Collusion and Concentration: Can Firm-Auditor Collusion Possibilities Cause Concentration in the Auditing Market? A Game Theoretic Analysis

Brishti Guha

Abstract
I consider a game where external auditors may collude with the client firms they audit. Given mandatory audit fee disclosure (required by the SEC since 2000) and an imperfect public signal capable of detecting cheating and collusion with nonzero probability, I show that the possibility of collusion results in a highly concentrated market structure for external auditors, with firms only wishing to hire those auditors who have many other clients. This results in an auditing oligopoly even in the absence of economies of scale in auditing or quality differences across auditors.

Keywords: Collusion, repeated games, auditing oligopoly, disclosure, fees, market concentration, public signal.

1. Introduction
While external auditors can play a role in enhancing the credibility of their client firms in the eyes of investors, this is only true as long as investors are satisfied that these auditors will not collude with the firms they audit. The possibility of collusion leads firms, investors and auditors to modify their behavior.

A seemingly unrelated fact is that the auditing market exhibits a very high degree of concentration. Currently, the “big four” – KPMG, Price WaterhouseCoopers, Ernst and Young and Deloitte – control the lion’s share of the market, catering, for instance, to 99 of the 100 FTSE in the UK (the Economist, 2011). Trends were similar in other countries. For instance, in the US the Big Four audited 99% of all public company sales in 2003 (Cunningham 2006). The Herfendahl – Hirshman index measuring market concentration was well above 1800 for auditing in 2006, indicating a very highly concentrated market structure2. The four-firm concentration ratio in auditing during 2002-04 averaged over 90% in several EU countries (92.3% in Italy, 93.2% in Luxembourg, 90.7% in the Netherlands and 92.2% in Spain (Ballas and Fafaliou

---

1 School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903. Email: bguha@smu.edu.sg.
2 Cunningham (2006).
Moreover, banks are reluctant to lend to firms unless these firms hire one of the big four as an external auditor.\(^3\)

While there is a rich economics literature on collusion, including collusion in the context of an external auditor, none of this literature to my knowledge shows that auditor-client collusion possibilities can in fact result in an oligopolistic market structure for the auditing industry. In this paper, I show that in a context where audit fee data are publicly observable (as is true following corporate governance reforms in 2000 after which the SEC made it mandatory for firms to disclose the fees they paid their auditors\(^4\)), the very possibility of collusion creates audit market concentration, provided there is a nonzero probability of collusion being exposed. Moreover, this remains true even in the absence of factors like economies of scale in auditing or productivity differences between auditing firms that are usually pictured as the causes of the auditing oligopoly.

The rest of the paper is organized as follows. Section 2 discusses some related literature. In Section 3 I lay out a game theoretic model with firms, auditors and investors playing a repeated game, and derive my main result, showing that firms prefer to hire those auditors who have a large number of other clients. Section 4 contains an intuitive discussion and interpretation of the results and concludes.

2. **Some Related Literature**

This paper is connected to the theoretical literature on collusion (starting with Tirole 1986), particularly on collusion in the firm-auditor context (including Khalil and Lawarree (2005, 2006) and Kofman and Lawarree (1996, 2003, 2006)). In these papers, the auditor is usually hired by the principal instead of by the firm. The principal often cannot detect collusion (unless an external signal confirms or disproves the auditor’s report as in Khalil and Lawarree 2006) and equilibrium often involves both collusion and shirking – with the agent cheating and the auditor colluding. Sometimes, in the absence of an external signal, it is best not to hire an auditor at all (Khalil and Lawarree 2006). Usually the principal’s strategy to deter collusion is to offer the auditor an extra reward if he reports wrongdoing on the part of the firm (instead of confirming the firm’s report): however this is a costly strategy and often not worthwhile for the principal.

\(^3\) The Economist (2011).
\(^4\) Francis and Wang (2005). In Section 4 I briefly discuss the consequences of relaxing the mandatory disclosure assumption.
Some of these papers involve a random external audit. In Kofman and Lawarree (1996) a second internal auditor – identical to the first one – is brought in for independent verification. The principal cannot tell if collusion has occurred. However, if one auditor confirms the firm’s report while another disputes it, the principal rewards the second auditor while punishing the first. Both auditors may be bribed by the agent (the firm). Kofman and Lawarree show that the only way of preventing collusion in this setup is for the principal to create some uncertainty by only probabilistically informing an auditor whether he is the first or the second auditor. Kessler (2004) studies a situation where an auditor may collude by either “forging” information (i.e. lying) or “concealing” it (pleading ignorance) and derives a tradeoff between the optimal frequency and intensity of audits. Baiman, Evans and Nagarajan (1991) do not model the incentives for collusion, but allow nature to determine whether self-enforcing collusive arrangements can prevail. Faure-Grimaud et al (1999) analyzes the cost of delegated auditing in a three-tier hierarchy. There is also a more general literature dealing with incentive compatible contracting between a principal and a monitor hired by him to supervise an agent.

These papers, however, do not focus on the impact of firm-auditor collusion possibilities on the market structure of the auditing industry – which is my focus. Another important difference with some of these papers is that in my model auditors are paid by the firms they audit, and not by investors. Finally, I examine a setup where the public (potential investors) can observe data on audit fees. This is a realistic feature given corporate governance reforms whereby the SEC mandated public disclosure of audit fees. Interestingly, the SEC’s justification for the measure was that this data might place investors in a better position to judge the likelihood of collusion between a firm and its external auditor (Francis and Wang 2005).

The accounting literature has largely looked at the implications of, rather than the causes of, audit market concentration. The limited literature on the causes of audit market concentration includes Dopuch and Simunic (1980), DeAngelo (1981), Danos and Eichenseher (1986) and Doogar and Easley (1998). Doogar and Easley (1998) argue that audit market shares are jointly determined by contracting practices, client size distributions and differences in auditor productivity. Dopuch and Simunic (1980) and Danos and Eichenseher (1986) emphasize the role of economies of scale in audit production while DeAngelo (1981) emphasizes quality differentiation across auditing firms in the presence of client-specific startup costs. My model,

5 Tirole(1986), Krasa and Villamil (1992) and Rajan and Winton (1995) are a few examples.
unlike these papers, assumes neither economies of scale in auditing nor client-specific startup costs nor quality differentiation. This is not to deny the importance of these factors in causing audit market concentration, but rather to focus on the issue of whether an independent causal role can be played by auditor-client collusion possibilities in an environment where these factors do not necessarily hold. As already emphasized, I also incorporate the realistic feature that investors are (now) able to observe information on audit fees.

My paper is also related to literature on multiplicity of transactions ("diversification") and imperfect correlation of the rewards and penalties from them (Laux 2001, Cerasi and Daltung 2000, Diamond 1984) in a way which will only become clear after the results are presented. Therefore I defer a more detailed discussion of this literature to section 4.

3. The Game
In the next subsection I examine some assumptions, in particular about the set of players and the information structure underlying the game. In a subsequent subsection, I discuss the timing of moves and players' feasible strategies. Finally I go on to prove the main results.

3.1 Assumptions: Players, Technology and Information Structure
All agents are risk-neutral and infinitely lived. There is a large number $E$ of entrepreneurs. Entrepreneurs have no capital of their own, but are financed by investors. In each period, every entrepreneur has access to one project, with a fixed capital requirement normalized to 1. Each such project yields gross returns of $G$ with probability $p<1$ and zero otherwise, amounting to an expected return of $H = pG$; project returns are unobservable by investors and unverifiable. Moreover, $H>R$ where $R$ is the risk-free rate of return on capital outside industry (and is also the investors’ outside option for their capital). Projects in subsequent periods require refinancing by investors. If investors ever decide not to refinance an entrepreneur, an investment covenant$^6$ prohibits future investment by this entrepreneur, so that the firm is dissolved.

---

$^6$ See Bolton and Scharstein (1990) for a discussion of investment covenants, which are feasible provided investment is verifiable. Investment covenants allow us to ignore the possibility that entrepreneurs continue operating on a private basis – possibly with the help of funds stolen from investors in earlier periods – even when denied financing by investors.
At the beginning of each period, entrepreneurs promise investors a return $g < G$ in the event their project is successful, and zero if not. For the purposes of our model, $g$ is exogenous.\(^7\)

It must however be high enough to meet the investors’ participation constraint, that is

$$pg \geq R$$

However, since project returns are unobservable by investors and unverifiable, the entrepreneurs’ promise is not contractible. Even if their project succeeds, they can claim it failed, reaping cheating gains of $g$ in the process. I will refer to this practice as “cheating”. While investors observe their own repayment – or lack of it – at the end of each period, they cannot automatically tell if not receiving a payment reflects cheating or genuine project failure. However, if an entrepreneur cheats, a public signal detects this and reveals its occurrence to investors at the end of the period with probability $q < 1$.

There is a pool of potential auditors, consisting of $M$ individuals possessing auditing expertise. Each inelastically offers his services to firms; each auditor is capable of servicing many clients up to a technological maximum of $L < E$. Moreover, neither the pool of potential auditors nor this technological maximum is too small, so that

A1: $M > E/L$

If an auditor investigates a case where investors are not paid returns, he can determine with certainty whether cheating has taken place. An honest auditor simply earns an endogenous fee $V$ from each client; the fee has to be paid at the beginning of any period when the auditor is hired, prior to any potential investigation by the auditor of the firm’s accounts. Moreover, investors can observe whether a firm has hired an auditor, and can also observe audit fees, which are verifiable.

However, an auditor need not act honestly. He has two courses open to him, collusion and extortion. Collusion involves making a secret offer to his clients of the following form, “You may cheat with impunity; I will certify that you have not cheated, synchronizing the delivery of my favorable audit report with a bribe that you must pay me”. The offer has to be made at the beginning of any period so that entrepreneurs have the opportunity to cheat should they wish to take up the auditor’s offer. The bribe however is paid only later once an investigation is launched and is synchronized with the delivery of a favorable audit report. This offer of collusion can be

\(^7\) Theoretically, $g$ may vary between $R/p$ – when the bargaining power lies with the entrepreneurs – to $G$ – when investors have all the bargaining power – and is realistically likely to lie in between these two extremes. However the results in the model do not depend on the exact value of $g$. 


simultaneously made to multiple clients (indeed, I will show that if an auditor colludes at all, it is always most profitable for him to collude with all his clients) but each client decides independently on whether to take up the offer.

Let $N$ denote the (endogenous) total number of clients that a single auditor has, and let $n \leq N$ denote the number of his clients with whom he is attempting to collude. If he were colluding with only one client, this would be detected and exposed by the same public signal that exposes cheating by firms, that is, with probability $q$. However, his probability of being caught colluding is not perfectly correlated across clients, so that the more clients he colludes with, the higher his probability of being caught; $q(n)$ is an increasing function of $n$ and is strictly greater than $q$ for $n > 1$. An example is a scenario in which the probability of not being caught while colluding with any one firm is independent of the probability of not being caught while colluding with any other. In this event, we would have $q(n) = 1-(1-q)^n$. If an auditor is ever caught attempting collusion (whether with one client or several) his credibility is immediately destroyed ruining his relationship with all his $N$ clients.

“Extortion” by an auditor refers to the possibility that an auditor discovers, after an investigation, that his client has not cheated but simply been unlucky, but threatens to falsely label the client as a cheat unless paid an extra bribe. Again, if such behavior occurs, it can be detected by the public signal with a probability that is increasing in the number of clients the auditor attempts to extort from. I will show later that extortion is never subgame perfect.

Before going on to discuss the timing of the game, and specifying strategy sets in more detail, we need to specify some more notation and parameter restrictions. First, we need a parameter restriction to ensure that auditing is necessary in the first place. That is, the possibility of detection by the public signal and the consequent loss of the stream of discounted future returns to honesty should not in itself deter entrepreneurs from cheating (if it did, honesty would be ensured through simple repeated game theoretic considerations even in the absence of auditors). At the same time, if cheating could be discovered with probability 1 (as by an auditor), the loss of the stream of discounted future rents should be high enough to deter cheating. Using $P_h$ to denote the discounted value of the entrepreneur’s future payoffs from honesty, this is ensured through

A2: $qP_h = q \delta p(G-g)/(1-\delta) < g < P_h$
This restriction allows us to concentrate on the “interesting” parameter space where there is a positive demand for auditing.

A final parameter restriction – again ensuring that \( L \), the technological maximum number of clients that a single auditor can service, is not too small\(^8\) – is

\[
A3: q(L)P_h > (1+q(L) - q)g
\]

recalling that \( q(L) \) is the probability that an auditor colluding with \( L \) clients gets caught. All parameters – \( p, q \), the form of the function \( q(n) \), \( G, g, \delta, L \) – are common knowledge.

3.2 Timing and Feasible Strategies

Entrepreneurs, investors and (hired) auditors engage in an infinitely repeated game. I describe the sequence of moves in any arbitrary period \( t \), describing the feasible strategies open to players:

1. *Entrepreneurs* decide whether to hire auditors (from the pool of \( M \) potential auditors). If so, they decide which auditor to hire, paying an agreed (endogenously determined) fee \( V \). *Investors* observe whether an auditor has been hired. If so, they also observe the identity of the auditor and how much he is being paid in fees. They then decide whether to finance the entrepreneurs. *Auditors*, if hired, accept the fee \( V \) and decide whether or not to make a secret collusion offer to their clients. If they do make such an offer, they decide how many clients to make the offer to.

2. If investors have not financed the firms in stage 1, the game ends. If they have, the project is implemented and payoffs are realized. *Investors* receive a payment of either \( g \) or 0. If they receive \( g \), the *auditors* do nothing. If they receive 0, auditors investigate and learn whether or not cheating has occurred.

3. If *auditors* find in stage 2 that cheating has occurred, they can either report this truthfully to investors, or collude with their cheating clients by delivering a clean report card in exchange for a bribe. If they find in stage 2 that cheating has not occurred, they can either reveal this truthfully, or threaten to blacklist the client unless paid a bribe. *Entrepreneurs* decide whether to give in to such demands. *Investors* then receive the auditor’s report.

4. A public signal reveals with probability \( q \) whether entrepreneurs have cheated; it also exposes collusion or extortion by auditors with probability \( q(n) \) where \( n \) is the number of

---

\(^8\) One can check that given \( A2: P_h > g \) – the LHS of the inequality in A3 is increasing more rapidly in \( L \) than its RHS. Hence A3 is equivalent to assuming that \( L \) is not too small.
clients that an auditor has tried to collude with or extort from. Investors now know (a) the payment they have received, (b) the auditor’s report, and (c) the realization of the public signal. This information influences their decision about whether to invest in period t+1.

5. If at the end of period t, investors decide not to refinance, the game ends. Otherwise, the game continues into period t+1 and steps 1-4 are repeated.

3.3 Main Results

My solution concept is a subgame perfect Nash equilibrium in pure strategies. I first describe a candidate equilibrium below, before proving its existence and deriving the results of interest. For convenience I call this candidate equilibrium “the honest equilibrium.”

Definition 1: In the “honest equilibrium”
(a) all firms hire auditors and act honestly,
(b) there is no collusion or extortion
(c) investors know this and finance the industry,
(d) the investors’ off-equilibrium strategy is to shun any firm which does not hire an auditor, to shun firms if the auditor’s fee does not lie within a prescribed range, to stop financing any firm the auditor labels a cheat or any firm exposed by the public signal, and to mistrust auditors if and only if they are revealed by the public signal to be colluding with or blackmailing clients.

Proposition 1: (i) Firms can satisfy investors that there will be no collusion by paying their auditor a sufficiently high fee. This can be done most cheaply if each hired auditor has as many clients as possible, ie, \( N = L \). This creates an auditing oligopoly.

(ii) The honest equilibrium of Definition 1 exists when only \( E/L < M \) auditors are hired, while the rest are unemployed.

(iii) The off-equilibrium investor strategies described in Definition 1 are subgame perfect.

Proof (i) The proof of part 1 of Proposition 1 proceeds in several steps.

Step 1. Denote by \( X \) the maximum bribe that an entrepreneur would be willing to pay an auditor for allowing him to cheat and giving him a clean report in exchange for the bribe. Now \( X \) is the entrepreneur’s cheating gain less expected loss due to possible exposure plus expected saving in that event of future audit fees. By colluding, the entrepreneur gains cheating gains of \( g \), but his cheating will be exposed by the public signal with probability \( q \), in which event – according to
the investors’ off-equilibrium behavior prescribed in Definition 1 - he is never refinanced, losing his future discounted stream of payoffs, $P_h$, minus the discounted value of audit fees in future periods. Thus

$$X = g - q\{P_h - \delta V/(1-\delta)\}$$

(2)

Step 2. Consider a hired auditor’s incentives to collude. For the time being, suppose that an auditor with $N$ clients who wants to collude at all will want to collude with all $N$ clients, so that $n = N$. In step 3 we will see that this is indeed the case. Then this auditor’s maximum bribes from colluding with all $N$ clients are $NX$ where $X$ is given by (2). However, his probability of getting caught is then $q(N)$; once caught, his reputation is destroyed, and in accordance with investors’ off-equilibrium strategy of mistrusting such an auditor, no firm ever hires him again, so that he loses the future discounted stream of audit fees from all his clients. Thus, the auditor’s no-collusion constraint may be written as

$$Nq(N)\delta V/(1-\delta) \geq N\{ g - q\{P_h - \delta V/(1-\delta)\}\}$$

or, canceling common terms and simplifying,

$$V \geq (1-\delta)(g-qP_h)/\delta q(N)-q) = V$$

(3)

From A2, we see that $V$ is strictly positive. This imparts an efficiency wage flavor to the model; to be credible, auditors must be paid a fee (wage) strictly above their opportunity cost of zero. In the spirit of efficiency wage models of, there is some “unemployment” or excess supply of potential auditors; competition among auditors drives audit fees down to $V$, but no lower.

Step 3. We now demonstrate that indeed $n = N$, that is, if an auditor colludes at all, it is best for him to collude with all his $N$ clients. Now, the expected income of an auditor who colludes with $n$ of his $N$ clients is

$$Y(n) = nX - q(n)N\delta V(n)/(1-\delta) = nX - Nq(n)(g-q\delta P_h)/(q(n)-q) = nX - N(g-q\delta P_h)/(1-(q/q(n)))$$

(4)

shows that while such an auditor would receive bribes from the $n$ clients he colludes with, he would, with probability $q(n)$, be exposed by the public signal in which event he loses his future fees from all his $N$ clients. These fees, are in turn $V(n)$, the “credibility wage floor” when (3) holds as an equality, substituting $n$ (the number of clients with whom the auditor colludes) for $N$. Now we can check that $Y(n)$ is unambiguously increasing in $n$. Hence, the auditor, if he colludes at all, has every incentive to set $n = N$.

---

Step 4. From steps 2 and 3, each hired auditor must be paid $V(N)$; no entrepreneur wants to pay higher than necessary, and paying any lower – or not hiring an auditor at all – results in investors not financing the entrepreneur at all. However, we see that $V$ is a decreasing function of $q(N)$, and therefore of $N$; the more clients an auditor has, the smaller the credibility wage each of his clients needs to pay him. Thus, each firm seeks to hire an auditor with as many clients as possible; leading to the emergence of an audit oligopoly\(^{10}\) where each hired auditor services his technological maximum of $L$ clients; $E/L < M$ auditors are thus hired, while the rest remain unemployed. This proves part 1 of the Proposition.

(ii) If $E/L$ auditors are hired, with each auditor servicing $L$ clients, the credibility wage is driven down to $V(L)$, which is what each such auditor is paid. Now investors can satisfy themselves that the no-collusion constraint is being met. However, they still need to consider whether (a) a firm may hire an auditor who does not collude, and cheat in spite of certain exposure, and (b) whether an auditor may attempt to extort from an honest entrepreneur. Now (a) can be ruled out if and only if

$$P_h - \delta V(L)/(1 - \delta) > g$$

The LHS of (5) depicts the discounted stream of future losses – the payoffs to honesty net of future audit fees – that a cheating firm will face when exposed by its auditor; these must outweigh the cheating gains on the RHS to rule out cheating. Now, substituting in for $V(L)$ using (3) and replacing $N$ by $L$, and simplifying, we find that (5) is equivalent to

$$q(L)P_h > (1 + q(L) - q)g$$

which is always true by A3.

Next we consider the possibility of extortion. Auditors may attempt extortion from an honest but unlucky firm by threatening to falsely report that it had cheated. However, the firm being blackmailed would recognize this as an empty threat. It realizes that if it refuses to pay up, the auditor has no incentive to actually implement its threat: while the auditor does not gain anything from lying about the firm (given the latter’s refusal to pay), he stands to lose his reputation – and therefore his future clientele – if his lying is exposed by the public signal. Thus, in a subgame perfect equilibrium extortion is ruled out.

\(^{10}\) The oligopoly may, for instance, emerge as follows. Since entrepreneurs decide which auditors to hire, the first auditor to receive an offer of employment will then receive offers from the next $L-1$ firms to make offers; after which the process is repeated for a second auditor, and so on until each firm has an auditor and each hired auditor can take no more clients.
Thus, there is no collusion (because firms will refuse to hire auditors for fees below $V(L)$, given that then they will receive no investment), no extortion, no incentive for firms to hire an auditor and then cheat, and no incentive to not hire an auditor (given that, again, that would result in no investment), and the honest equilibrium of Definition 1 exists. Investors finance firms because they can observe that the audit fees meet the credibility floor and can also check that firms have no incentive to cheat, while auditors have no incentive to collude or extort. This proves part (ii) of the Proposition.

(iii) It only remains to check whether the off-equilibrium investor behavior described in Definition 1 is subgame perfect. First, if a firm does not hire an auditor, does it pay investors to deny the firm investment? The answer is yes: from A2, we recall that $qP_h - a$ firm’s expected discounted losses in the event of exposure by the public signal — is strictly less than the cheating gains $g$. Thus a firm not subject to auditor scrutiny will have every incentive to cheat even if this cheating is revealed with probability $q$ by the public signal. Knowing this, it makes sense for investors not to invest in such a firm.

Secondly, does it make sense for investors to deny firms investment if they see that audit fees violate either (3) or (5)? Again, the answer is yes, because a violation of (3) — too low a wage for auditors — would lead to auditor-client collusion, so that hiring an auditor is no guarantee of credibility. A violation of (5) on the other hand would mean that the credibility wage is so high that firms’ discounted future payoffs to being honest and hiring an auditor are low relative to their one-time cheating gains, prompting them to cheat even at the cost of certain exposure. Again, this would mean that it does not make sense for investors to invest. [We have shown in part (ii) that firms’ preferences for hiring auditors with as many clients as possible guarantees that if auditors are paid their credibility wage, firms have no incentive to cheat].

It is also credible for investors to refuse to finance any firm that is caught cheating, either by the public signal or by an auditor’s report. Given that extortion is not subgame perfect, a bad report by an auditor reflects genuine cheating on the firm’s part. To see that denying financing is then a credible threat, simply note that not investing is the Nash equilibrium of the stage game; if our game were played only once, firms would always have an incentive to cheat, leading to hold-up. A standard result in game theory is that the stage game Nash equilibrium is always a credible threat in a repeated game. Finally, not trusting an auditor who has been caught colluding makes sense because such an auditor’s report carries no value; hence a firm that hires such an auditor can
be considered on par with a firm that does not hire an auditor at all. We have shown that in the latter case, firms will always have the incentive to cheat; knowing this, investors refuse to finance any firms that hire auditors previously exposed by the public signal. This proves part (iii) of the Proposition. \textit{QED}

4. Discussion: Intuitions and Conclusion

In this section I discuss the intuition underlying two results from the previous section. The first result I discuss is the main one about firms’ preference for hiring auditors with as many clients as possible. The second is the result that it is better for an auditor to collude with all its clientele rather than with a strict subset of its clients. I briefly discuss likely implications if audit fee disclosure were not mandatory. I then discuss somewhat related ideas in the literature.

Given audit fee disclosure, firms can use their audit fees to commit to credibility. This commitment is however cheaper for a firm if it hires an auditor with many clients; such an auditor will refrain from colluding even if paid a relatively low “credibility wage”, and this fact is known to investors. This low credibility wage not only rules out auditor-client collusion at a cheap price (for the firm) but also guarantees that the firm will not find it worthwhile to cheat at the cost of being exposed by the auditor.$^{11}$

This of course begs the question as to why auditors with many clients are more likely to refrain from collusion, even if paid a (relatively) low fee. At first, we may be tempted to infer that this is because auditors with many clients have “more to lose” if their collusion is exposed. However, the argument is more subtle than this. Even though the fact that auditors with many clients will lose the fees from all these clients if exposed does tend to reduce these auditors’ incentives to collude, as against this these auditors’ ability to obtain bribes from all these clients (given multiple collusion) \textit{increases} their incentives to collude by an equal extent, leaving the no-collusion constraint unaffected. There is, however, a crucial third factor which ensures that auditors with many clients have relatively low incentives for collusion - which is that the probability of being detected and exposed in collusion is increasing in the number of clients an auditor has (given our other result that if he colludes, he colludes with all of his clients). This increased probability of detection implies that an auditor with many clients faces a larger expected

\footnote{That is, if the auditor’s fee is low enough, a firm’s stream of discounted payoffs to honesty net of these fees is attractive enough such that the firm will not want to lose them for the sake of one period cheating gains.}
loss from collusion, making him willing to refrain from collusion for the sake of a relatively low credibility fee.

I now briefly discuss the intuition underlying the subsidiary result that an auditor who colludes would want to collude with all his clients. While the amount of bribes collected is obviously increasing in the number of clients \( n \) the auditor colludes with, the probability of getting caught is also increasing in \( n \). However, the credibility fee – which is what is forfeited in the event of getting caught – is, at the same time, decreasing in \( n \), at a faster rate than the rate at which the probability of getting caught is increasing. For a given number of total clients \( N \), therefore, an auditor’s income from collusion is an increasing function of the number of these that he colludes with.

Mandatory disclosure of audit fees enables the investor to avoid firms and auditors who are likely to collude. Where disclosure is not mandatory, firms, in order to assure investors of the credibility of their auditing processes, would have an incentive to disclose audit fees voluntarily anyway (of course on the assumption that false disclosure would be detectable and punishable as fraud). Alternatively, they could hire only auditors with a large clientele.\(^\text{12}\) Either of these measures would have the same effect as mandatory disclosure in increasing concentration in the auditing business.

Multiplicity of transactions and imperfect correlation of the rewards and penalties from them have of course been at the heart of many phenomena in corporate governance. Managers working on multiple independent projects can be punished for neglecting a project by withholding from them the returns of their other projects without running into the constraint of limited liability: this induces them to perform better (Laux, 2001). Debt-financed bankers monitoring a diversified asset portfolio are likelier to be able to repay their debts: so they can expect a full-liability payoff on their monitoring efforts – a fact that improves their incentive to monitor (Cerasi and Daltung 2000, Diamond 1984). In our model, multiplicity of clients and imperfect correlation of the risk of detection of collusions facilitate the punishment of colluders, increase the commitment power of auditors with many clients, and lead to an oligopolistic market structure in auditing.

\(^{12}\) Since for a given audit fee, auditors with a large clientele have weaker incentives to collude, hiring them enhances firms’ credibility.
References


