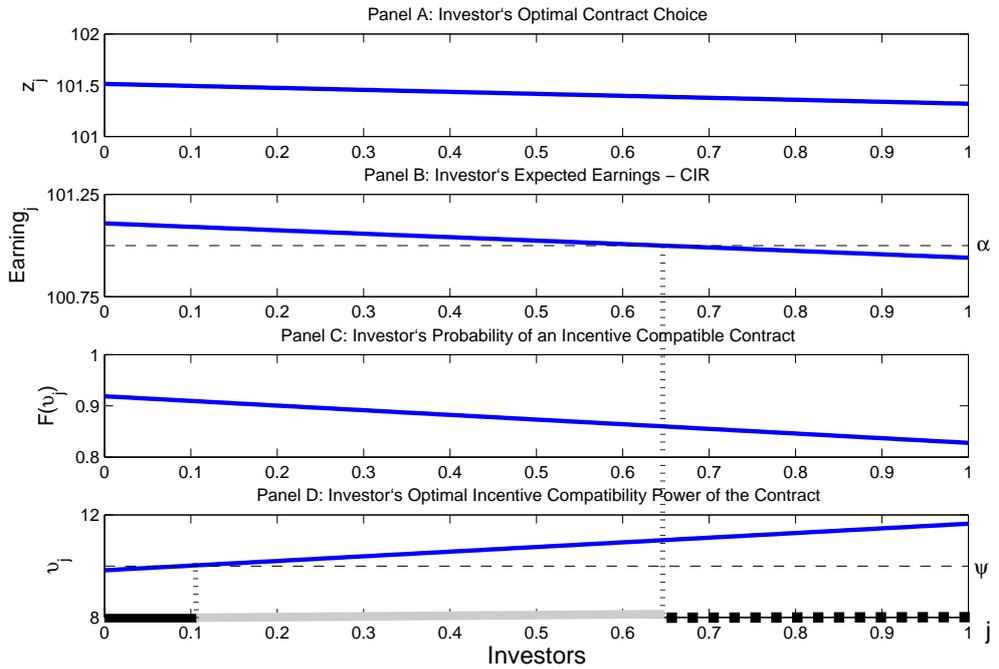
The average investor surplus is defined as the difference between average investor expected (actual) profits and the reservation profit. Parameters are such that $\bar{w}/\underline{w} - 1 = 2.5\%$, $\alpha/\underline{w} - 1 = 1\%$, $\pi = 20\%$, $\psi = 10$, $c = 0.04$, $\gamma = 6\%$, and $a/\psi = 10\%$.

Figure 1: Competitive versus Monopolistic Equilibrium



Equilibrium in the space (m, t) , top panel and (p, t) , lower panel. M and N represent the equilibrium in the monopoly and competitive case, respectively.

Figure 2: Investor Contracts



The x-axis shows investors ordered according to j , with the most “optimistic” on the left. Panel A shows the high-state contract across investors. Panel B shows expected earnings across investors and compares them with the reservation utility α . Panel C shows the subjective probability of having offered a truth-telling compatible contract, across investors. Panel D shows the contract that will actually imply or not imply a fraud: It compares v_j with ψ . On the x-axis, the light and dark thick solid lines represent contracts that are and are not incentive compatible, respectively; the thick dashed line represents those investors that do not participate.