

The Market for Liars: Reputation and Auditor Honesty *

by

Andrew McLennan

Department of Economics
University of Minnesota
271 19th Avenue South
Minneapolis, MN 55455

`mclennan@atlas.socsci.umn.edu`

and

In-Uck Park

Department of Economics
University of Bristol
8 Woodland Road
Bristol BS8 1TN, U.K.

`i.park@bristol.ac.uk`

June 2003

this version: **October 2004**

* 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 the Institute for Social Economic Research (University of Osaka) Keio University, the University of Pittsburgh, the University of Illinois, 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).

Running Title: **The Market for Liars: Reputation and Auditor Honesty**

Corresponding Author:

In-Uck Park

Department of Economics

University of Bristol

8 Woodland Road

Bristol BS8 1TN, U.K.

`i.park@bristol.ac.uk`

Journal of Economic Literature Classification Numbers M41, D82, G14.

The Market for Liars: Reputation and Auditor Honesty

by

Andrew McLennan and In-Uck Park

“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.

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 does not create a good reputation. The main findings are: (a) honest auditors charge higher fees, and have economic profits accruing to reputation; (b) as the fraction of auditors who are honest increases, the premium charged by reputable auditors eventually decreases, which diminishes the incentive to refuse bribes; (c) if the fraction of honest auditors exceeds an upper bound, there does not exist a two-tier equilibrium; (d) thus the reputation mechanism may be undermined by entry into the honest segment of the industry, if it is possible; (e) increasing auditor independence increases the upper bound.

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, 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 empirical literature is surveyed in Section 5.) The membership of this group 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 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

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

(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.

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 other ways so that the desire to preserve one’s reputation is the only motivation for behaving in the manner one’s reputation predicts. Wilson (1983) proposes that auditor’s reputations are of this sort; many of the themes announced in that paper take on concrete and particular forms in our work. In contrast with some other models of reputation (e.g., Kreps and Wilson (1982), Milgrom and Roberts (1982), Rogerson (1983)) there is no hidden state variable, e.g., a “commitment type” or capital suitable for making products of a particular quality, that persists across periods.

Klein and Leffler (1981), Shapiro (1983), and Allen (1984) are also models of pure reputation in this sense. 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 greater value. The provision of high quality goods is not guaranteed, even when their provision is socially desirable, either because the fly by night strategy is not adequately deterred, or because entry into the industry can undermine the stream of future returns.

Shapiro (1983) provides a precise model in which firms must establish a reputation by producing high quality goods at a loss for one period. Even though firms with established reputations expect a stream of future revenues in excess of cost, and even though all firms have the same technology, which involves only variable costs, the price of the high quality goods stays above its cost of production because the cost of establishing a reputation is a sufficient barrier to new entry. In Allen (1984) there is an entry cost, and in equilibrium the sum over future periods of the difference between cost and revenue exactly covers this cost⁽⁴⁾. Allen also considers the possibility that consumers may observe the firm’s

⁽⁴⁾ This assumption seems conceptually superior to Shapiro’s device for creating reputation since, in Shapiro’s model, consumers do not expect high quality from producers who have not produced high quality before, even though (at least in the equilibration process) some producers do provide high quality without

quantity, and refuse to buy if, for the observed quantity, a deviation to low quality would be more profitable. Although one might expect that such monitoring by consumers would make it easier to sustain high quality production, for some parameter values observability undermines the existence of a high quality equilibrium.

Our model resembles those of Klein and Leffler, Shapiro, and Allen, insofar as there is the possibility that a high quality producer may switch to low quality, but it differs in certain respects that reflect specific conditions of the auditing industry. First of all, we assume that the supply of high quality auditors is determined exogenously. In contrast to the models of Shapiro and Allen, in which the provision of high quality is equilibrated by entry, and also to earlier theoretical studies (DeAngelo (1981) and Dye (1993)) of auditing, the form of reputation studied here depends on an entry barrier that prevents competition from eroding the profits of the reputable auditors.

Several considerations suggest that it is reasonable to assume that entry into the Big n is impossible. There is the reality that for at least the last half century there has, in fact, been no entry into the Big n , and Big n firms are highly profitable, suggesting that entry barriers are at least extremely high, if not literally insurmountable. One method of creating a reputation for high quality is to post a bond, but it seems plausible that the amounts of money at stake may be far beyond the wealth of even the largest potential entrants. An interesting paradox is that if there were some rule of longevity by which auditors who had not suffered a scandal for a long period, say fifty years, were inducted into the Big n , firms who were close to this threshold would have a strong incentive to behave honestly, and this would be known by others, so that they would, in effect, already have attained a Big n reputation. Thus no such rule is sustainable in an equilibrium in which Big n auditors are more profitable than others. (Section 4 discusses additional arguments supporting the existence of barriers to entry.)

Another special feature of the auditing industry stressed by Wilson (1983) and present in our model, but not in earlier work on pure reputation, 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. In particular, the actual quality provided by disreputable audit firms is endogenous, being a matter of the average quality of their clients. This is the primary source of the main conclusions derived in our analysis, since it points to two ways in which an equilibrium with both types of auditors may fail to exist. First, it may be impossible to sustain an equilibrium in which a large fraction of

having done so previously.

the market is served by high quality auditors, since the quality of the marginal client firm (i.e., its probability of being found to be viable) is then low. For such a client, purchasing high quality audit services has a low probability of resulting in ultimate success, so it is willing to pay only a small fee differential to obtain high quality auditing. An equilibrium with both types of auditors will collapse if high quality auditors are offered “bribes” (in practice these can take the form of continued patronage or purchase of consulting services) in excess of their expected future profits, which are proportional to this fee differential.

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. There seems to be little reason to doubt that this expansion has been as inexorable in recent decades as it was earlier. Our formal model considers only stationary equilibrium with static underlying parameters, but one may interpret it as providing a sequence of snapshots of this growth process. Specifically our assumption that the size of the reputable segment is fixed approximates 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 view, this slow expansion was accompanied by a diminution of the fee differential, which led eventually to a situation in which Arthur Andersen was tempted to risk its reputation.

The collapse of Arthur Andersen led to much discussion of reforms aimed at strengthening “auditor independence.” Such reforms may postpone the collapse of the reputation mechanism, but we know of no proposed reform that would arrest the internal expansion of the Big n firms. 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 such exits from the Big n create antitrust concerns that become increasingly severe as n shrinks.

A second way in which the fee differential may be too small is that the pool of client firms is very good, even after removing the client firms audited by reputable auditors, so that the financial markets believe that a client firm audited by a disreputable auditor is still very likely to be a good investment. That is, reputable auditing is unimportant, and the welfare loss resulting from its collapse quite small, because the phenomenon it detects is rare. Of course this second possibility seems more an artifact of theory than a relevant concern!

The remainder has the following structure. Sections 2 and 3 present and analyze the formal model of financial auditing. Section 4 contrasts our model with related theoretical

research on the auditing profession. Section 5 reviews related empirical literature, and Section 6 concludes.

2. The Model

This section describes the formal model of the auditing industry, and the different types of equilibria that are possible.

In each 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. The firm has an unknown *state*, which may be either “good” or “bad,” so the set of possible states is $S := \{G, B\}$. The firm hires an auditor who investigates the firm’s condition, and this investigation reveals the state of the firm to both the firm and the auditor, but not to the financial market. The auditor then publicly states whether the firm is good or bad. The auditor must state that the condition of the firm is good when that is in fact the case. If the state is found to be bad, the audit firm may state this (in the terminology of the industry, the audit firm “qualifies the accounts”) but there is also the possibility that the client may offer inducements to the audit firm to not qualify the accounts, as we explain in more detail below. Thus, when the auditor reports that the firm’s state is bad, that is certainly the case, but in response to a favorable report the financial market must assess the probability that the state is, in fact, good.

After the auditor issues her report, the repayment promise r required for the firm’s loan is determined in the financial market. (It may happen that there is no repayment promise sufficient to induce lenders to make the loan.) 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 . (The state is a sufficient statistic for the probability of success in the sense that, conditional on it, the firm’s type is uninformative.) We assume that $p_B < p_G$. In the event that the firm succeeds its value is h (which is fixed and exogenous) so that the profits resulting from the attempt at survival, net of the audit fee f and the loan repayment, are

$$h - f - r.$$

If it fails, then its value is zero and it makes no repayment of the loan⁽⁵⁾. Regardless of the outcome, we assume that the firm and the auditor have no dealings in later periods.

⁽⁵⁾ That is, we are assuming that f is treated as an operating expenses, hence is 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.

In the event that the firm fails, there is an investigation that reveals whether the state was actually G or B . The auditor is forced to leave the industry if she is revealed to have failed to qualify the accounts when the state was bad.

Although, in the equilibria we emphasize, different auditors will have different reputations, all auditors have the same cost of performing an audit, which is assumed to be zero for the sake of notational simplicity. Each auditor is capable of auditing one firm in each period. The interaction between the auditor and the client firm is assumed to be governed by a contract, specifying a payment prior to the audit, and there is also the possibility of a renegotiation leading to a bribe once the actual state has been revealed. Since the auditor is bound to report that the state is good when the actual state is good, there is no possibility of renegotiation in this circumstance.

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.

There is a continuum of auditors with total measure one. Of these, measure $m_R < 1$ are *reputable auditors*. In the equilibria we focus on reputable auditors do not accept bribes because the maximum bribe a firm in a bad state would be willing to pay is insufficient to compensate for the risk to a reputable auditor's reputation. The constant fee charged by reputable auditors is denoted by f_R . All auditors value expected audit fees and bribes in future periods by applying the discount factor $\delta \in (0, 1)$, so the value of a reputable auditor's future stream of fees is $\delta f_R / (1 - \delta)$. The minimum bribe that a reputable auditor would accept is $(1 - p_B)\delta f_R / (1 - \delta)$, which is the probability of being caught times the

loss in that event.

The prior reputation for honesty of reputable auditors is an exogenously given (i.e., historically determined) fact: reputable auditors retain their reputation unless they are caught, and the model does not permit an auditor to acquire a reputation for honesty. Thus the quantity m_R is given and fixed, instead of being determined by equilibration⁽⁶⁾. The essence of reputation is a coordination of intentions and inclinations on the part of the audit firm and expectations on the part of the financial markets. In general one can try to create a reputation, in the sense of inducing desired expectations concerning one's behavior, by taking steps to make honest behavior attractive and dishonest behavior risky or irrational. The implicit hypothesis of our model is that the expectations of financial markets are unresponsive to publicly available information suggesting honest behavior, so that creating "a Big n reputation" from scratch is either infeasible or excessively costly. 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 bribes only because they are insufficient to compensate for the risk to her reputation.

The other auditors are described as *disreputable*. Prior to each period new auditors can enter the disreputable segment of the industry by paying a fixed cost C . The fee actually paid to the disreputable auditor will be f_{DG} or f_{DB} according to the state. Here f_{DG} is the contractual fee, and the difference

$$b := f_{DB} - f_{DG}$$

is the amount of the *bribe*. (In practice such payments 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.) Our definition of equilibrium specifies that the bribe should be the minimum amount that is acceptable to the auditor as compensation for the risk of getting caught⁽⁷⁾.

⁽⁶⁾ One might ask whether there could be 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.

⁽⁷⁾ This assumption can be justified by the following bargaining protocol. Prior to performing the audit, the firm pays the auditor a sum that is necessarily independent of the state. After the audit, if the state is bad, the firm may make a take-it-or-leave-it offer of a bribe to the auditor. Naturally, other bargaining protocols imply different outcomes. It may be of interest to investigate other possibilities along these lines.

The possibility of entry is the mechanism that equates the total supply for auditors with total demand, and it also determines the expected revenue of disreputable auditors. The value of a disreputable auditor's business is equal in equilibrium to the cost C of entry into the industry, and a client firm whose state has been found to be bad is willing to pay a bribe that is sufficient to compensate a disreputable auditor for the risk of being forced to leave the industry. The value of being an undiscredited disreputable auditor is the discounted value of the entry cost C , so the minimum acceptable bribe, plus the value of the possibility of not being detected, must equal the value of refusing the bribe, i.e., $b + p_B \delta C = \delta C$. Thus

$$b = (1 - p_B) \delta C.$$

Since there is free entry, the value of receiving f_{DG} in every subsequent period is C , so

$$f_{DG} = (1 - \delta)C \quad \text{and} \quad f_{DB} = f_{DG} + b = (1 - p_B \delta)C.$$

The type of the auditor is the only information available to the financial market at the time the financial contract is negotiated. The financial market is competitive and risk neutral, so that the firm will receive a loan if there is a repayment amount that does not exceed the value of the firm in the event of success (net of the audit fee and bribe) such that the expected repayment is equal to ℓ . Let r_R and r_D be the repayment amounts for firms who receive unqualified reports from reputable and disreputable auditors respectively. Since reputable auditors always report truthfully, in equilibrium $p_G r_R = \ell$, so that $r_R = \ell/p_G$. We assume that $p_G h > \ell$, so that a firm that is known to be good is a worthwhile gamble⁽⁸⁾.

The repayment amount for a firm receiving an unqualified audit from a disreputable auditor (in equilibria where this happens) is r_D . Equilibrium requires that

$$(\text{Prob}(G|D)p_G + \text{Prob}(B|D)p_B)r_D = \ell, \tag{1}$$

where $\text{Prob}(G|D)$ and $\text{Prob}(B|D)$ are the probabilities of the good and bad state conditional on receiving an unqualified audit from a disreputable auditor. We assume that, regardless of the audit fee, there is no repayment amount satisfying this condition when the state is known to be bad. That is, $p_B h < \ell$, i.e., the expected revenues from continued operation are less than the loan amount.

⁽⁸⁾ Since $(1 - \delta)C$ is the minimum audit fee that could be observed in equilibrium, it would be natural to assume that $p_G(h - (1 - \delta)C) > \ell$, but this strengthening of the assumption would not simplify the analysis by ruling out obnoxious and irrelevant special cases.

For the most part we will focus on equilibria in which firms found to have a bad state are willing to pay the bribe that suffices to elicit a false report from a disreputable auditor, but are not willing to pay the amount required to induce a false report from a reputable auditor. We describe such an outcome as a *two-tier equilibrium*. Another outcome that might be described as an equilibrium is a collapse of the reputation mechanism, resulting in a situation in which formerly reputable auditors accept bribes, so that, in effect, there are only disreputable auditors.

A firm that is audited by a disreputable auditor and is found to be bad must decide whether it is worthwhile to pay the bribe. Since, in our model, all firms who have been found to be in a bad state are, at that point in time, identical, either all firms bribe or none do. The profit of such a firm, in the event that it succeeds, is $h - f_{DB} - r_D$.

Prior to choosing which auditor to hire, the firm is privately informed of its *type*, which is the probability $t \in T := (0, 1)$ that the state will be good. If the firm hires a reputable auditor its expected payoff is

$$tp_G(h - f_R - r_R) = t(p_G(h - f_R) - \ell). \quad (2)$$

Provided that it is worthwhile to bribe a disreputable auditor, the expected payoff resulting from hiring a disreputable auditor is

$$tp_G(h - f_{DG} - r_D) + (1 - t)p_B(h - f_{DB} - r_D). \quad (3)$$

Comparing the results of hiring the two sorts of auditors, we find that the reputable auditors are hired by firms with favorable types.

Lemma 1: *Assume that $h - f_{DB} - r_D > 0$, and suppose there is a probability \tilde{t} such that a firm of type \tilde{t} is indifferent between hiring a reputable auditor and hiring a disreputable auditor. Then firms of type $t > \tilde{t}$ strictly prefer hiring a reputable auditor to hiring a disreputable auditor, and the opposite is the case for firms of type $t < \tilde{t}$.*

Proof: Since $h - f_{DB} - r_D > 0$, the equality

$$\tilde{t}p_G(h - f_R - r_R) = \tilde{t}p_G(h - f_{DG} - r_D) + (1 - \tilde{t})p_B(h - f_{DB} - r_D)$$

implies that $h - f_R - r_R > h - f_{DG} - r_D$. ■

The distribution of firm types (and thus the prior beliefs of the financial markets) are given by a probability distribution μ on $(0, 1)$. We assume that μ is atomless and has full support. These assumptions imply that there is a unique number t^* such that

$\mu((t^*, 1)) = m_R$. Now Lemma 1 implies that, in equilibrium, all types in $(0, t^*)$ either hire disreputable auditors or do not attempt to stay in business, and all types in $(t^*, 1)$ hire reputable auditors. (Since μ is atomless, what a firm of type t^* does is inconsequential.)

We now explain the determination of the repayment terms of clients of disreputable auditors and the fees of reputable auditors. Under the assumptions that (a) all types of firm hire auditors and attempt to stay in business, and (b) the set $(t^*, 1)$ of types hiring reputable auditors is known, when a firm hires a disreputable auditor, the financial market's beliefs about its probability $\tau(t^*)$ of being good will be the average type of firms with types in $(0, t^*)$: set

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

The equilibrium value of r_D is now determined by (1):

$$r_D = \rho_D(\tau(t^*)),$$

where, for $0 \leq \alpha \leq 1$

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

is the equilibrium repayment amount when the probability of a good state is α .

If t^* is small, so that r_D is large, it may happen that $h - f_{DB} - r_D < 0$, in which case even disreputable auditors refuse the largest bribes that the client firm can afford. The outcome seems indeterminate: if the financial market came to believe that all auditors were refusing bribes, the repayment amount r_D would fall, restoring the bribability of disreputable auditors.

An even more extreme possibility is that $h - f_{DG} - r_D < 0$, 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. In any event we will not give further consideration to these possibilities. Henceforth we will assume that

$$h - f_{DB} - r_D > 0. \quad (4)$$

In view of Lemma 1, the fee of reputable auditors will be determined by the condition that a firm of type t^* is indifferent between the two types of auditors. Equating (2) and (3) and isolating f_R yields:

$$f_R = f_{DG} + r_D - r_R - \frac{(1 - t^*)p_B(h - f_{DB} - r_D)}{t^*p_G}. \quad (5)$$

At this point we have shown how to determine the equilibrated variables— f_R , f_{DG} , f_{DB} , r_R , and r_D —in a two-tier equilibrium.

3. Equilibrium Analysis

This section presents the analysis of parameter values for which two-tier equilibria exist. In such an equilibrium a disreputably audited firm must be willing to pay the bribe b , which will be the case if (4) holds. Note that $\tau(t^*)$ is an increasing function of t^* , and $\rho_D(\tau(t^*))$ is a decreasing function of $\tau(t^*)$, so there is a unique number $\underline{t} > 0$ such that

$$\rho_D(\tau(\underline{t})) = h - f_{DG} - b = h - f_{DB} = h - (1 - p_B\delta)C. \quad (6)$$

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

We now examine the incentive compatibility condition under which reputable auditors cannot be bribed. A firm whose state has been found to be bad by a reputable auditor would be willing to pay up to $h - f_R - r_R$ to obtain an unqualified audit report. Thus the maximum bribe plus the probability of success in the bad state times the value of the stream of future audit fees must be less than the value of having the future audit fees with certainty:

$$h - f_R - r_R + p_B \frac{\delta}{1 - \delta} f_R < \frac{\delta}{1 - \delta} f_R,$$

or $f_R \geq f_R^{\min}$ where

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

Comparing this with (5), an equilibrium in which reputable auditors do not accept bribes exists if and only if $t^* \geq \underline{t}$ and

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

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

$$\bar{\rho}(t) := \frac{f_R^{\min} - f_{DG} + r_R - (h - f_{DB})}{1 + \frac{(1-t)p_B}{tp_G}} + h - f_{DB} \quad (7)$$

is the repayment amount for disreputably audited firms that 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 2: $f_R^{\min} - f_{DG} + r_R < h - f_{DB}$, so that $\bar{\rho}$ is a decreasing function.

Proof: Above we assumed the existence of \underline{t} satisfying (6), so that $h - f_{DG} - r_R > h - f_{DG} - \rho_D(\tau(\underline{t})) = b$. Thus $h - r_R - b > f_{DG}$. Since b is just sufficient to compensate an auditor for the risk of losing a stream of revenues of f_{DG} 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 $f_R = h - r_R - b$, this implies that $f_R^{\min} < h - r_R - b = h - r_R + f_{DG} - f_{DB}$. ■

Note that $\bar{\rho}(t^*)$ decreases from $h - f_{DB}$ to $f_R^{\min} - f_{DG} + r_R$ 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

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

so that the image of $\bar{\rho}$ is contained in the interior of the image of ρ_D , which means that we may define $\bar{\tau} : (0, 1) \rightarrow (0, 1)$ by $\bar{\tau}(t^*) := \rho_D^{-1}(\bar{\rho}(t^*))$. Of the four asserted inequalities, the first follows from (4) and $r_D > r_R = \ell/p_G$, while the second follows from $h < \ell/p_B$. The third follows from $\ell/p_G = r_R$ and $f_R^{\min} > f_{DG}$, while the fourth follows from $f_R^{\min} < h - r_R$ and $h < \ell/p_B$. Since ρ_D and $\bar{\rho}$ are both decreasing functions, we have:

Lemma 3: $\bar{\tau}$ 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 $\rho_D(\tau(t^*)) \geq \bar{\rho}(t^*)$, i.e., $\tau(t^*) \leq \bar{\tau}(t^*)$.*

The inequality $\tau(t^*) \leq \bar{\tau}(t^*)$ holds whenever $\bar{\tau}(t^*) \geq t^*$, since $t^* \geq \tau(t^*)$. The case $\bar{\tau}(t^*) < \tau(t^*) \leq t^*$ corresponds to an exotic possibility: the distribution of firm types is so heavily concentrated near t^* , and t^* is so close to one, that the financial markets extend loans to disreputably audited firms at terms that are only slightly worse

than the repayment terms for reputedly 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, and this aspect of the model seems both unrealistic and unproblematic, but 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 whenever m_R is in this interval.

Lemma 4: *There is a unique number $0 < \hat{t} < 1$ such that $\bar{\tau}(\hat{t}) = \hat{t}$.*

Proof: Since $\bar{\tau} : (0, 1) \rightarrow (0, 1)$ is continuous, existence follows from the intermediate value theorem. It is easy to see that a function from $(0, 1)$ to $(0, 1)$ cannot have more than one fixed point if it is either convex or concave, so uniqueness is a consequence of the next result. ■

Lemma 5: *Either $\bar{\tau}$ is concave or $\bar{\tau}$ is convex.*

Proof: The definitions of $\bar{\tau}$ and ρ_D , together with (7), imply that

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

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

$$\bar{\tau}(t)\Delta p = \frac{\ell(p_B + t\Delta p)}{t(\alpha + \beta\Delta p) + \beta p_B} - p_B. \quad (8)$$

Differentiating (8) with respect to t twice, we get

$$\begin{aligned}
\bar{\tau}'(t)\Delta p &= \frac{t(\alpha + \beta\Delta p)\Delta p + \beta p_B\Delta p - (p_B + t\Delta p)(\alpha + \beta\Delta p)}{(t(\alpha + \beta\Delta p) + \beta p_B)^2} \cdot \ell \\
&= \frac{\beta p_B\Delta p - p_B(\alpha + \beta\Delta p)}{(t(\alpha + \beta\Delta p) + \beta p_B)^2} \cdot \ell \\
&= \frac{-\ell\alpha p_B}{(t(\alpha + \beta\Delta p) + \beta p_B)^2}
\end{aligned} \tag{9}$$

and

$$\begin{aligned}
\bar{\tau}''(t)\Delta p &= \frac{2\ell\alpha p_B(t(\alpha + \beta\Delta p) + \beta p_B)(\alpha + \beta\Delta p)}{(t(\alpha + \beta\Delta p) + \beta p_B)^4} \\
&= \frac{2\ell\alpha p_B(\alpha + \beta\Delta p)}{(t(\alpha + \beta\Delta p) + \beta p_B)^3}.
\end{aligned} \tag{10}$$

Note from (8) that, for all $t \in (0, 1)$, $t(\alpha + \beta\Delta p) + \beta p_B > 0$, since otherwise $\bar{\tau}(t) < 0$ contrary to the definition of $\bar{\tau}$. Hence, for all t , $\bar{\tau}''(t)$ has the same sign as the constant $\alpha(\alpha + \beta\Delta p)$. ■

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: *If $\mu((\bar{t}, 1)) \leq m_R \leq \mu((\underline{t}, 1))$, i.e., $\underline{t} \leq t^* \leq \bar{t}$, then there is a two-tier equilibrium.*

Proof: We wish to show that $\underline{t} \leq t^* \leq \bar{t}$ implies that $\tau(t^*) \leq \bar{\tau}(t^*)$. For $0 \leq t^* \leq \hat{t}$ we have $\tau(t^*) \leq t^* \leq \bar{\tau}(t^*)$. For $\hat{t} \leq t^* \leq \bar{t}$ we have $\tau(t^*) \leq \hat{t} = \bar{\tau}(\hat{t}) \leq \bar{\tau}(t^*)$. ■

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

Lemma 6: $\underline{t} < \bar{t}$.

Proof: If $\tau(1) \leq \hat{t}$, 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 $\tau(\underline{t}) < \hat{t}$. From Lemma 2 we have $f_R^{\min} - f_{DG} + r_R < h - f_{DB}$, so (7) implies that $\bar{\rho}(\tau(\underline{t})) < h - f_{DB}$, whence $\rho_D^{-1}(\bar{\rho}(\tau(\underline{t}))) > \rho_D^{-1}(h - f_{DB})$ because ρ_D^{-1} is decreasing. The definition of $\bar{\tau}$ gives $\bar{\tau}(\tau(\underline{t})) = \rho_D^{-1}(\bar{\rho}(\tau(\underline{t})))$, and (6) implies that $\rho_D^{-1}(h - f_{DB}) = \rho_D^{-1}(\rho_D(\tau(\underline{t}))) = \tau(\underline{t})$. Therefore $\bar{\tau}(\tau(\underline{t})) > \tau(\underline{t})$, and consequently $\tau(\underline{t}) < \hat{t}$ since \hat{t} is the unique fixed point of $\bar{\tau}$. ■

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 (7), $\bar{\rho}(0) = h - f_{DB}$. From (6), therefore, $\tau(\underline{t}) = \rho_D^{-1}(h - f_{DB}) = \rho_D^{-1}(\bar{\rho}(0))$, and $\bar{\tau}(0) = \rho_D^{-1}(\bar{\rho}(0)) = \tau(\underline{t})$ follows from the definition of $\bar{\tau}$. 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 $\bar{\tau}$, perhaps more than once, as shown, or it may happen that $\tau(t) < \bar{\tau}(t)$ for all $0 \leq t \leq 1$. If the graph of $\bar{\tau}$ moved up when δ increased, increasing δ would unambiguously expand the set of values of t^* 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 $\bar{\tau}$ 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 7: \underline{t} is a decreasing function of δ .

Proof: From (6) 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. ■

4. Related Theoretical Work

This section compares our model to two other models of how the Big n auditors might be different, in equilibrium, from other auditors. The first paper, due to DeAngelo (1981), proposes that large auditors have a larger client base, hence more to lose in the event of a scandal, due to the phenomenon known as “lowballing.” The second paper, Dye (1993), proposes what has come to be known as “the deep pockets hypothesis.” The main idea is that large firms have more wealth, hence are at greater risk in the event of law suits by creditors or investors.

Another possibility that is sometimes mentioned, but is not considered in our model or by DeAngelo or Dye, is that Big n auditors may simply have superior technology and more talented employees, so that (independent of incentives arising out of moral hazard considerations) their reports are more reliable because they are, in effect, more precise

instruments for examining client firms. While it may be true that Big n firms are more capable, the apparent difficulty of entering the Big n , and the continued existence of less reputable auditors, seem difficult to explain in terms of these factors alone.

DeAngelo was responding to industry critics who argued that large audit firms benefitted unfairly and unjustifiably from discrimination by underwriters and other actors in financial markets. Those critics maintained that large and small audit firms provided essentially the same service, and that the preference of customers for large audit firms was the result of discriminatory practices unrelated to any underlying difference in quality. She countered by arguing that large firms had more to lose in the event that they were found to have failed to honestly report the condition of one of their clients, and that this made their reports more reliable. Thus large and small audit firms were selling different products which could reasonably be expected to be regarded as distinct by customers and those who might be influenced by auditors' reports.

DeAngelo describes a world in which competition in the market for new clients drives expected profits, at the time of first audit, to zero. Since the costs of performing an audit for a new client are greater than the costs of auditing a client one has audited before, audit firms can extract quasi-rents from existing clients who would incur costs if they switched auditors. Since these quasi-rents are anticipated, competition leads auditors to offer a fee for the initial audit that is below cost. This practice, which is known as *lowballing*, implies that relationships with existing clients are valuable to the audit firm, and the value of these relationships is placed at risk whenever the audit firm behaves dishonestly. Large firms have more clients, hence more to lose.

The following argument suggests that one should not expect to observe both large and small audit firms 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. In equilibrium the firms that choose disreputable auditors must be precisely those who prefer to do so, given the signal to financial markets that results from this choice. The "best" type in the pool choosing a disreputable auditor would prefer to be audited by a reputable auditor, if the fees were the same, since the alternative is to be treated as an average member of the pool. A process of continued expansion of the high quality tier, and the best firms in the low quality tier switching, 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) who consider a setting in which, prior to the formulation of a contract between two agents (here the client firm and the financial market) one of the agents has the opportunity to disclose private information. They argue that if credible full revelation is not more expensive

than nondisclosure, then it will occur. More specifically, there cannot be an equilibrium in which some types choose not to disclose, since the most favorable nondisclosing type would prefer disclosure to the beliefs the other agent has about nondisclosing types. Their result has the implication that, across a wide range of models, the coexistence of reputable and disreputable auditors implies a fee differential, which in turn implies either that the costs of high quality auditing are higher or that there are barriers to entry.

Dye (1993) develops 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⁽⁹⁾, 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 market, unless small auditors enjoyed a cost advantage. 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. 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.

5. Empirical Literature

In this section we give a brief guide to relevant empirical literature. We organize our discussion around four issues: a) Is it actually the case that the audit reports of large and/or reputable audit firms lead to better treatment of client firms on financial markets? b) Do large firms charge higher fees? c) What is the relative importance of reputation in comparison with deep pockets? d) Have the answers to these questions changed over time,

⁽⁹⁾ In fact Arthur Andersen was a partnership.

and has the market share of the Big n been expanding?

A. *Does Auditor Reputation Benefit Clients?*

Balvers et. al. (1988) study the impact of auditor reputation on underpricing of IPO's. In their model the audit firms' reputations are common knowledge, but the reputations of investment bankers are less well known. In equilibrium high reputation auditors match with high reputation bankers, which acts as a signal to the markets conveying information concerning bankers' quality. Using a data set consisting of 1182 IPOs between 1981 and 1985, they find support for the following propositions: a) high reputation auditors are matched with high quality investment bankers; b) there is less underpricing of IPOs associated with high quality audit firms; c) there is less underpricing of IPOs associated with high quality investment banks.

Using data on 103 IPOs in New Zealand between 1983 and 1986, Firth and Smith (1992) find that IPO firms that are less well positioned to give credibility to their accounts (e.g., firms whose management has low level of share ownership in the company, firms with high debt to equity ratio, and firms with short history) tends to hire Big 8 audit firms. In a separate regression with underpricing of IPO as the dependent variable, they find that reputation of the underwriter is a significant explanatory variable, but that whether the auditor is a Big n firm is not significant.

Beatty (1989) raises doubts about the classification of Big 8 as having reputation capital and non Big 8 as entirely without reputation, because "it appears that smaller Big 8 and larger non-Big 8 exhibit similar reputations in the IPO market." Using two-stage least squares, he first measures reputation as excess compensation for the auditor, then tests the relation between reputation capital, measured in this way, and underpricing of IPOs. He finds that auditors' reputation is inversely related to underpricing of IPOs.

Beatty's finding may also be construed as supporting the view that high reputation auditors charge higher fees, but since fees are used to measure reputation, the more cautious interpretation would seem to be that, for some reason, there is an inverse relationship between the fees charged by auditors and underpricing. Reputation is one possible explanation, but Beatty's finding is potentially consistent with other theories of auditor "quality."

B. *Is There a Big n Premium?*

The literature on the existence of a Big n premium is large, with few conclusions that are robust across all studies. Simunic (1980) and Chaney et. al. (2002) are studies that fail to find a Big n premium. Studies that find a positive Big n premium include Francis (1984), Francis and Stokes (1986), Palmrose (1986), Francis and Simon (1987),

Simon and Francis (1988), Ettredge and Greenberg (1990), Chan et. al. (1993), Pong and Whittington (1994), Craswell et. al. (1995), and Ireland and Lennox (2002)

In some cases, such as Francis and Stokes (1986) and Palmrose (1986), there is a premium for smaller client firms, but not for large clients. On the other hand, Rubin (1988) finds a Big 8 premium for large clients (which are city governments in his study) but not for small clients.

Any measurement of a Big n premium depends on some measure of auditor costs, and potentially other factors that might affect either the choice of auditor or the relationship between auditor and client. While the studies that we are aware of do not present structural models, they attempt to various extents, and in various ways, to incorporate additional explanatory variables, and to correct for phenomena that could bias regression analysis. In particular, Ireland and Lennox (2002) and Chaney et. al. (2002) are explicitly concerned with biases resulting from the possibility that, for whatever reason, the client firms choosing large firms might be “easier” to audit, in which case the actual premium over costs enjoyed by Big n firms might be larger than it appeared. Ireland and Lennox (2002) find a Big n premium, and argue that correction for selection bias results in a larger estimate of the premium, but Chaney et. al. (2002) find no premium. While the accumulated evidence does not seem to resolve the question definitively, it seems that most of the studies to date find that there is a Big n premium. In addition, there is the possibility that some authors’ failure to find a premium is due to biased statistical methods, but with respect to this point as well, the literature has not yet reached a clear consensus.

C. Deep Pockets vs. Reputation

Although we have not done so, it seems reasonable to suppose that one could construct a theoretical model incorporating both reputation and the deep pockets hypothesis. Similarly, it seems reasonable to suppose that both effects are present in reality. In addressing the empirical literature, probably the correct attitude is to try to assess the relative importance of the two effects.

The empirical literature is not especially large, nor does it point clearly in a particular direction.

Using litigation data during 1960-85 for the largest 15 audit firms in the US (who audit more than 90% of US public companies) Palmrose (1988) found that the “Pretty Big 9-15,” as a group, had higher litigation occurrence rates than the Big 8. She interprets this as evidence that the Big 8 provide, in some sense, higher quality service, but she does not take a position concerning the nature or source of this quality differential. Using multivariate regression (earlier studies used univariate regression) Stice (1991) studies the

effects of 9 characteristics of the client and auditor on the likelihood of lawsuits against auditors. The variables that were found to be significant concern the client firm: the client’s asset structure (higher ratio of accounts receivable and inventory to total assets) financial condition, variability of its returns, and market value. In contrast to Palmrose (1988), he did not find that a Big 8 dummy variable was significant.

Firth (1990) examines the effects of auditor criticisms reported by the UK Department of Trade after investigations, finding that i) there is a small but significant reduction in stock prices of other clients of the criticized auditor immediately after the criticism, and ii) relative to comparable uncriticized auditors, the criticized auditors experienced loss of clients and a reduction in new clients’ fees, but no reduction of existing clients’ fees. As Firth points out, these findings are consistent with the reputation hypothesis⁽¹⁰⁾. Similarly, Chaney and Philipich (2002) find significant negative returns of Arthur Andersen clients during the three days following the disclosure that documents had been shredded, and of course the damage to Arthur Andersen went far beyond lost clients and lower fees. Insofar as criticism does not directly result in diminished auditor wealth, it seems difficult to explain these phenomena using only the deep pockets hypothesis.

In contrast to Palmrose (1988), Lennox (1999) finds that large auditors are more likely to be sued and/or criticised. In contrast to Firth (1990) and Chaney and Philipich (2002), he finds that criticised auditors did not suffer drops in demand as a result of criticism. This paper presents complicated issues concerning the data—reports in the UK financial press—which may be biased in the direction of emphasizing large events. One point of interest is that, whereas his discussion treats reputation as a continuous variable, so that increased criticisms should lead to eroded reputation and demand, in our model it is a binary variable. That is, our model is consistent with “run-of-the-mill” criticism having slight or no effect on demand, while “decisive” events result in occasional catastrophic losses of reputation, as in the case of Arthur Andersen.

D. Evidence of a Changing Environment

One of our important theoretical findings is that incentives to maintain reputation will diminish if the Big n premium is eroded as a result of competition. While entry into the Big n seems to be very difficult, all these firms are now very different creatures from the firms that opened branches in multiple cities during the first quarter of the twentieth

⁽¹⁰⁾ Strictly speaking, these findings relate to a different model of reputation than the one considered herein, in which reputation is nothing more than the belief that the audit firm is motivated to behave in a way that will tend to sustain its reputation, and does not represent a belief about an unobserved persistent feature of the auditor. Since all agents in our model have rational expectations concerning auditor behavior, a loss of reputation by an auditor would not be interpreted as revealing a hidden state variable, and consequently it would not result in new doubts about that auditor’s other clients.

century. (See Edwards (1960) for a description of the evolution of the auditing industry during this period.) In particular, there is the possibility that internal expansion of these firms could erode the Big n premium, and for this reason it is interesting to examine the evolution of their market share over time. Most of the studies cited above deal with a particular time period, and comparison across studies is complicated by differences in the nation from which the data is taken, and the criteria used to select the sample of audit clients.

We know of two studies that are explicitly focused on the evolution of market share in the US. Using a sample of 299 S&P firms drawn from a selection of industries chosen to include both regulated and unregulated industries, and industries with both high and low propensity to use Big n auditors, Danos and Eichenseher (1986) find that the Big 8 market share increased from 51.8% in 1965 to 58.2% in 1972 and then to 66.6% in 1980. For all audit clients listed in the NYSE, NASDAQ, or AMEX (excluding those that were acquired, went private, or for which there was inadequate information) Wolk, Michelson and Wootton (2001) find that the Big n market share was well over 80% in the years 1988, 1991, 1996, and 1999.

6. 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. Lowering the the quality of the marginal firm also results in a lower average quality, and less favorable financing, for disreputably audited firms. We find that the former effect is dominant, except in unusual and unrealistic circumstances, so that expansion of the supply of high quality auditing leads to a reduction of the fee differential. If the supply of high quality auditing passes a certain threshold, the fee differential will be so small that the value of reputation is insufficient to motivate honest behavior.

Our model shares various features with the models of DeAngelo (1981) and Dye (1993). Unlike those models, our model has equilibria in which reputable and disreputable auditors coexist in spite of the disreputable auditors having equal or greater costs.

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. 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 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.
- 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.
- 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., and K. L. Philipich (2002). "Shredded reputation: the cost of audit failure," *Journal of Accounting Research* **40**, 1221–1245.
- Chaney, P. K., Jeter, D. and Shivakumar, L. (2002). "Audit Pricing in Private Firms," mimeo, Vanderbilt University and London Business School.
- 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, East Lansing 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. (1990). “Auditor reputation: The impact of critical reports issued by government inspectors,” *RAND Journal of Economics* **21**, 374–388.
- 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 Judgment’: 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.
- Lennox, C. (1999). “The Relationship Between Auditor Accuracy and Auditor Size: An Evaluation of Reputation and Deep Pockets Arguments,” *Journal of Business, Finance and Accounting* **26**, 779–806.
- 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. (1982). “Reputation and Imperfect Information,” *Journal of Economic Theory* **27**, 253–279.
- 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.
- 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.
- Rubin, M. (1988). "Municipal Audit Fee Determinants," *The Accounting Review*, **63**, 219-236.
- 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," *Quart. J. Econ.* **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

