The Market for Liars: Reputation and Auditor Honesty*

by

Andrew McLennan
Department of Economics
University of Queensland

In-Uck Park
University of Bristol
Sungkyunkwan University

This version: **August 2015**

Abstract. In the model there are two types of financial auditors with identical technology, one of which is endowed with a prior reputation for honesty. We characterize conditions under which there exists a "two-tier equilibrium" in which "reputable" auditors refuse bribes offered by clients for fear of losing reputation, while "disreputable" auditors accept bribes because even persistent refusal will not create an improved reputation. Such an equilibrium may fail to exist because competition among reputable auditors drives fees below the level at which refusing bribes is optimal. Sustaining such an equilibrium requires both that entry into the reputable segment is impossible or unprofitable, and also that the reputable segment is prevented from expanding past the point where the equilibrium collapses. (*JEL* Classification Numbers M41, D82, G14.)

Keywords: auditor reputation, fee premium, independence, audit insurance.

"The National Association, representing the majority of its members' views, always has held firmly to the belief that all human nature is weak at best, and that that weakness extends even to Public Accountants, certified or otherwise."
W.R. Anderson, "The Surety Bond," *The Certified Public Accountant Bulletin*, IV-4, 1925.

1. Introduction

During the first three decades of the twentieth century the auditing industry grew rapidly, and several audit firms opened branches in multiple cities(1). Since that period most observers have recognized a clear distinction between a small number of large firms, described as the Big Eight, the Big Six, the Big Five, or the Big Four, depending on the era,

* This project was initiated when the first author was a Benjamin Meaker Visiting Professor at Bristol University. Both authors are grateful for the hospitality of the Institute for Social and Economic Research of Osaka University. We have benefitted from conversations with Ian Jewitt, detailed remarks by Jennifer Ireland, and the comments of seminar participants at Australian National University, the Institute for Social Economic Research (University of Osaka), Keio University, the University of Pittsburgh, the University of Illinois, the University of Sydney, the Midwest Economic Theory Meetings (Spring 2003) and the Canadian Economic Theory Conference (Spring 2003). The second author gratefully acknowledges the research support of the ESRC, U.K. (Grant Number R000222716). Emails: a.mclennan@economics.uq.edu.au and i.park@bristol.ac.uk.

(1) This process in described in Edwards (1960).

who are generally acknowledged to have a special status within the industry. (Herein these will usually be described as the Big n firms.) Various studies suggest that smaller audit firms charge lower fees and their clients received less favorable treatment from financial markets in connection with IPOs and perhaps other transactions. The membership of the Big n seems to be very stable: two bilateral mergers in 1989 reduced the number from eight to six, and a merger in 1998 reduced the number to five[(2)]. Although the collapse of Arthur Andersen in connection with the Enron scandal (revealed in 2001) provides an example of departure from this group, it seems to be very hard to enter the Big n[(3)]: all of its current members are descended from the firms that expanded on a national scale during the early part of the last century. These firms have also undergone considerable internal expansion in recent decades.

This paper presents a model in which there are two types of audit firms, which we describe as "reputable" and "disreputable." These firms have identical technology, but behave differently in equilibrium. The reputable audit firms are more fearful of losing their reputation and less easily induced to misreport the condition of the client. As a result their reports are, in fact, more credible, and more valuable because of the impact on the client firms' valuations on asset markets, or because they allow new financing to be secured on more favorable terms. For this reason reputable firms are able to charge higher fees, and the capitalized value of expected future fees is what makes them more reluctant to jeopardize their reputations. Thus ours is a model of "pure" reputation: agents of different reputation do not differ in any fundamental way, so that the desire to preserve one's reputation is the only motivation for behaving in the manner one's reputation predicts.[(4)] Wilson (1983) proposes that auditors' reputations are of this sort; many of the themes announced in that paper take on concrete and particular forms in our work.

In some circumstances capital (thought of in its general sense as a commitment extending across periods) can serve as a device that decreases the cost of providing high quality goods, in comparison with a "fly-by-night" strategy of selling low quality goods

---

(2) Arthur Andersen, Deloitte & Touche, Ernst & Young, KPMG, and PricewaterhouseCoopers constituted the Big 5 in 1999.

(3) E.g., "Too Few Accountants," *The Economist*, January 31, 1998.

(4) Our mechanism appears to be similar to the statistical theory of discrimination where the market's negative perception on the characteristics of a particular group is self-fulfilling, because anticipation of the market's unfavorable treatment due to negative perception, reduces the investment motives of this group, which confirms the perception. (See, for example, Phelps (1972), Arrow (1973), Coate and Loury (1993).) Our mechanism differs, in particular, in that, there being no investment decisions to be made, reputable and disreputable firms have the same audit technology. Furthermore, the differential in their audit-reporting practices is not sustainable in one-shot settings where the damage from an uncovered audit fraud is the same across all firms. Unlike the statistical theory of discrimination, therefore, the differentiated audit-reporting behavior we analyze is supported in repeated settings but not in one-shot settings.

at high quality prices until consumers' beliefs change. Klein and Leffler (1981) point out that, when a fly-by-night strategy could yield positive profits, a firm with a reputation for high quality must be expecting a stream of future returns on its reputation of equal or greater value. Capital can also serve as a barrier to entry that prevents a high quality firm's expectation of future profits from being eroded, so that the firm can rationally expect that its initial investment in high quality capital will have a nonnegative return. All this is relatively straightforward when the capital can be observed by consumers. When consumers cannot observe capital the provision of high quality goods must be sustained by beliefs. Kreps and Wilson (1982a), Milgrom and Roberts (1982), and Rogerson (1983) provide models in which reputation is a matter of belief about a hidden state variable.

Like the model presented here, the models of Klein and Leffler (1981), Shapiro (1983), and Allen (1984), and also earlier theoretical studies (DeAngelo (1981) and Dye (1993)) of auditing, are models of pure reputation insofar as there is no hidden state variable. Klein and Leffler's discussion, which is relatively informal, can be seen as motivating the formal elaboration of their ideas in subsequent research. Shapiro (1983) provides a precise model in which any firm can establish a reputation by producing high quality goods at a loss for one period, which will be amortised by the future income stream that reflects the obtained reputation. Allen (1984) explores further the extent to which such equilibria may be sustained when consumers may observe the firm's quantity (as well as the delivered quality ex-post).

A key feature differentiating our model from those described above is that we assume, for reasons to be explained shortly, that firms with bad reputation are powerless to change the market's perception. In the sort of language used at the time of the papers mentioned above, this would be considered a different equilibrium concept: they assume a free entry condition and we impose the opposite assumption. Expressed in terms of the terminology derived from the sequential equilibrium concept of Kreps and Wilson (1982b), the earlier literature studies one type of equilibrium of the underlying game and we study another.

In Shapiro's model a firm can choose its reputation directly because consumers always expect a firm to produce the quality observed in the last period. Of course these expectations must be fulfilled along the equilibrium path, and in addition this equilibrium condition depends crucially on consumers actually being able to observe quality after a one period lag. For the market studied here, and in many other settings involving reliability and safety, low quality results in infrequent failures, so low quality might go unnoticed for some time. Observation of quality is consequently a matter of strategic and statistical inference. In such a setting, producing high quality in any given period sends a very weak signal to the market, so one might think that any strategy for acquiring a reputa-

tion that depends on consumer observation of quality must necessarily involve the gradual acquisition of reputation over a number of periods.

But such gradual acquisition of reputation seems problematic because, in equilibrium, a firm has a good reputation precisely when it has an incentive to maintain its reputation. If a firm was thought to be on the cusp of acquiring a reputation for high quality, it would have a strong incentive to not jeapordize its position, and in this sense it would already have the reputation it sought. More generally, whenever a firm's unique best response was to embark on or continue along a path of reputable behavior leading eventually to a good reputation, it would already be reputable, in the sense that everyone would already expect it to behave reputably.

It is certainly possible, both in our model and in those of earlier papers, that there could be random events, either in the model itself or in the larger world, that according to some social convention lead to some firm having a good reputation. In this sense our model does allow equilibria in which new firms enter the high quality tier, but as a consequence of luck rather than willfully pursuing a strategy leading to that situation. However, our purpose in this paper is not to explain such emergences of reputation but to examine the extent to which reputation thus obtained is indeed sustainable in equilibrium, which may depend on the specific market in question.

The remarks to this point pertain quite generally to industries in which the observation of quality is infrequent and/or stochastic. These considerations apply with special force to the auditing industry. There is, first of all, the reality that for at least the last half century there has, in fact, been no recent entry into the Big n. Big n firms are highly profitable, suggesting that entry barriers are at least extremely high, if not literally insurmountable. Top audit firms, whose clients are the largest corporations, are conceivably exposed to correspondingly large inducement payments that may be offered by distressed client firms to influence the auditor's behavior, which we refer to as "bribes" for brevity. (In practice, bribes are difficult to identify as such, since they are embedded in the continuing relationship between the firm and the auditor, which may involve purchases of consulting services or other interactions.) The probability of detection is more difficult to estimate; probably it varies quite a bit on a case-by-case basis, and is sometimes rather small. In this connection it is interesting to note that, at least in auditors' self descriptions, maintaining a reputation for high quality services is in large part a matter of cultivating a corporate culture, one part of which consists of deterrents to cheating even when the circumstances are particularly favorable. This appears especially pertinent in light of the fact that each partner's temptation to accept bribes would not internalize the negative externalities to all other partners in case of exposure.

Accepting that firms' reputation is more of a result of historic accident and luck rather than that of strategic pursuance by the firms in such industries like auditing, one is led to the question of to what extent the reputation bestowed on firms may be sustainable. In particular, would it be possible that reputable auditors coexist alongside with disreputable auditors? For this to be the case, the premium that reputable auditors command must be high enough for them to resist the maximum bribe the firms would be willing to offer but, at the same time, it should be low enough so that enough firms are willing to hire them instead of the disreputable auditors who charge a lower fee. Note that unlike moral hazard models, such as Shapiro (1983), where the short-term gain from "cheating" is exogenously given cost savings, in our context the short-term gain is the amount of bribe that firms would be willing to offer, which is endogenously determined simultaneously with how favorably they would be treated in the financial markets. Thus, verifying the incentive compatibility conditions surrounding viable levels of high quality premium is more delicate and it is not clear a priori whether and when the two types of auditors may coexist.

Our analysis provides a delineation of the conditions under which the two types of auditors coexist. We now describe the key features of equilibria in which both types of auditors coexist, which we refer to as "two-tier equilibria." Subsequently we discuss more informal implications concerning circumstances under which reputable audit services may be sustained in the market.

In each period of our model, a continuum of firms privately informed of the likelihood that they are in a good state, referred to as their "type", appoint one auditor each for a fee from a pool of auditors, of which a fixed fraction are perceived as "reputable." In a two-tier equilibrium, in every period all reputable auditors are hired at a high fee and report the client firm's state truthfully, while all disreputable auditors are hired at a lower fee and accept bribes and misreport when the client firm's state is bad (and are forced to close the business if the firm fails subsequently). Because firms reported good by reputable auditors get most favourable terms of loan and firms of higher types are more likely to receive this benefit, the equilibrium has a cutoff structure, with client firms hiring reputable auditors if and only if their types are above a certain threshold. The key equilibrium conditions are that (i) firms of these types exhaust all reputable auditors, (ii) the threshold type is indifferent between hiring a reputable auditor and a disreputable one, and (iii) the maximum bribe a client firm in a bad state is willing to offer will not be accepted by a reputable auditor because it does not cover the foregone future stream of premium fees in case the firm fails. Our analysis delineates when a two-tier equilibrium exists that satisfies these conditions and when it does not.

One feature of our model is that the actual "quality" provided by disreputable audit

firms is endogenous, since it is the average *ex ante* quality of their clients. This is the primary source of some key conclusions derived in our analysis, since it points to two ways in which the fee premium charged by high quality auditors may be too small to deter acceptance of bribes, in which case there does not exist a two-tier equilibrium:

(a) It may be impossible to sustain an equilibrium in which a large fraction of the market is served by high quality auditors, since the quality of the marginal client firm (i.e., its probability of being *in a good state*) is then low, which results in a fee premium for reputable auditing that is too small.

(b) The pool of client firms is so good, even after removing the client firms audited by reputable auditors, that the financial markets believe that a client firm audited by a disreputable auditor is still very likely to be a good investment. In this case only a small amount of information is conveyed by a high quality audit, and again the fee premium for reputable auditing is too small.

In case (b) reputable auditing is unimportant, and the welfare loss resulting from its collapse quite small, because the phenomenon it detects is rare. This second possibility appears to be more an artifact of theory than a relevant concern, and our discussion below will focus on the other possibility.

In case (a), for a marginal client firm, purchasing high quality audit services has a low probability of resulting in ultimate success, and the option of offering a bribe is valuable, so it is willing to pay only a small fee increment to obtain high quality auditing. An equilibrium with both types of auditors will collapse if high quality auditors are offered bribes in excess of their expected future profits, which are proportional to this fee increment.

Going a bit beyond the formal structure of our model, if we view it as a static description of the industry in a single era it becomes possible to ask if the collapse of Arthur Andersen can be understood as the result of the gradual evolution of the industry over a long period of time, to the point that case (a) became relevant. Some data suggests that the market share of the top audit firms has, indeed, increased substantially during recent decades. Although there has been no entry of new firms into the Big n, the member firms have expanded greatly since the early part of the Twentieth Century, when they acquired a national presence by opening offices in multiple cities, and there seems to be little reason to doubt that this expansion has been as inexorable in recent decades as it was earlier. In our formal analysis the size of the reputable segment is fixed, but one may interpret it as providing a snapshot of this growth process over a small number of periods. Specifically, a series of such equilibria with increasing reputable segment size corresponds, approximately, to internal expansion that is constrained to be slow, for instance because the corporate culture of reputable auditing can only be sustained in an environment in

which veterans greatly outnumber new employees. In this scenario, slow expansion was accompanied by a gradual diminution of the fee increment, which led eventually to a situation in which Arthur Andersen was tempted to risk its reputation.

The collapse of Arthur Andersen led to reforms aimed at strengthening "auditor independence." The Sarbanes-Oxley Act has numerous provisions, some of which restrict the auditor and the client firm from engaging in other sorts of business. Others mandate stringent record keeping requirements or strengthen penalties for top officers of client firms that have committed accounting fraud. The perspective provided by our model suggests that such reforms may postpone the collapse of the reputation mechanism by increasing the costs of bribing the auditor. However, these reforms seemingly do little to arrest the internal expansion of the Big n firms, which suggests that we may again face a similar crisis in the future. The collapse of Arthur Andersen probably diminished the size of this segment, even though many former Arthur Andersen employees have found employment at other Big n firms, but this retrenchment could easily be short-lived.

A more radical reform would be to require publicly traded firms to purchase audit insurance, and to require the auditor to be hired by the insurer. Indeed, a special feature of the auditing industry stressed by Wilson (1983) is that the purchasers of the service (the client firms) and the true "consumers" of the auditor's product (actors in financial markets) are different. The temptation to provide low quality arises out of the conflict of interest between these two groups of agents. Of course if the reform was implemented the client firm might still try to bribe the insurer-auditor partnership, so the real import of this proposal may be to require the financial structure of the insurance industry (including reinsurance) to be used to create "deep pockets" that prevent the client firm from using bankruptcy to shift losses to debt and equity markets. In assessing this proposal one should recall that the institutions of limited liability and bankruptcy can be understood as a "dirty" compromise necessitated by the limited financial and monitoring capacity of potential insurers, and it is not obvious that audit insurance would be a viable exception to this general principle. On balance we think this proposal merits further, cautious, consideration.

It is natural and timely to ask whether our model could also apply to credit rating agencies, which, like auditors, are employed by the issuers of financial assets to provide information used by others. (Until the early seventies agencies sold their ratings to investors, but this business model ceased to be viable with the advent of photocopying.) Recent anecdotal evidence suggests that rating agencies are at least sometimes tempted to provide ratings biased in favor of their clients, but that they also have some concern about the general reputation of their ratings. However, our model seems to fit that indus-

try less well for at least two reasons. First, the discovery of dubious ratings has not led to exit from the industry. Second, there are only a few ratings agencies, and face little competition with a disreputable segment of the market and extreme barriers to entry. Indeed, the position of the ratings agencies is enshrined in the SEC's Nationally Recognized Statistical Rating Organization designation, and further protected by various institutions that require investors to invest in assets that are rated as safe. Bolton et. al. (2012) and Mathis et. al.(2009) provide theoretical models of reputation in this industry.

The remainder has the following structure. Sections 2 and 3 formalize our model and equilibrium concept, respectively. Section 4 studies the extent to which reputable auditing is sustainable. Section 5 concludes.

2. The Model

In every period there is a continuum of client firms with total measure one. A client firm cannot stay in business unless it succeeds in obtaining financing in the form of a loan whose amount ℓ is fixed and exogenous. Our analysis assumes that an external audit is necessary for the firm to seek outside financing. This assumption can be justified by appeal to tax law and regulations governing publicly traded companies, but in fact our main findings are not sensitive to this assumption, as we explain in Remark 1 toward the end of Section 4.

Each client firm is privately informed of its *type* $t \in (0,1)$. The distribution of firm types is given by an atomless probability distribution μ on $(0,1)$ with full support. The firm has an underlying state, which may be either good (G) or bad (B). The firm's type is the prior probability that the state is good, and after the state is known the type does not convey any additional useful information about the firm.

At the beginning of every period there is a continuum of existing auditors. The measure of reputable auditors is m_R, there is a certain measure of disreputable auditors, and new auditors can enter the industry by paying a fixed cost C. All auditors are identical in the sense that they employ the same audit technology at the same cost, which is normalized to zero. We normalize auditor capacity, assuming that each auditor is capable of auditing one firm in each period[(5)].

[(5)] This is an extreme assumption, which is difficult to defend on the basis of realism. Disreputable auditors have an incentive to be as small as is consistent with physical economies of scale, because this minimizes the value risked by dishonest behavior, and thus technological factors determine their size. Reputable auditors have the opposite incentive as they are better motivated to behave when more is at stake. In this light, a reputable audit firm may be understood as a confederation of such enterprises, and in effect our model assumes that each of these units understands that the confederation's reputation increases the expected future revenues of that unit, but the unit is not incentivized to value the reputation beyond that. At the expense of some algebraic complication, we could introduce a factor $\zeta \geq 0$ such that a unit

After learning its type, the client firm chooses an auditor and enters into a contract that specifies a contractual fee f for the audit[6]. The value of f, which may vary across auditors depending on the anticipated quality of service, will be determined in the audit service market at a competitive level that equates supply and demand of each type of auditor. The audit then reveals whether the state of the firm is good or bad to both the firm and the auditor, but not to the financial market.

The auditor then makes a public report concerning the state of the firm. It must report that the state of the firm is good when that is in fact the case. If the state is found to be bad, the client firm has a chance to make a take-it-or-leave-it offer[7] of inducements, or a bribe for short, to the audit firm to report that the state is good. Denote the amount of the bribe by b. If the audit firm accepts the bribe it must report that the state is good, i.e., issue an "unqualified" audit report in the language of audit industry.

After observing the auditor's identity and report the financial market (which is assumed to be competitive) may either refuse to extend financing a loan of ℓ or extend it in exchange for a repayment promise r. If the firm does not obtain the loan it is liquidated with zero scrap value. If the firm obtains financing, it either succeeds or fails. The probability of success is either p_G or p_B, according to whether the firm's state is G or B, where $p_B < p_G$.

In the event that the client firm succeeds its value is h (which is fixed and exogenous) so that the profit resulting from the attempt at survival, net of the total cost of audit and the loan payment, is $h - f - r$ if no bribe was paid and $h - f - b - r$ if there was a bribe.

---

with reputation value v acts as if the value is actually $(1 + \zeta)v$, with ζv representing the extent to which the larger firm is able to induce the unit to lean in the direction of honesty. This would not change the main features of the analysis or the conceptual conclusions, so we have chosen to simplify the presentation. However, the reader should bear in mind that the ability of audit firms to impose a "corporate culture" on their constituent components is an important issue in this industry.

[6] We assume that the contractual fee cannot depend on the reported state. An agreement to pay a higher fee when the state is bad amounts to a vehicle for shifting a greater fraction of the expected fee onto the firm's creditors when the firm is bankrupt, so it is natural to expect that such an arrangement would be disallowed by the bankruptcy court.

On the other hand an agreement to pay a higher fee when the state is good is disadvantageous from at least two points of view. First, in practice, though not in the model due to the assumption of rational expectations, if the financial markets can observe the terms of such a contract, they will tend to have doubts about the auditor's independence. Second, the effect of such a contract is to shift some of the auditor's expected payments into events in which the firm does not declare bankruptcy, which is unnecessarily generous to the firm's creditors. In the equilibria we study a disreputably audited firm always receives financing. The creditors end up paying the audit fee when the firm fails, and the equity holders pay when the firm succeeds. Under a contract in which an auditor's fee was lower in the bad state, and the expected fee was equilibrated, hence higher in the good state, the expected payment of the creditors would be less, and the expected payment of the equity holders would be more, to the extent that the probability of success was lower in the bad state.

[7] Naturally, other bargaining protocols imply different outcomes. It may be of interest to investigate other possibilities along these lines.

If the firm fails, then its value is zero and it makes no repayment of the loan. That is, we are assuming that the auditor's fee and bribe are treated as operating expenses, hence are paid regardless of whether the firm succeeds, but in the bankruptcy proceedings of a failed firm the assets are insufficient to repay all debt more senior than the loan under consideration. After either success or failure, the client firm leaves the market.

If a firm fails after obtaining an unqualified audit report, there is an investigation that determines its state. If the state was bad, but the auditor reported that it was good, then the auditor is forced to leave the industry. An auditor that is found to have misreported the state is allowed to keep the fee and bribe. (If we assumed that the auditor was forced to return the fee and bribe, certain algebraic expressions would change, of course, but the main features of our analysis would continue to hold.) All other auditors continue operation in the next period. The auditor's payoff is the sum over all periods of the discounted fees and bribes it receives.

Since client firms only live for a single period, we are assuming away the possibility that client firms and auditors may be able to engage in continuing relationships. In reality auditors and clients may have an incentive to stay together because an initial audit is more expensive than subsequent audits; as DeAngelo (1981) describes, this can result in lowballing. It is natural to suppose that a firm's type may be positively correlated from one year to the next. A continuing relationship presents various means of disguising bribes, and indeed the very concept of a bribe may not be well defined because for both parties the value of the relationship itself is always at stake. All of these seem to us to be interesting aspects of the industry. Our sense is that a range of models are possible and potentially interesting research projects, but we do not foresee that alternative models would cast a drastically different light on the issue.

Finally, we assume that honest audit is socially valuable. Specifically, we assume that

$$p_B h \;<\; \ell \;<\; p_G h. \tag{1}$$

When the first inequality does not hold all firms will be financed because even a firm that is known to be bad is a worthwhile gamble. When the second inequality fails none will be financed because even if a firm was known to be good, and the audit itself was costless, the loan amount would exceed the expected revenues from continued operation.

3. Equilibrium

Our focus is on *two-tier equilibria*, which are stationary equilibria in which there are reputable and disreputable auditors. A firm that is found to be in a bad state is willing to

pay a bribe that is large enough to induce a false report from a disreputable auditor, but is unwilling to pay a bribe large enough to elicit a false report from a reputable auditor. We should stress that the reputable auditor's behavior is an equilibrium phenomenon, not a matter of inherent honesty; an incentive compatibility condition will state that she declines any bribe any firm may offer only because it is insufficient to compensate for the risk to her reputation.

We represent a two-tier equilibrium by a five-tuple $(f_D, f_R, b, r_D, r_R) \in \mathbb{R}^5_+$ that satisfies the equilibrium conditions laid out below. The variables f_D and f_R are the fees offered to the disreputable and reputable auditors respectively, b is the bribe offered to disreputable auditors, and r_D and r_R are the repayment promises of the firms that have received unqualified audit reports from disreputable and reputable auditors respectively. We are interested in understanding, in particular, which values of m_R are consistent with two-tier equilibrium.

We first describe the equilibrium conditions for disreputable auditors. Since there is free entry into the disreputable segment, the value of a disreputable auditor's business is equal, in equilibrium, to the entry cost C. The bribe offered by a client firm in a bad state should be the minimum amount that compensates for the risk of getting caught, so the audit firm should be indifferent as to whether to accept the bribe. (Rejecting the bribe with positive probability is inconsistent with equilibrium because the client firm could offer slightly more to insure acceptance.) Therefore the value of the disreputable audit firm should be its discounted future revenue if it always refuses bribes, which is $f_D/(1-\delta)$ where $\delta \in (0,1)$ is the discount factor. If it accepts a bribe it survives into the next period with probability p_B, so it is indifferent if $b + p_B\delta C = \delta C$. Thus

$$b = (1-p_B)\delta C \quad \text{and} \quad f_D = (1-\delta)C. \tag{2}$$

The financial market is competitive and risk neutral, so the loan to the client firm will have a repayment amount such that the expected repayment is equal to ℓ. The auditor's report and her reputation are the only information available to the financial market at the time the financial contract is negotiated. Since reputable auditors always report truthfully, in equilibrium $p_G r_R = \ell$, i.e.,

$$r_R = \ell/p_G. \tag{3}$$

Similarly, the repayment amount for a disreputably audited firm will be

$$r_D = \ell/p_D. \tag{4}$$

where p_D is what the financial markets believe to be the probability that such a firm repays the loan.

In a two-tier equilibrium loans are extended to firms certified to be good in audit reports. These comprise firms that were indeed found to be good, and disreputably audited firms that were found to be bad. In both cases the repayment amount must be no more than the value of a successful client firm, net of audit fee and bribe. Certain uninteresting complications arise when these conditions hold with equality, and to avoid these we only consider the case where both hold strictly:

$$h - f_R > r_R \quad \text{and} \quad h - f_D - b > r_D. \tag{5}$$

The probability p_D that a disreputably audited firm repays the loan depends on the average type of disreputably audited firms. First of all, observe that the inequality above implies that there are no firms that decide to go out of business without incurring the audit fee because the value of going out of business is zero, but hiring an auditor results in a positive return with positive probability. If a firm of type $t \in (0,1)$ hires a reputable auditor its expected payoff is

$$tp_G(h - f_R - r_R). \tag{6}$$

The expected payoff resulting from hiring a disreputable auditor is

$$tp_G(h - f_D - r_D) + (1-t)p_B(h - f_D - r_D - b). \tag{7}$$

Both of these are continuous functions of t, so in order to have a two-tier equilibrium it must be the case that there is some type $\tilde{t} \in (0,1)$ that is indifferent between the two sorts of auditors, which is to say that (6) and (7) are equal when $t = \tilde{t}$. Since $h - f_D - r_D - b > 0$, this cannot be the case unless $h - f_R - r_R > h - f_D - r_D$. In turn, this implies that the expected payoff from hiring a reputable auditor increases more rapidly with t than the expected payoff of hiring a disreputable auditor, so a firm prefers to hire a reputable (disreputable) auditor if its type is above (below) $\tilde{t}$. (Since μ is atomless, what a firm of type $\tilde{t}$ does is inconsequential.)

Since μ is atomless and has full support, there is a unique number t^* such that $\mu((t^*,1)) = m_R$, and in a two-tier equilibrium we must have $\tilde{t} = t^*$. The average type of a disreputably audited firm is consequently

$$\tau(t^*) := \frac{\int_0^{t^*} t \, d\mu}{\mu((0,t^*))}, \tag{8}$$

and the probability that a disreputably audited firm repays its loan is

$$p_D = \tau(t^*)p_G + (1 - \tau(t^*))p_B.$$

The fee of reputable auditors, f_R, will be determined by the condition that a firm of type t^* is indifferent between the two types of auditors. Equating (6) and (7) and isolating f_R yields:

$$f_R = f_D + r_D - r_R - \frac{(1-t^*)p_B(h - f_D - b - r_D)}{t^* p_G}. \tag{9}$$

At this point we have shown (in (2), (3), (4), and (9)) how to determine the equilibrated variables–f_D, f_R, b, r_D, r_R–in a two-tier equilibrium.

It remains to ensure incentive compatibility. The equilibration of the bribe offered to a disreputable auditor has already been described. We now discuss two other types of possible deviation.

Consider the possibility that a client firm might offer a fee other than f_R or f_D. Hypothetically an audit firm's willingness to accept such an offer might depend on the induced belief about the client firm's type because it affects expectations about how the subsequent interaction will unfold. However, client firms never offer bribes that do more than make the audit firm indifferent about acceptance, so in equilibrium the expected payoff from accepting a client firm's fee offer does not depend on the client firm's type. This justifies our assumption that each type of auditor has a fee that is equilibrated in the market, and that client firms cannot profit by deviating.

The value of a reputable auditor's future stream of fees is $\delta f_R/(1-\delta)$, so a reputable auditor would accept a bribe if it exceeded $(1-p_B)\delta f_R/(1-\delta)$, which is the probability of being caught times the loss in that event. Since a reputably audited firm would be willing to pay a bribe up to $h - f_R - r_R$ upon being revealed to be bad, incentive compatibility for a reputable auditor requires

$$(1-p_B)\frac{\delta}{1-\delta} f_R \;\geq\; h - f_R - r_R. \tag{10}$$

This is the case if and only if $f_R \geq f_R^{\min}$ where

$$f_R^{\min} \;:=\; \frac{h - r_R}{1 + (1-p_B)\frac{\delta}{1-\delta}}.$$

Definition 1: *A two-tier equilibrium is a five-tuple (f_D, f_R, b, r_D, r_R) that satisfies (2)–(5) and (8)–(10).*

Other Equilibria. Before analyzing two-tier equilibria, we discuss other possible equilibria. A trivial and uninteresting equilibrium is a collapse of the reputation mechanism, resulting in a situation in which all auditors are disreputable, i.e., $m_R = 0$. In this equilibrium, f_D and b are determined by (2) and $r_D = \ell/\tau(1)$.

Another possibility is that $h - f_D - r_D < 0$, in which case disreputable auditors will not be hired since hiring one cannot lead to a situation in which continued operation of the firm is worthwhile. One might expect the fee of reputable auditors to rise to the point where all firms are indifferent between immediate bankruptcy and hiring a reputable auditor, so that $h - f_R - \ell/p_G = 0$, hence $f_R = h - \ell/p_G$. At this fee all firms are indifferent as to whether to hire a reputable auditor, and which firms actually hire auditors is indeterminate. Moreover, a firm whose state has been found to be bad is unwilling to pay any bribe, since all the surplus has already gone into the audit fee. Such equilibria are quite extreme, with very high audit fees and bankruptcy of many firms for which it would be worth the expense of an audit if there could be a credible report in the event that the state was found to be good, so they might be regarded as unrealistic in the sense that there would be pressure for institutional reforms that would undermine the features of the model leading to this outcome.

Next, one can think of variations of two-tier equilibria, in which the identities of reputable and disreputable auditors change across periods and the manner in which they do is common knowledge. For example, it is commonly known that one group of auditors are reputable in all odd periods and another group are in all even periods. Although other variations are feasible, such changes of identities, in addition to being hardly justifiable, mean that reputable auditors are more susceptible to bribes because their future income prospects are lower. Also note that such variations are inconsistent with the aforementioned notion of ours that firm's reputation is historically determined market's perception which cannot be changed in the short-run.

Finally, one might ask whether there could be equilibria involving a third class of auditors with an intermediate reputation. Since they sometimes accept bribes and sometimes refuse, a client in a bad state must be unwilling to induce sure acceptance by slightly increasing the bribe. Thus the client's option of offering a bribe has no value, so the difference between this auditor's fee and the fee of the reputable auditor must exactly offset the difference in reputation. Varying the intermediate reputation between the two extremes, the intermediate value theorem implies the existence of a reputation at which the client firm is indifferent about offering the minimum bribe the auditor would accept, so such equilibria exist. But such an equilibrium is highly unstable: a small improvement (decline) in the intermediate auditor's reputation results in refusal (sure acceptance) of bribes, due to the increased (decreased) value of future fees, after which this group would merge with the reputable (disreputable) auditors.

We note that of all equilibria above, honest audit is most widely sustained in a two-tier equilibrium because the incentive compatibility is most easily satisfied. That is, in

any environment if reputable auditors are viable in any of the equilibria discussed above, so are they in a two-tier equilibrium.

4. Equilibrium Sizes of Reputable Auditors

In this section we describe the range of parameter values that permit a two-tier equilibrium. Formally our model is static, but our inclination is to view it as describing the state of a slowly changing industry at a point in time. As we explained in the introduction, informal and anecdotal evidence suggests that the fraction of the market hiring reputable auditors has been increasing from one era to the next, and our main focus is on how this expansion might eventually result in collapse because parameter values move into a region that does not support two-tier equilibria.

There are two main reasons that a two-tier equilibrium may cease to be viable. First, the average quality of disreputably audited firms may decline to the point that there is no repayment promise that attracts financing. Second, because the fees of reputable auditors are determined by the marginal client firm's indifference between the two types of auditors, if the quality of the marginal client firm declines, the fees of reputable auditors may decline to the point where they become willing to take bribes. Our analysis shows that in the simplest and most typical case, the first possibility takes precedence over the second, so that collapse of the disreputable segment occurs before reputable auditors are seriously tempted. This may seem counterfactual, insofar as the collapse of Arthur Andersen was not an indirect effect of the collapse of the disreputable. Perhaps the simplest and most plausible way to reconcile this observation with the analysis here is to observe that our model assumes that all firms are of the same size, but Enron was much larger than all but a few American corporations, and therefore capable of offering very large bribes. However, we will also sketch a more complicated scenario in which audit firms become tempted to accept bribes before the disreputable segment of the market evaporates.

The equilibrium conditions listed in Definition 1, apart from (5) and (10), simply determine the equilibrated variables. Examining the incentive compatibility conditions surrounding bribe offers, (5) and (10), this section characterizes the environments in which a two-tier equilibrium exists and thereby, the viable sizes of reputable auditors.

In a two-tier equilibrium disreputably audited firms must be willing to pay the bribe b, which will be the case if (5) holds. Note that $\tau(t^*)$ is an increasing function of t^*, and the equilibrium repayment amount when the probability of a good state is $\alpha \in (0,1)$,

$$\rho_D(\alpha) := \frac{\ell}{\alpha p_G + (1-\alpha)p_B},$$

is a decreasing function of α, so there is a unique number $\underline{t} > 0$ such that

$$\rho_D(\tau(\underline{t})) = h - f_D - b = h - (1 - p_B\delta)C. \tag{11}$$

Disreputably audited firms will be able to obtain financing that justifies paying a bribe when $t^* \geq \underline{t}$ and not otherwise.

As for reputable auditors, the incentive compatibility condition under which they cannot be bribed is (10). Combining this with (9) where $r_D = \rho_D(\tau(t^*))$, an equilibrium in which reputable auditors do not accept bribes exists if and only if $t^* \geq \underline{t}$ and

$$f_D + \rho_D(\tau(t^*)) - r_R - \frac{(1-t^*)p_B(h - f_D - b - \rho_D(\tau(t^*)))}{t^* p_G} \geq f_R^{\min}.$$

Isolating $\rho_D(\tau(t^*))$, this inequality is equivalent to $\rho_D(\tau(t^*)) \geq \overline{\rho}(t^*)$ where, for $t \in (0,1)$,

$$\overline{\rho}(t) := \frac{f_R^{\min} + r_R - h + b}{1 + \frac{(1-t)p_B}{tp_G}} + h - f_D - b \tag{12}$$

is the repayment amount for disreputably audited firms that, for t to be the threshold type, results in a fee for a reputable auditor that makes her indifferent about whether to accept the maximum bribe a client firm would be willing to pay.

Lemma 1: *$f_R^{\min} + r_R - h + b < 0$, so that $\overline{\rho}$ is a decreasing function.*

Proof: Above we assumed the existence of $\underline{t}$ satisfying (11), so that $h - f_D - r_R > h - f_D - \rho_D(\tau(\underline{t})) = b$. Thus $h - r_R - b > f_D$. Since b is just sufficient to compensate an auditor for the risk of losing a stream of revenues of f_D in each future period, b is not sufficient to induce acceptance by an auditor who could receive a fee of $h - r_R - b$ in each future period. Since b would be the maximum possible bribe if $f_R = h - r_R - b$, the minimum level of f_R that deter acceptance of bribes is lower, i.e., $f_R^{\min} < h - r_R - b$. ∎

Note that $\overline{\rho}(t^*)$ decreases from $h - f_D - b$ to $f_R^{\min} - f_D + r_R$ as t^* increases from 0 to 1, and that $\rho_D(\alpha) = \ell/(\alpha p_G + (1-\alpha)p_B)$ decreases monotonically from ℓ/p_B to ℓ/p_G as α increases from 0 to 1. We claim that

$$\frac{\ell}{p_G} < h - f_D - b < \frac{\ell}{p_B} \quad \text{and} \quad \frac{\ell}{p_G} < f_R^{\min} - f_D + r_R < \frac{\ell}{p_B},$$

so that the image of $\overline{\rho}$ is contained in the interior of the image of ρ_D, which means that we may define $\overline{\tau} : (0,1) \to (0,1)$ by $\overline{\tau}(t^*) := \rho_D^{-1}(\overline{\rho}(t^*))$. Of the four asserted inequalities, the first follows from (5) and $r_D > r_R = \ell/p_G$, while the second follows from (1). The third

follows from $\ell/p_G = r_R$ and $f_R^{\min} > f_D$, while the fourth follows from $f_R^{\min} < h - r_R$ and $h < \ell/p_B$. Since ρ_D and $\overline{\rho}$ are both decreasing functions, we have:

Lemma 2: *$\overline{\tau}$ is an increasing function.*

Summarizing the discussion above:

Theorem 1: *There exists a two-tier equilibrium if and only if $t^* \geq \underline{t}$ and $\rho_D(\tau(t^*)) \geq \overline{\rho}(t^*)$, i.e., $\tau(t^*) \leq \overline{\tau}(t^*)$.*

The inequality $\tau(t^*) \leq \overline{\tau}(t^*)$ holds whenever $\overline{\tau}(t^*) \geq t^*$, since $t^* \geq \tau(t^*)$. The case $\overline{\tau}(t^*) < \tau(t^*) \leq t^*$ corresponds to the possibility that the financial markets extend loans to disreputably audited firms at terms that are only slightly worse than the repayment terms for reputably audited firms, with reputable auditors receiving a premium that is insufficient to dissuade them from accepting a bribe from a firm that is revealed to be bad. Assuming that the reputation mechanism collapsed, the distribution of types of disreputably audited firms would become even more favorable. The parameter values leading to this outcome are extreme (specifically, the distribution of firm types is so heavily concentrated near t^* and t^* is close to one) and this aspect of the model seems both unrealistic and unproblematic. However, it leads to a certain qualification of our comparative statics: changing δ has an unambiguous effect on $\underline{t}$, but we are not able to say that it has a monotonic (in the sense of set inclusion) effect on the set of values of t^* supporting a two-tier equilibrium.

To prepare for the discussion of the possible consequence of changes in m_R and δ we develop a sufficient condition (Theorem 2 below) for the existence of a two-tier equilibrium. This will have the implication that for any specification of h, ℓ, C, p_B, p_G, and δ such that firms known to be good receive financing and those known to be bad do not, there is a nonempty interval such that a two-tier equilibrium exists for every m_R in this interval.

Lemma 3: *There is a unique number $0 < \hat{t} < 1$ such that $\overline{\tau}(\hat{t}) = \hat{t}$.*

Proof: Since $\overline{\tau} : (0,1) \to (0,1)$ is continuous and its image is a strict subset of $(0,1)$, existence of a fixed point follows from the intermediate value theorem. It is easy to see that such a function $\overline{\tau}$ must have a unique fixed point if

$(*)$ $\overline{\tau}$ is strictly convex, strictly concave, or affine over its entire domain.

The remainder of proof shows $(*)$. The definitions of $\overline{\tau}$ and ρ_D, together with (12), imply that

$$\begin{aligned}\frac{\ell}{p_B + \overline{\tau}(t)(p_G - p_B)} &= \frac{f_R^{\min} + r_R - h + b}{\frac{p_B + t(p_G - p_B)}{tp_G}} + h - f_D - b \\ &= \frac{tp_G(f_R^{\min} + r_R - h + b)}{p_B + t(p_G - p_B)} + h - f_D - b.\end{aligned}$$

To simplify notation let $\beta = p_G(f_R^{\min} + r_R - h + b)$, $\gamma = h - f_D - b$ and $\Delta p = p_G - p_B$. Substituting, then rearranging, yields

$$\overline{\tau}(t)\Delta p = \frac{\ell(p_B + t\Delta p)}{t(\beta + \gamma\Delta p) + \gamma p_B} - p_B. \tag{13}$$

Differentiating (13) with respect to t twice, we get

$$\begin{aligned} \overline{\tau}'(t)\Delta p &= \frac{t(\beta+\gamma\Delta p)\Delta p + \gamma p_B \Delta p - (p_B + t\Delta p)(\beta+\gamma\Delta p)}{\big(t(\beta+\gamma\Delta p) + \gamma p_B\big)^2} \cdot \ell \\ &= \frac{\gamma p_B \Delta p - p_B(\beta+\gamma\Delta p)}{\big(t(\beta+\gamma\Delta p) + \gamma p_B\big)^2} \cdot \ell \;=\; \frac{-\ell\beta p_B}{\big(t(\beta+\gamma\Delta p) + \gamma p_B\big)^2} \end{aligned} \tag{14}$$

and

$$\overline{\tau}''(t)\Delta p \;=\; \frac{2\ell\beta p_B\big(t(\beta+\gamma\Delta p) + \gamma p_B\big)(\beta+\gamma\Delta p)}{\big(t(\beta+\gamma\Delta p) + \gamma p_B\big)^4} \;=\; \frac{2\ell\beta p_B(\beta+\gamma\Delta p)}{\big(t(\beta+\gamma\Delta p) + \gamma p_B\big)^3}. \tag{15}$$

Note from (13) that, for all $t \in (0,1)$, $t(\beta+\gamma\Delta p) + \gamma p_B > 0$, since otherwise $\overline{\tau}(t) < 0$ contrary to the definition of $\overline{\tau}$. Hence, for all t, $\overline{\tau}''(t)$ has the same sign as the constant $\beta(\beta+\gamma\Delta p)$. ∎

If $\tau(1) > \hat{t}$ let $\bar{t} := \tau^{-1}(\hat{t})$, and otherwise set $\bar{t} := 1$. We now have the following result:

Theorem 2: *For every m_R such that $\mu((\bar{t},1)) \le m_R \le \mu((\underline{t},1))$, there is a two-tier equilibrium in which the measure of reputable auditors is m_R.*

Proof: Recall that $m_R = \mu((t^*,1))$ so that $\mu((\bar{t},1)) \le m_R \le \mu((\underline{t},1)) \;\Leftrightarrow\; \underline{t} \le t^* \le \bar{t}$. We wish to show that $\underline{t} \le t^* \le \bar{t}$ implies that $\tau(t^*) \le \overline{\tau}(t^*)$. For $0 \le t^* \le \hat{t}$ we have $\tau(t^*) \le t^* \le \overline{\tau}(t^*)$. For $\hat{t} \le t^* \le \bar{t}$ we have $\tau(t^*) \le \hat{t} = \overline{\tau}(\hat{t}) \le \overline{\tau}(t^*)$. ∎

The interval $(\underline{t}, \bar{t})$ is always nonempty:

Lemma 4: $\underline{t} < \bar{t}$.

Proof: If $\tau(1) \le \hat{t}$, then $\bar{t} := 1 > \underline{t}$. Otherwise $\hat{t} = \tau(\bar{t})$, so that the claim follows from the monotonicity of τ once we show that $\tau(\underline{t}) < \hat{t}$. In light of Lemma 1, (12) implies that $\overline{\rho}(\tau(\underline{t})) < h - f_D - b$, whence $\rho_D^{-1}(\overline{\rho}(\tau(\underline{t}))) > \rho_D^{-1}(h - f_D - b)$ because ρ_D^{-1} is decreasing. The definition of $\overline{\tau}$ gives $\overline{\tau}(\tau(\underline{t})) = \rho_D^{-1}(\overline{\rho}(\tau(\underline{t})))$, and (11) implies that $\rho_D^{-1}(h - f_D - b) = \rho_D^{-1}(\rho_D(\tau(\underline{t}))) = \tau(\underline{t})$. Therefore $\overline{\tau}(\tau(\underline{t})) > \tau(\underline{t})$, and consequently $\tau(\underline{t}) < \hat{t}$ since $\hat{t}$ is the unique fixed point of $\overline{\tau}$. ∎

The picture developed by these results is shown in Figure 1.

[Insert Figure 1 near here.]

To understand the graphical derivation of $\underline{t}$ observe that, by (12), $\overline{\rho}(0) = h - f_D - b$. From (11), therefore, $\tau(\underline{t}) = \rho_D^{-1}(h - f_D - b) = \rho_D^{-1}(\overline{\rho}(0))$, and $\overline{\tau}(0) = \rho_D^{-1}(\overline{\rho}(0)) = \tau(\underline{t})$ follows from the definition of $\overline{\tau}$. It is possible for $\underline{t}$ to be less than, equal to, or greater than $\hat{t}$. The graph of τ can cross the graph of $\overline{\tau}$, perhaps more than once, as shown, or it may happen that $\tau(t) < \overline{\tau}(t)$ for all $0 \leq t \leq 1$. If the graph of $\overline{\tau}$ moved up when δ increased, increasing δ would unambiguously expand the set of sizes of reputable auditors for which a two-tier equilibrium exists by virtue of the next result. But examples show that it is possible that an increase in δ can result in the graph of $\overline{\tau}$ moving down in some intervals.

Although we treat δ as a discount factor, it can be construed more generally as a measure of auditor independence, insofar as its function in the model is primarily to measure the relative importance of the auditor's expected fees from other clients in comparison with the temptations offered by the current client. Note that the definition of τ is independent of δ.

Lemma 5: *$\underline{t}$ is a decreasing function of δ.*

Proof: From (11) we have

$$\underline{t} = \tau^{-1}(\rho_D^{-1}(h - (1 - p_B\delta)C)).$$

Since τ is an increasing function, ρ_D is a decreasing function, and neither function changes when δ varies, this quantity decreases when δ increases. ∎

Remark 1: For expositional ease we conduct our analysis presuming that external audit is necessary for the firm to seek outside financing. However, our main findings are not sensitive to this presumption. In the two-tier equilibrium analyzed above, disreputable auditors' reports carry no informational content for the financial market, hence can be dispensed with unless a mandatory audit requirement is in place, without affecting the equilibrium information transmission. However, this feature is an artifact of the two-state model we adopted for expositional ease, and disappears, for instance, when the model is minimally extended to allow for a third state, called say ugly, as follows: In addition to good and bad states, a firm may assume an ugly state with a fixed probability $t_U > 0$, in which case the success probability (if a loan ℓ is obtained and the project is undertaken) is p_U where $p_U < p_B$. The analysis above in this section extends straightforwardly to this case so long as p_U is not too large, and a two-tier equilibrium exists in which all

firms hire auditors of their choice, i.e., all firms of type t^* or above hire reputable auditors who make qualified reports when the firm's state is bad or ugly, in which case the firm cannot get a loan; all other firms hire disreputable auditors who issue unqualified reports for good firms (without a bribe), accept a bribe and issue unqualified reports for bad firms, and do not accept a bribe and issue qualified reports for ugly firms. Unlike in the two-state model, firms hire disreputable auditors voluntarily because their reports carry informational contents that warrant more favorable treatments in the financial market than what firms would get without an audit report.

We close this section with a brief comparison of our model with two other models in the accounting literature, of how the Big n auditors might be different from other auditors. DeAngelo (1981) describes a world in which competition in the market for new clients drives expected profits, at the time of first audit, to zero: specifically, auditors offer a fee for the initial audit that is below cost such that the loss will be covered by future quasi-rents from existing clients who would incur costs if they switched auditors. This practice, which is known as *lowballing*, implies that relationships with existing clients are valuable to the audit firm and thus, that large firms with more clients have more to lose by behaving dishonestly because then more of such relationships would be put to risk.

However, if the capitalized value of existing accounts is the only force motivating differential honesty and there is free entry into the high quality tier of the auditing industry, so that the two types of auditors have equal expected fees, then one should not expect to observe both large and small audit firms. This is so because the "best" type in the pool choosing a disreputable auditor would prefer to be audited by a reputable auditor, given that the fees are the same, since the alternative is to be treated as an average member of the pool. Such a process of continued expansion of the high quality tier would lead eventually to the evaporation of the market for disreputable auditors. This point is made in great generality by Okuno-Fujiwara, et. al. (1990).

Dye (1993) develops what has come to be known as "the deep pockets hypothesis" in a model in which wealthier audit firms are more strongly motivated to be diligent in their investigations of client firms because their greater wealth makes them more vulnerable to lawsuits. Even if both large and small auditors are incorporated, hence protected by limited liability[(8)], smaller firms may be presumed to have much less wealth, so that a judgment against the auditor in a lawsuit might be a hollow victory for the plaintiff. One significant point emerges from Dye's model: if the deep pockets hypothesis was the only important factor, large auditors would dominate small ones, and drive them out of the

[(8)] In fact Arthur Andersen was a partnership.

market, unless small auditors enjoyed a cost advantage. We do not know of formal studies supporting the view that such cost differentials are substantial, but possibly this is due to difficulty in obtaining data.[(9)]

5. Conclusion

This paper develops a model in which the audit firms previously endowed with a reputation for honesty do, in fact, behave honestly, because they are afraid to jeopardize their reputation. The value of reputation is proportional to the difference between the fees charged by reputable auditors and those charged by auditors without prior reputation. In equilibrium this fee differential leaves the marginal client firm indifferent between the two types of auditors. High quality client firms have a stronger preference for reputable auditors, so that expansion of the reputable segment of the market leads to a marginal client firm with lower willingness to pay for a high quality audit, diminishing the fee differential. Lowering the quality of the marginal firm also results in a lower average quality for disreputably audited firms, eventually driving disreputable auditors out of business. Either effect undermines the mechanism sustaining honest behavior of reputable auditors if the supply of high quality auditing passes a certain threshold, although we find it likely that the latter effect takes precedence over the former.

A large body of empirical research examines the hypotheses that large audit firms charge higher fees and are viewed as more reliable by actors in financial markets.[(10)] Several studies support these conclusions, but the evidence is mixed, suggesting that the work to date has not settled these questions definitively. The data of Danos and Eichenseher (1986) and Wolk, Michelson and Wootton (2001) suggest that, as a result of internal expansion of the top firms, the market share of Big n auditors has increased substantially during recent decades.

Our comparative statics result may be viewed as supportive of policy reforms (e.g., prohibitions on auditors supplying consulting services or periodic mandatory rotation of

[(9)] Dye (1993, p. 890) mentions that liability can arise out of failure to follow proper procedures, and that such failure can result in lower costs. For example, Generally Accepted Auditing Standards require that auditors inspect warehoused inventories. But one might expect (we have no specific information) that to a large and increasing extent the relevant data is already available to the auditor in electronic form, and can be manipulated with no more effort than is required to modify a spreadsheet. In addition, the hypothesis that wealthy firms use more expensive procedures is counter to the usual presumption of scale economies. On the other hand, even prior to the Enron affair there was substantial anecdotal evidence (e.g., Grout et. al. (1994)) concerning the importance of pressure to give a favorable presentation of the client firm's accounts. In the contemporary popular press there is little discussion of costs of diligence.

[(10)] The list includes Simunic (1980), Francis (1984), Francis and Stokes (1986), Palmrose (1986, 1988), Francis and Simon (1987), Balvers et. al. (1988), Simon and Francis (1988), Beatty (1989), Ettredge and Greenberg (1990), Stice (1991), Firth and Smith (1992), Chan et. al. (1993), Pong and Whittington (1994), Craswell et. al. (1995), and Ireland and Lennox (2002), Chaney et. al. (2004).

auditors) that enhance auditor independence by diminishing anticipated revenues from the client firm, as a fraction of the auditor's current and future business. But such reforms are a stop gap response to the erosion of the value of reputation that results from an increase in the market share of the top firms. This suggests giving consideration to more drastic reforms, such as requiring publicly traded firms to purchase audit insurance, with the auditor hired by the insurer, in order to align the auditor's incentives with its true "customers," namely those who base financial decisions on audit reports.

References

Allen, F. (1984). "Reputation and product quality," *RAND Journal of Economics* **15**, 311-327.

Arrow, K. (1973). "The Theory of Discrimination," in O. Ashenfelter and A. Rees, ed., *Discrimination in Labor Markets,* Princeton, NJ: Princeton University Press, pp. 3–33.

Balvers, R., McDonald, B. and Miller, R. (1988). "Underpricing of new issues and the choice of auditors as a signal of investment banker reputation," *The Accounting Review* **63**, 605–621.

Beatty, R. (1989). "Auditor reputation and the pricing of initial public offerings," *The Accounting Review* **64**, 693–709.

Bolton, P., Freixas, X., and Shapiro, J. (2012). "The Credit Ratings Game," *Journal of Finance* **67**, 85–111.

Chan, P., Ezzamel, M. and Gwilliam, D. (1993). "Determinants of audit fees for quoted UK companies," *Journal of Business, Finance and Accounting* **20**, 765–786.

Chaney, P. K., Jeter, D. and Shivakumar, L. (2004). "Self-Selection of Auditors and Audit Pricing in Private Firms," *The Accounting Review* **79**, 51–72.

Coate, S., and Loury, G. C. (1993). "Will Affirmative-Action Policies Eliminate Negative Stereotypes?," *Amer. Econ, Rev.* **83**, 1220–1240.

Crawell, A., Francis, J. and Taylor, S. (1995). "Auditor brand name reputations and industry specializations," *Accounting and Business Research* **19**, 23–31.

Danos, P., and J. W. Eichenseher (1986). "Long-term trends toward seller concentration in the U.S. audit market," *The Accounting Review* **61**, 633–650.

DeAngelo, L. (1981). "Auditor size and audit quality," *Journal of Accounting and Economics* **3**, 183-199.

Dye, R. (1993). "Auditing standards, legal liability and auditor wealth," *J. Pol. Econ.* **101**, 887–914.

Edwards *History of Accounting in the United States*, MSU Business Studies, Easy Lancing Michigan, 1960.

Ettredge, M., and Greenberg, R. (1990). "Determinants of Fee Cutting on Initial Audit Engagements," *Journal of Accounting Research*, **28**, 198–210.

Firth, M. and Smith, A. (1992). "Selection of auditor firms by companies in the new issue market," *Applied Economics* **24**, 247–255.

Francis, J. (1984). "The Effect of Audit Firm Size on Audit Prices: A Study of the Australian Market," *Journal of Accounting and Economics,* **6**, 133–151.

Francis, J. and Simon, D. (1987). "A Test of Audit Pricing in the Small-Client Segment of the U.S. Audit Market," *The Accounting Review*, **62**, 145–157.

Francis, J. and Stokes, D. (1986). "Audit Prices, Product Differentiation, and Scale Economies: Further Evidence from the Australian Audit Market," *Journal of Accounting Research,* **24**, 383–393.

Grout, P., Jewitt, I., Pong, C. and Whittington, G. (1994). "'Auditor Professional Judgement': Implications for Regulation and the Law," *Economic Policy,* Issue 19, 308–344.

Ireland, J. and Lennox, C. (2002). "The Large Audit Firm Fee Premium: A Case of Selectivity Bias?," *Journal of Accounting, Auditing and Finance,* **17**, 73–91.

Klein, B. and Leffler, K. (1981). "The role of market forces in assuring contractual performances," *J. Pol. Econ.* **89**, 615–641.

Kreps, D. and Wilson, R. (1982a). "Reputation and Imperfect Information," *Journal of Economic Theory* **27**, 253–279.

Kreps, D. and Wilson, R. (1982b). "Sequential Equilibrium," *Econometrica* **50**, 863–894.

Mathis, J., McAndrews, J. and Rochet, J.-C. (2009). "Rating the Raters: Are Reputation Concerns Powerful Enough to Discipline Rating Agencies," *Journal of Monetary Economics* **56**, 657–674.

Milgrom, P. and Roberts, J. (1982). "Predation, Reputation, and Entry Deterrence," *Journal of Economic Theory* **27**, 280–312.

Okuno-Fujiwara, M., Postlewaite, A., and K. Suzumura (1990). "Strategic Information Revelation," *Review of Economic Studies* **57**, 25–47.

Palmrose, Z. (1986). "Audit Fees and Auditor Size: Further Evidence," *Journal of Accounting Research,* **24**, 97–110.

Palmrose, Z. (1988). "An analysis of auditor litigation and audit service quality," *The Accounting Review* **63**, 55–73.

Phelps, E. (1972). "The Statistical Theory of Racism and Sexism," *Amer. Econ. Rev.* **62**, 659-661.

Pong, C. and Whittington, G. (1994). "The determinants of audit fees: some empirical models," *Journal of Business Finance & Accounting* **21**, 1071–1095.

Rogerson, W. (1983). "Reputation and product quality," *BELL Journal of Economics* **14**, 508–516.

Simon, D., and Francis, J. (1988). "The Effects of Auditor Change on Audit Fees: Tests of Price Cutting and Price Recovery," *The Accounting Review*, **63**, 255-269.

Simunic, D. (1980). "The Pricing of Audit Services: Theory and Evidence," *Journal of Accounting Research,* **18**, 161–190.

Shapiro, C. (1983). "Premiums for high quality products as returns to reputations," *Quarterly Journal of Economics,* **98**, 659–679.

Stice, J. (1991). "Using financial and market information to identify pre-engagement factors associated with lawsuits against auditors," *The Accounting Review* **66**, 516–533.

Wilson, R. (1983). "Auditing: perspectives from multi-person decision theory," *The Accounting Review* **58**, 305–318.

Wolk, C. M., Michelson, S. E., and C. W. Wootton (2001). "Auditor concentration and market shares in the US: 1988-1999 a descriptive note," *British Accounting Review* **33**, 157–174.

Figure 1