SMU ECONOMICS & STATISTICS WORKING PAPER SERIES

SINGAPORE MANAGEMENT UNIVERSITY

Honesty and Intermediation: Corporate Cheating, Auditor Involvement and the Implications for Development

Brishti Guha
September 2005

Paper No. 18-2005

HONESTY AND INTERMEDIATION: CORPORATE CHEATING, AUDITOR INVOLVEMENT AND THE IMPLICATIONS FOR DEVELOPMENT[1]

Brishti Guha

Singapore Management University

Abstract

We examine self-enforcing honesty in firm-investor relations in an imperfect public information game. Minimum firm size requirements and moral hazard limit ability to raise outside capital, yielding a floor on personal wealth required to enter entrepreneurship. Credible auditing could create efficiency gains. We propose mandatory disclosure of audit fees and an interpretation of international differences in shareholding patterns. We endogenize auditor-firm collusion and extortion by auditors. We embed our game-theoretic analysis in a general equilibrium model to generate unique equilibria that trace the impact of the distribution of wealth on the existence of the market and consequences for development.

Keywords : Corporate governance, moral hazard, vicious circles, inequality and development, general equilibrium, repeated games.

JEL Codes: D82, G3, O1

HONESTY AND INTERMEDIATION: CORPORATE CHEATING, AUDITOR

INVOLVEMENT AND THE IMPLICATIONS FOR DEVELOPMENT

1. Introduction

Recent corporate scandals have often featured firms cheating their shareholders, with or without auditor involvement. This environment raises issues of trust in firm-investor relations and of "self-enforcing honesty".   In particular, low shareholder trust obviously has adverse economy-wide repercussions. Once one expects to be dealt with dishonestly, this distrust discourages investors, leading to a collapse of industry and triggering off a recession. Honesty and trust, on the other hand, feed off each other. Focusing on the moral hazards faced by firm insiders, we ask when firm insiders will behave honestly and characterize possible equilibria in a world with stochastic market outcomes and imperfect information.   We also investigate when efficiency can be improved – through market creation or increased output and welfare - by employing informed intermediaries (like auditors) and when the presence of such intermediaries could lead to honest equilibria (taking into account the auditor's possible temptations to collude with clients planning fraud or to extort blackmail from honest ones).   We also discuss some policy applications of our model and predictions relating to disclosure of audit contracts, public transparency and the Sarbanes-Oxley reforms. We briefly discuss possible extensions of our work including suggestions on international differences in shareholding patterns.

The distinction between entrepreneurs who control firms and ordinary investors who do not, reflects the distribution of wealth in a society. Our exploration of this link

leads us into the realm of development economics. We show that the level and distribution of wealth affect the growth of corporations and indeed the very existence of the share market through an angle hitherto unnoticed in the development literature – that of credibility. This has obvious implications for the processes of industrialization and growth.

While our basic tool is the infinite-horizon one-sided prisoner's dilemma game, we embed this within a general equilibrium framework. This enables us to establish the existence of a unique equilibrium for any given set of parameters. It also makes it possible for us to endogenize the set of firms that participate actively in the game and the set of investors.

2. Assumptions.

We assume that all individuals are immortal and belong to a closed community with a given wealth distribution. Their wealth is inelastic and fixed in supply, there is no saving or depreciation. Further, they cannot borrow. However, they can lend (to the outside world or to government) at a fixed rate of return R. All agents are risk-neutral.

Individuals can become entrepreneurs and set up firms. However, any enterprise requires a minimum investment of $\underline{I}$. Entrepreneurs have no access to any external source of funds other than investors. Problems of collective action and contract enforcement are sufficient to prevent them forming partnerships among themselves. Groups that can solve these problems effectively (such as the extended business families of pre-war Japan, contemporary Korea, India and the Chinese Diaspora) can be viewed as collective entities with a combined wealth that is the relevant factor in this case. Individuals – or groups of

this kind – with personal funds $F < \underline{I}$ must therefore go public in order to engage in any enterprise.

Each enterprise lasts for one period after which investors can retrieve their capital in its original form. A fresh enterprise requires refinancing by investors.

Firms experience good luck and earn a rate of profit of G on total capital with probability p, and bad luck with a rate of profit of B otherwise (G > B). H = pG + (1 – p)B > R is the firm's profit expectation on total capital. G and B are exogenous. One may, for example, assume that output is subject to exogenous shocks while prices are fixed in the world market as in a small open economy.

Investors and firms have an understanding that all shareholders should receive an expected return of D ("dividend") on their capital. The entrepreneur in addition should receive an "autonomous" expected payment of A, not linked to the volume of insider capital. Both D and A are determined endogenously. For the bulk of this paper we shall assume that these expectations are to be achieved by paying outside investors ("shareholders" who contribute a total of S to the firm's capital) an amount DSG/H when luck is good and DSB/H when it is bad. The firm is to retain (A + DF)G/H and (A + DF)B/H under good and bad luck respectively. The outside investor's expected income then works out at

$$pDSG/H + (1 – p)DSB/H = DS[pG/H + (1 – p)B/H] = DS$$

while the insider's income expectations add up similarly to A + DF. These income expectations sum to A + D(F + S) which is equal to the total profits of the enterprise H(F + S), implying A = (H – D)(F + S). In essence, we confine ourselves to a pure equity contract for most of the analysis that follows.

However, we shall later consider other contracts that promise the same return to the investor and endogenize the firm's choice between these alternatives.

We assume that $DB/H < R$, so that investors prefer not to enter the industry if they expect to receive only their bad luck dues. A sufficient condition for this to hold for all non-negative D and positive R is that B should be non-positive.

The parameters F, S, G, B and p are common knowledge. But the actual realization of the firm is not observable by outsiders or legally verifiable. Therefore, the understanding is not a fully enforceable contract. The firm can cheat by distributing DSB/H to its outside investors even when luck is good.

There is a publicly observable signal that detects cheating by investors with an accuracy (probability) of q with a one period lag and there is collective memory of past cheating.

We retain all these assumptions throughout the subsequent analysis. However, we do have two separate models of the firm, each with its own distinctive characteristics. In the first model, we assume that the firm derives its identity from the investment of its own capital F in firm-specific assets (a la Oliver Williamson) that are valueless outside the firm; and that outside investors have a contractually guaranteed priority of claim in the event of liquidation of the firm. The latter is enforceable because the assets of a firm that has shut shop are verifiable and its payouts are common knowledge. In the second model, we withdraw these assumptions.

Of our assumptions, the one that needs further discussion is that of no savings. We assume zero savings so as to focus on the distinctive consequences of cheating without our

results being obscured by the changing dynamics of the accumulation and distribution of wealth. A justification for the assumption is provided in the Appendix.

3. The Model with Firm-specific Assets

Cheating offers the firm an instant gain of

$$G(F + S) - (A + DF)G/H - DSB/H$$

$$= (G - B)DS/H.$$

But it also involves the risk of detection by the public signal. Exposure of a firm as a cheat triggers instant collective withdrawal of all investment. Thus, detection implies closure of the firm and loss of its capital as well as of its future stream of profits from honest operation. Cheating therefore is deterred if

$$(1) \qquad \frac{(G-B)DS}{H} \le q\delta[\frac{\{A+DF\}}{(1-\delta)}].$$

Substituting for A from the relationship $A = (H - D)(F + S)$ and rearranging,

$$(2) \qquad q \ge \frac{(1-\delta)DS(G-B)}{\delta H[(F+S)H - DS]} = q*(D).$$

The firm will have no incentive to cheat if the probability of detection exceeds the ratio of cheating gains to the present value of future profits lost in the event of detection. Alternatively, we can derive a ceiling for D:

$$(3) \qquad D \le \frac{q\delta H^2(F+S)}{q\delta HS + LS(1-\delta)} .$$

Here L = G-B.

Writing s for S/F, (3) translates into

$$(4) \qquad s \le \frac{q\delta H^2}{[D\{q\delta H + L(1-\delta)\} - q\delta H^2]} = s*.$$

(3) or equivalently, (4) represents a credibility constraint. The insider knows that, if s exceeds the limit set by (4), investors will expect him to cheat and will not therefore invest. (3) and (4) are subject to the insider's participation constraint $A \geq 0$ – without which the insider would prefer to invest his capital in other firms rather than go into business himself since his income as an outside investor would exceed his income as an insider. $A \geq 0$ implies

(5) $H \geq D.$

However, as long as $D < H$, the insider's expected gains will rise with outside investment: so profit maximization will drive s up to the limits of credibility, turning the inequality (4) into an equation indicating the insider's demand for outside capital. In the space (s, D), (4) determines a function s*(D) that can be represented by a rectangular hyperbola asymptotic

to the vertical D-axis and to the horizontal $D = \dfrac{q\delta H^2}{q\delta H + L(1-\delta)}$. Intuitively, the moral hazard of the insider (his temptation to cheat) is an increasing function of D (the rate of return expected by outside investors) as well as of s (the ratio of outsider to insider capital). Any increase in D must therefore be offset by a decrease in s if the firm is to be deterred from cheating.

D (or s) also determines the minimum wealth that an entrepreneur needs to enter the field. Since the firm needs to reach a minimum size $\underline{I}$ to be functional, it cannot come into being unless

(6) $\dfrac{\underline{I}\text{-}F}{F} \leq s^*.$

This implies

(7)
$$F \geq \frac{I}{1+s^*} = F^*.$$

F^* is the minimum wealth requirement for entry and depends on the level of s (or D).

Now suppose K is the aggregate wealth of the economy and P(W) the fraction of total wealth owned by those with wealth below W. Then the total demand for outside capital generated by the entrepreneurs who can enter is

(8)
$$X_d = K[1 - P\{\frac{I}{1+s^*(D)}\}]s^*(D)$$

Here, the term in square brackets represents the ratio of entrepreneurial capital to the total wealth of the economy. The RHS, therefore, represents the amount of outside capital that entrepreneurs can apply for without compromising their credibility. As D falls, not only can each firm credibly apply for more outside investment, but also more firms can enter and create additional demand for capital.

The total supply of outside capital is the total wealth of those below the minimum threshold required for entry:

(9)
$$X_s = KP[\frac{I}{1+s^*(D)}].$$

This is subject to the outside investor's participation constraint $D \geq R$ that assures investors of their opportunity

With the supply and demand functions for outsider capital thus defined, three kinds of equilibrium are possible:

First, (Case 1) a regular interior equilibrium with $H > D > R$ (Fig. 2) characterized by

(10)

$$s^*(D) = \frac{P(\dfrac{I}{1+s^*(D)})}{1 - P(\dfrac{I}{1+s^*(D)})}.$$

A striking characteristic of this is that, for given I, the equilibrium value of s* is uniquely determined by the distribution of wealth alone. If we write P as a function of s*, P = Q(s*), equation (10) reduces to

(10a)   s* = Q(s*)/{1 − Q(s*)},

which has a unique solution. In this equilibrium, outside capital is fully employed without being in excess demand. The optimal ratio of outsider to entrepreneurial capital exactly corresponds to the ratio of wealth owned by those below the minimum wealth requirement for entry to that owned by those above. The interior equilibrium occurs for parameter ranges such that

$$s^*(H)[1 - P(\frac{I}{1+s^*(H)})] < P(\frac{I}{1+s^*(H)}),$$

and

$$P(\frac{I}{1+s^*(R)}) < s^*(R)[1 - P(\frac{I}{1+s^*(R)})].$$

Second, (Case 2) an equilibrium in which D = H, the participation constraint of the firm is binding and capitalists are indifferent between setting up their own businesses and investing in others (Fig. 3). This equilibrium occurs for parameters such that

$$s^*(H)[1 - P(\frac{I}{1+s^*(H)})] > P(\frac{I}{1+s^*(H)}).$$

Third, (Case 3) an equilibrium in which D = R, the participation constraint of the investor binds and investors are indifferent between investing in the firm and in their

outside option[1] (Fig. 4). Here there is an excess supply of capital which has taken shelter in its outside option. This equilibrium occurs for parameter ranges such that

$$P(\frac{I}{1+s*(R)}) > s*(R)[1 - P(\frac{I}{1+s*(R)})].$$

There also remains a possibility (Case 4) that no market may exist. If W_{max} is the wealth of wealthiest individual, $P(W_{max}) = 1$. Suppose that $s*(D_{min}) = (\underline{I}/ W_{max}) - 1$. Then for all $D \geq D_{min}$, $X_d = 0$. The demand curve lies in the positive quadrant only for $D < D_{min}$. Now, if $D_{min} < R$, the minimum supply price of outside capital, the demand and supply curves will not intersect. No equilibrium will be possible (Fig. 5). Note that a low W_{max} implies a higher $s*(D_{min})$, and therefore a low D_{min}. Thus if the society is a poor one without any rich individuals, a market is not likely to exist.

To sum up, without auditing, and when insiders' assets are firm-specific, only those with wealth above a certain floor can become entrepreneurs. Moreover, a market may not exist if no one in the society has wealth above a particular floor. However, if a market exists, there can be three different types of equilibria depending on the distribution of wealth. Only one of these is an interior equilibrium (with insiders earning strictly more than outsiders, who in turn have invested all their wealth in the enterprise and earn strictly more than the outside opportunity R). In the second equilibrium, insiders are just indifferent between becoming an entrepreneur and becoming an outside shareholder in another firm : while in the third, outsiders are indifferent between financing the enterprise and the outside opportunity – but some outsider capital is necessarily not invested in the enterprise.

4.The Model with Redeployable Assets

In this case, firms that are exposed as cheats and dissolved due to loss of investor confidence do not lose their capital. They can in fact withdraw their capital without cost and reinvent themselves as outside investors in other firms. Now the other option firms have is simply to get the outside opportunity cost on their funds. They will take whichever option gives them more.

Cheating will be deterred only if

(11)
$$\frac{(G-B)DS}{H} \leq \frac{q\delta\{A+DF-\max\{D,R\}F\}}{(1-\delta)}.$$

Assume first that D > R. This implies

$$A \geq \frac{LDS(1-\delta)}{q\delta H}.$$

Substituting for A in terms of D, one derives the inequality

(12)
$$D \leq \frac{H^2(F+S)q\delta}{(F+S)q\delta H + LS(1-\delta)} = D**.$$

Alternatively, we can express this as a ceiling on s:

(13)
$$s \leq \frac{q\delta H(H-D)}{[DL(1-\delta)\} - q\delta H(H-D)]} = s**(D),$$

or as a lower limit to q:

(14)
$$q \geq \frac{s(1-\delta)LD}{\delta H(H-D)(1+s)} = q**(D).$$

As in the model with firm-specific assets, the profit maximizing firm will expand up to the limit represented by (13), converting this inequality into an equation. s**(D) can then be represented in the positive quadrant of the (s, D) space by a curve that intercepts the

vertical D-axis at D = H and declines monotonically to a horizontal asymptote at

$$D = \frac{q\delta H^2}{q\delta H + L(1-\delta)}.$$

If however this asymptote is less than R, we must redo our calculations from the point at which the s**(D) curve dips below D = R. Assuming D < R, inequality (11) now yields

$$D \leq \frac{q\delta H[(F+S)H - RF]}{q\delta HS + LS(1-\delta)} \leq R$$

$$\text{Or } q < \frac{RLS(1-\delta)}{\delta H(H-R)(F+S)}$$

$$\text{Or } s \leq \frac{q\delta H(H-R)}{[RL(1-\delta)\} - q\delta H(H-R)]} = s**.$$

These limits are independent of D, and therefore constant at the values reached by inserting D = R in (13) and (14).

As in the model with firm-specific assets, all this imposes a minimum wealth requirement for entry and, in interaction with the distribution of wealth in the economy, generates a demand curve for, and a supply curve of, outside capital. The demand curve $X_d(D)$ differs from that in the previous model in that it (1) intercepts the vertical axis at D = H instead of having a horizontal segment at this level, and (2) becomes vertical below D = R (Fig. 6). In consequence of the first property, one of the three kinds of possible equilibria for the previous model disappears in this: there cannot be an equilibrium in which the firm's participation constraint binds – D cannot equal H. The second property does not however imply any further restriction on the varieties of possible equilibria. The case where no equilibrium exists in the previous model may also be replicated in this.

Here in fact if s**(R) implies a minimum entry requirement in excess of W_{max}, no demand curve for outside capital will exist at all.

To sum up, without auditing, and when insiders' assets are redeployable, there is again a certain floor on the wealth needed to become an entrepreneur. Again, a market may not exist if no one in the society has wealth above a particular floor. If a market exists, there can be two different types of equilibria depending on the distribution of wealth. One of these is an interior equilibrium (with insiders earning strictly more than outsiders, who in turn have invested all their wealth in the enterprise and earn strictly more than the outside opportunity R).   In the other, outsiders are indifferent between financing the enterprise and the outside opportunity but some outsider capital is necessarily not invested in industry.

5. Inequality, the Share Market and Development

Our model sheds some light on a major barrier to industrialization in many poor countries. In such countries, even when enough aggregate wealth exists to support large-scale industry, its mobilization and concentration is often impossible if it is spread too widely. If not too many individuals are wealthy enough to set up businesses on their own, others may find it difficult to raise supplementary funds to do the same. The constraints on borrowing are well-known and have been extensively studied. We show that raising capital from the share market runs into a similar credibility constraint, a limit on the ratio of outsider to insider capital set by concerns over cheating. This accounts in part for the relative underdevelopment of share markets in most poor countries[2]. It explains why most firms in such countries are set up by extended family groups (the Zaibatsu of nineteenth and early twentieth century Japan, the Korean chaebol and the family firms of India and of

the Chinese business sphere) within which capital can flow freely because of a degree of trust and reciprocity among the members. It also accounts for the efforts of governments like the Korean to deliberately foster inequality so as to facilitate the accumulation of personal fortunes that could help in building up credible large-scale industries[3].

If, however, a market exists, and particularly if we have an interior equilibrium of the kind depicted in Fig. 2, the role of wealth inequality changes. Here, the equilibrium ratio of outsider to insider capital s* is, for any given I, uniquely determined by the distribution of wealth; and it can be shown that the more egalitarian the distribution of a given aggregate wealth (in the sense of a higher P( ) for any W), the higher must s* be in equilibrium: the demand curve for outside capital $X_d(D)$ will lie further to the left, the supply curve $X_s(D)$ further to the right, so that D will be lower – not surprisingly, given the negative relationship between D and s* and the higher equilibrium value of s*. Indeed, as the distribution of wealth becomes more equitable, the equilibrium level of D may fall to R (as in Fig. 4); increase in equity beyond this point leads to the disappearance of the market.

This view of equity as an inhibitor of industrialization contrasts starkly with the received wisdom articulated, for example, by Murphy, Shleifer and Vishny [1989]. They see equity as laying the foundations of industrialization by creating a large homogeneous mass market for manufactures. This, however, is essentially a demand effect. It affects production only in a closed economy, where the consumption and production patterns must coincide. In a small open economy, the two are independent.

A wealthier economy with the same degree of inequality – i. e. with higher aggregate wealth but the same Lorenz curve – would have the opposite characteristics.

The fraction of total wealth P(W) owned by people with less than a given level W of personal wealth will be smaller; given $\underline{I}$, the minimum investment requirement for each firm, outsider capital requirements will be lower in equilibrium, thus permitting a higher expected income D for outside investors. It becomes easier to sustain a credible capital market – an addition to the long list of factors that tend to make industrialization a cumulative process and create vicious circles that retard the development of poorer economies.

Finally, an increase in minimum firm-size requirements $\underline{I}$ with the same level and distribution of wealth will increase P for any given s*, thereby depressing the demand curve for outside capital and driving up the supply curve, reducing D and increasing s* in equilibrium. Thus, technological indivisibilities make it more likely that markets would collapse.

6. The Role of the Auditor

The induction of an auditor who can detect and expose cheating by firms changes the picture. It opens the door to Pareto improvements in equilibrium. It accomplishes this essentially by relaxing the credibility constraint. s*(D) or s**(D) increase for any level of D. Firms can mobilize more outside capital for a given rate of payout. Also, the minimum wealth level required for entry falls, so that more entrepreneurs enter. The demand curve for outside capital shifts upward; so does the supply curve (since more entrepreneurs means fewer outside investors) – though their intercepts on the D-axis do not shift. We prove this below and then discuss the consequences for the different kinds of equilibrium specified above.

Let V denote the fee to be paid each period to a firm-hired auditor. Auditors are paid at the beginning of each period. If the investors receive low returns, the auditor investigates and then delivers an audit report revealing whether the firm cheated or has just been unlucky. At this stage, the firm and the auditor can bargain with each other, the firm offering a bribe in exchange for a favorable report, the auditor demanding extra payment for such a report. We discuss below the feasible strategies open to each of the actors in this scenario.

First, *firms* may or may not engage auditors at time t = 0, paying an agreed fee. In either event, they may or may not cheat. If they have hired an auditor and cheated, they may or may not choose to bribe him by offering an extra payment coincidentally with the delivery of a favourable audit report. They could make the offer right at the outset of the period (say at the time of hiring) or later after the cheating and perhaps the investigation has occurred. If they have hired an auditor and not cheated, and the auditor demands an extra payment for certifying to the fact, they may pay or not pay.

Second, the *auditor* accepts a fee at t = 0 and investigates whether cheating has occurred. If it has, he could truthfully report the fact or suppress it for a bribe. If it has not, he could truthfully report this without additional demands or demand a payment for such a report. He also however has the option of negotiating at the outset of the period with the firm, offering it a clean report card in exchange for a bribe, both to be delivered at the end of the period.

Third, *investors* observe at t = 0 whether firms have hired auditors or not. Depending on the information regime assumed, they may or may not get to know the auditor's fees. They then decide whether to invest in a firm or not. At t = 1, they may or

may not reinvest. The information available to them at this point includes the return they have received in the last period, the ratio of insider to outsider capital, the auditor's report and the public signal.

We now show that with auditing, and when insider assets are redeployable, there exists an equilibrium where (a) all firms become clients of the auditor and act honestly, (b) the auditor does not collude with firms planning to cheat, nor does he extort by threatening to blacklist honest firms, (c) investors know this and finance the industry, (d) the investors' off-equilibrium strategy is to shun any firm which is not the auditor's client, to stop financing any firm the auditor labels a cheat, and to mistrust the auditor if and only if he is revealed by the public signal to be colluding with or blackmailing a client. Moreover, if audit fees lie in a certain range, and this is disclosed to investors, the honest equilibrium is guaranteed.

For given D, credible auditing (1) increases the limit on s below which investors will be able to finance the industry without fear of being cheated and (2) lowers the floor on wealth required to become an entrepreneur − even if we account for the auditor's propensities to collude and extort if he has more than one client.

Proof: Hiring an auditor and staying honest thereafter is more attractive for firms than not becoming a client if and only if

(15) $$A + \{D - \max (D, R)\}F - V > 0$$

or substituting for A, and using D≥R in equilibrium,

(16) $$(H-D)(F + S) - V > 0.$$

To ensure that the firms have no incentive to cheat after becoming the auditor's client (and risking certain exposure by the auditor), we require:

(17)
$$\frac{[(H-D)(F+S)-V]\delta}{1-\delta} > \frac{DSL}{H}.$$

If V were 0, the LHS would represent the present value of future honesty which we denote (for simplicity of notation) by P_h, so that (17) now reads

$$P_h - \frac{\delta V}{1-\delta} > \frac{DSL}{H}.$$

This caps the fees that the auditor can get for an honest equilibrium to be sustainable with auditing. The auditor's fees however must not only be positive: they must also suffice to deter collusion or extortion by him if he is to be credible.

Collusion makes firm and auditor vulnerable to exposure by the public signal. The firm takes this into account: the maximum bribe it is willing to pay the auditor is its cheating gain less expected loss due to possible exposure plus expected saving in that event of future audit fees. The auditor compares this bribe with his possible loss of future fees from all his N clients[4] who will dismiss him, once exposure undermines shareholder confidence in him. In the circumstances, the auditor may prefer to try to extract bribes, not from one, but from all his clients by offering to collude with all[5]. Collusion offers must be made in advance of the action (so that firms can cheat if they so desire) – but implemented only at the end of the period by simultaneous exchange of bribes for good reports.

Suppose then that the auditor can make a secret offer to all his clients: 'If all of you bribe me, I will certify that none of you have cheated and you can then cheat with impunity'—the offer being made before the firms announced their payouts and the payment of each bribe synchronized with the delivery of a favorable report. If this were possible, the auditor might be able to extract bribes from all his clients (rather than from

just one) and would stand to lose fees from all of them if caught. Each client of course decides independently on the auditor's offer.

A crucial question here is what happens to q, the probability of collusion being detected, as the number of colluding firms increases. Detection in a single case destroys the credibility of the auditor and his relationship with all his clients. Suppose the auditor's probability of being caught is not perfectly correlated across firms: q for the auditor is an increasing function of N, q(N), where N is the number of clients who decide to collude with him. We defer for the moment the question of how N is determined. An extreme example is a scenario in which the probability of not being caught while colluding with any one firm is independent of the probability of not being caught while colluding with any other. In this event, we would have

$$q(N) = 1 - \{1 - q(1)\}^N.$$

For the individual firm, however, the maximum bribe each firm will offer is equal to gains from cheating, minus expected loss if caught (the discounted value of future payoffs the entrepreneur could have got as an insider, less the fees he would have had to pay the auditor if both were still in business): the relevant level of public signal accuracy here is q(1), the signal that guides investors. Thus the no collusion constraint is:

$$\frac{q(N)\delta V}{1-\delta} > \frac{DSL}{H} - q(1)P_h + \frac{q(1)\delta V}{1-\delta}$$

or

(18) $$\frac{[q(N) - q(1)]\delta V}{1-\delta} > \frac{DSL}{H} - q(1)P_h.$$

In combination with the incentive constraint on firms to deter cheating, this gives us the following range of inequalities on auditor fees:

$$(19) \qquad P_h - \frac{DSL}{H} > \frac{\delta V}{1-\delta} > \frac{\dfrac{DSL}{H} - q(1)P_h}{q(N) - q(1)} .$$

This range is non-empty if and only if

$$(20) \qquad q(N)P_h > [q(N) + 1 - q(1)]\frac{DSL}{H} .$$

If q(N) = q(1), this last inequality implies that firms would have no incentive to cheat even without an auditor. If, q(N) > q(1), firms might cheat in the absence of an auditor, but not under the eyes of one who receives a fee in the appropriate range.

The implied upper limit on the ratio of outsider to insider capital is

$$(21) \qquad s^m = \frac{q(N)\delta H(H-D)}{DL(1-\delta)[1 + q(N) - q(1)] - q(N)\delta H(H-D)} .$$

Now

$$s^m > s** = \frac{q(1)\delta H(H-D)}{DL(1-\delta) - q(1)\delta H(H-D)}$$

if and only if

$$q(N)\,[\,DL(1-\delta) - q(1)\delta H(H-D)\,] > q(1)\,[\,DL(1-\delta)[1 + q(N) - q(1)] - q(N)\delta H(H-D)\,]$$

or if and only if

$$(22) \qquad (1 - q(1))(q(N) - q(1))(1-\delta)DL > 0$$

Thus the auditor raises the limit on outsider financing compatible with honesty – and lowers the floor on wealth required to become an entrepreneur – provided the public signal is imperfect and his probability of being caught increases with the number of clients he attempts to collude with. Likewise for the model with firm-specific assets.

Not surprisingly, then, auditing has a larger role the noisier the public signal (the smaller is q(1)) – provided collusion can be prevented. Control of collusion, on the other

hand, is facilitated by fees in the appropriate range, by a more accurate public signal and by greater patience on the firm's part: s_m, the credible limit on the ratio of outsider to insider capital, is an increasing function of $q(1)$ and $q(N)$ and of δ.

Turn now to the determination of N. Firms decide independently on the collusion proposal and there is no guarantee of unanimity among them. However, they differ only in size. Moreover, all firms have an incentive to drive s to the common credibility limit determined by the market parameter D. Thus, ultimately, firms differ only in the volume of entrepreneurial capital F and all differences in their behavior should be traceable to differences in F. Now, every expression involving capital in all our behavioral inequalities is linear homogeneous in F and S – therefore in F (since the ratio of S to F is the same for all firms). Accordingly, any proposal that makes auditing fees and bribes proportional to insider capital and is acceptable to one firm will be acceptable to all. N will correspond to the entire clientele of the audit firm.

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

We need not investigate the market structure of the auditing industry. If audit fees are observable, they must lie within the range in (19): outside it, there is no demand for auditing because auditing will not be credible. Competition no doubt affects auditing fees:

free entry into auditing, for example, reduces audit fees to the lower limit of the range –
but not below it: auditors prepared to work for less will not be believed by investors and
will not therefore be employed (a la Shapiro-Stiglitz). However, our results are not tied to
any particular specification of the structure of the auditing industry.

A comparison with the case of insider asset specificity

In the model with firm-specific assets, a similar analysis applies, but now the
condition for the firm not to cheat is:

(23)
$$\frac{q(N)\delta V}{1-\delta} > \frac{DSL}{H} - q(1)P_h' + \frac{q(1)\delta V}{1-\delta}$$

where P_h' is the present value of future honesty in this particular model:

$$P_h' = P_h + \frac{\delta DF}{1-\delta}.$$

The critical range for credible auditing is then

(24)
$$P_h' - \frac{DSL}{H} > \frac{\delta V}{1-\delta} > \frac{\dfrac{DSL}{H} - q(1)P_h'}{q(N) - q(1)}.$$

The upper limit to the ratio of outsider to insider capital becomes

(25)
$$s^{m1} = \frac{q(N)\delta H^2}{DL(1-\delta)[1+q(N)-q(1)] - q(N)\delta H(H-D)}$$

$$> \frac{q(1)\delta H^2}{DL(1-\delta) - q(1)\delta H(H-D)} = s*.$$

Thus again, the credibility constraint is relaxed and the floor on wealth required to
become an entrepreneur is lowered.

Are auditors happier when insiders' assets are firm-specific and are lost when the firm is dissolved – i.e. in model 1 than in model 2? We might think so because the cap on auditor fees for a credible honest equilibrium has increased : however, at the same time the lower limit has fallen. Moreover, given D and s, the critical q*(D) (implied by (2)) is lower than q**(D) in (14). Given D and s, there is a smaller range of q in model 1, for which auditors' services are in demand.[6]

Discussion

In both models above, auditing relaxes the credibility constraint and lowers the floor on wealth required to become an insider. This raises both the supply and demand curves for capital with the following consequences for the different equilibria specified.

First, the change in the interior equilibrium will replicate the same ratio of outsider to entrepreneurial capital s*(determined by the wealth distribution independently of all other factors) but at a higher payoff D for investors. Entry requirements (determined by s*) are invariant, so is the set of firms. Since all available capital was fully invested in the industry and remains so after the change, there is no impact on output, only a redistribution from firms to investors and auditors.

Second, if equilibrium without auditing lay on the flat segment of the demand curve, so will equilibrium with auditing, but with more firms and entrepreneurial investment just compensating for lower outsider investment (so that all capital continues to be employed in the industry). Auditing relaxes entry requirements, but capitalists remain indifferent between entering and not entering. Everyone continues to earn H – so there is neither a redistributive nor an output effect.

Third, if equilibrium occurred earlier on the flat segment of the supply curve, the rise in both curves will absorb some of the excess supply of capital into the industry (at the same payoff R) or all of it (at a higher payoff). In the former case, higher s* will result in lower entry requirements, more firms, and increased mobilization of outside capital without change in its payoff. Outside investors will be no worse than before, while firms benefit – incumbents from a higher s, new entrants from a rise in payoff above the opportunity cost level R. The Pareto improvement is possible because excess capital earlier reduced to its outside option is now in the industry, increasing its output by more than its displaced earnings. In the latter case, there is a shift to a regular interior equilibrium. s* rises, leading to new entry and more outside investment in each firm at a higher payoff D > R. Outside investors and new firms benefit; whether incumbents benefit or not depends on whether higher s compensates for the higher payout D. In any event, industry output increases. So the gains of the gainers will more than offset the losses of the losers, and the changed scenario is optimal according to the Hicks-Kaldor-Scitovsky compensation criteria.

Finally, where no market exists, the change could create one if it raises s*(R) above $(\underline{I}/W_{max}) - 1$, increasing credibility to the point where a firm can offer investors their opportunity cost. The Pareto-improvement is clear; so is its source – the emergence of the market.

From this we can gather the following. It seems that in case 2 – where equilibrium without auditing lies on the flat segment of the demand curve (only possible with firm-specific assets) – auditing has no role as it has neither an output nor a redistributive effect. In case 4 – where Pareto gains are to be achieved – there are clear gains to hiring an

auditor, and this happens, as we discussed before, when no one in the society has very much wealth. In situation 3, too an auditor adds to net social welfare. Situation 3 – that of excess supply of capital – is likely to arise when there is a strong middle class so that the fraction of wealth owned by those with less wealth than F* or F** is substantial. Because a high enough proportion of total wealth is owned by those not qualified to become insiders, an excess supply of capital emerges. Thus another situation when auditing is beneficial is when there is a strong middle class. Here, we do not concern ourselves with how auditing emerges in the first place[7]. Finally, in situation 1, firms do not gain from auditing. Investors, however, do. In this situation, we could perhaps see intermediaries hired by investors instead of firm-hired auditors. This suggests that a firm-hired auditor is important either when every one in the society is strongly wealth constrained, or when there is a prominent middle class.

What policy conclusions can we draw from our analysis? We see that an honest equilibrium with auditing requires audit fees in a range definable in terms of the parameters (high enough to deter collusion, and low enough to deter cheating by the auditor's clients). Now mandatory disclosure of audit fees can *guarantee* an honest equilibrium (provided that the required range is non-empty) because investors can satisfy themselves that auditors' fees indeed lie in this range. Therefore, mandatory disclosure could ensure an honest equilibrium with the welfare gains alluded to above. Such disclosure would, of course, have to be backed by penalties severe enough to eliminate any net expected positive payoff to cheating based on false disclosure (after taking into account the consequences of possible exposure by the public signal).

Auditors dislike any rise in public transparency for two reasons. First, lower q increases the probability that auditors can significantly improve welfare over the public signal and so increases demand for them. Secondly, audit fees that guarantee an honest equilibrium are lower for an accurate public signal (particularly if accuracy increases speedily with the number of clients). The logic is that if the signal is inaccurate, auditors' future fees should be high for the expected loss of such fees (should collusion be exposed) to outweigh the current period bribes the auditor could extract. Auditors therefore would dislike changes such as disclosure of stock options as costs.[8]

In the light of our model, we now briefly discuss some suggested antidotes to corporate fraud such as the reforms partly embodied in the Sarbanes-Oxley Act. One such measure is the separation of audit from non-audit activities. The idea is that this would limit the scope for bribery(otherwise, a generous investment banking mandate, for example, could secure the auditor-cum-investment banker's collusion). Certainly, a legally enforced separation of audit from non-audit activities could increase the credibility of auditing by making covert bribery difficult and facilitating detection. In terms of our model, it would raise q, resulting in a wider range of fees satisfying the conditions for honest auditing.

An *additional* conclusion is that the inaccessibility of outside credit for entrepreneurs may significantly restrict market creation, particularly when credit market imperfection is compounded by a wealth distribution with too few wealthy individuals However, auditing can partially compensate for this market imperfection. Auditors, on the other hand, dislike a too-perfect credit market that enables firms to set up business for lower values of public transparency and thus minimizes the need for auditors.

7. Optimal contracts.

Equity contracts of course are only one of the possible ways of raising capital. Consider an alternative scenario in which the firm has a menu of contracts to choose from. In Model I, the most general form of contract that our liquidity-constrained firm can offer its investors is the promise to deliver min $[D_GS, G(F + S)]$ if it is lucky and min$[D_BS, B(F + S)]$ if unlucky (where $pD_G + (1 - p)D_B = D$). This is a contract that explicitly allows for the possibility of bankruptcy. Bankruptcy is possible if $D_BS > B(F + S)$. On the other hand, if $D_BS \le B(F + S)$ – implying $s \le s_l = \dfrac{B}{D_B - B}$ – bankruptcies are impossible and any claim of bankruptcy will be legally barred.

The possibility of bankruptcy raises the specter of false bankruptcies. The entrepreneur could claim misfortune even when he has been lucky, distribute $B(F + S) < D_BS$ to his investors and decamp with the spoils $(D_G - B)S - BF$. (Bankruptcy implies closure of the firm and loss of its subsequent income: so false bankruptcies are unprofitable if the present value of this loss exceeds its one-time cheating gain – if

$$\frac{\delta[HF + (H - D)S]}{1 - \delta} \ge (D_G - B)S - BF$$

$$\text{or } s \le \frac{\delta H + (1 - \delta)B}{(1 - \delta)(D_G - B) - \delta(H - D)} = s_b.$$

Of course, this implies a meaningful limit on s only if $s_b > 0$, which occurs if and only if

$$(1 - \delta)(D_G - B) > \delta(H - D).$$

If $s \le s_b$ or if $D_BS \le B(F + S)$, there will be no false bankruptcies. However, firms could still cheat by misrepresenting the fortunes of the business and offering

investors D_BS instead of their rightful dues D_GS, risking detection by the public signal and loss of future income. This variety of cheating would be unprofitable if and only if

$$\frac{q\delta[HF + (H-D)S]}{1-\delta} \geq (D_G - D_B)S$$

$$\text{or s} \leq \frac{\delta qH}{(1-\delta)(D_G - D_B) - \delta q(H-D)} = s_d.$$

An optimal contract is one that maximizes s while eliminating false bankruptcies as well as the incentive to dishonestly offer investors D_BS instead of D_GS. Bankruptcy would be impossible if the contract itself provides that, in the event of bad luck, the entire proceeds of the firm should go to the outside investor: $D_BS = B(F + S)$. This is indeed the maximum guarantee against misfortune that a liquidity-constrained firm can credibly promise the outside investor: it maximizes D_B subject to credibility. If max $D_B = B(F + S)/S \leq D \leq D_G$, the implication is that it minimizes the cheating premium $(D_G - D_B)$ for any given D – and thereby maximizes s_d. It follows that this is the optimal contract whenever $B(F + S)/S \leq D$. Given that investors are to receive $B(F + S)$ under bad luck, the contract must provide for a good luck payout that ensures an expected return D:

$$pD_GS + (1 - p)B(F + S) = DS$$

$$\text{or } D_GS = [DS - (1 - p)B(F + S)]/p.$$

With D_G and D_B thus determined, $(D_G - D_B)S$ reduces to $\{DS - B(F + S)\}/p$, yielding

$$s \leq \hat{s} = \frac{\delta pqH + (1-\delta)B}{(1-\delta)(D-B) - \delta pq(H-B)}.$$

Thus, the optimal contract also imposes a ceiling on s that is a decreasing function of D.

The picture is rather different if B(F + S)/S > D. With a contract that offers investors B(F + S) in the event of bad luck, this configuration tempts the firm to cheat the investor in the diametrically opposite way − by offering the smaller good luck payoff when in fact it has been unlucky. However, B(F + S)/S > D is equivalent to $s < \dfrac{B}{D-B}$, s is a variable controlled by the firm and the firm's profits are an increasing function of s (if D < H).[9] Since, $s < \dfrac{B}{D-B}$ also undermines its credibility by creating moral hazard, the firm has every incentive to increase s to the level $s \geq \dfrac{B}{D-B}$, at which point the firm will find it worthwhile to offer the optimal contract described above.

The credibility limit for the pure equity contract is s = s*. That for a pure debt contract is determined by the no-bankruptcy conditions − either $s \leq s_l$ or $s \leq s_b$, which reduce, for this contract (in which $D_B = D_G = D$), to $s \leq \dfrac{B}{D-B}$ and $s \leq$ $\dfrac{\delta H + (1-\delta)B}{(1-\delta)D - B - \delta(H-B)}$ respectively. All these limits are lower than ŝ, the ceiling on the ratio of outsider to insider finance under an optimal contract. In all cases, however, there is a limit to the size of the firm, proportional to the personal wealth of the entrepreneur. All our qualitative results will therefore continue to hold.

8. Extensions

Our model suggests that the ratio of outsider to insider capital influences insiders' temptation to cheat but that auditing mitigates this effect, making a higher ratio consistent with honesty. This has interesting implications for international differences in shareholding patterns. In the US, for example, most firms are widely held: s is higher than

in countries dominated by family-owned firms or those where large blocks of capital are owned by other large companies, or banks – well represented on the board of directors and therefore "insiders". Does this difference reflect differences in the auditing framework? Is firm ownership dispersed where better auditing safeguards small outside shareholders? We find much empirical support for this view. La Porta et al, [1999], find that firms are widely held only in countries which have good measures of shareholder protection. In other countries, one often observes family owned firms. This fits in with family financing and underdeveloped share markets wherever shareholder protection (eg. a better auditing framework or better transparency) is inadequate. La Porta et al [1997] also find that firms tend to go public in the first place only if good measures of shareholder protection are in place. Evidence from Korea (Joh [2003]) indicates that in a large sample of externally audited firms, misappropriation by controlling insiders was severe wherever these insiders had a low ownership stake. This bears out our result that a high ratio of outsider financing intensifies moral hazard.

Of course, a widely held shareholding pattern also implies that when auditing is faulty, there is a great danger of investors being duped. In our model shareholders are aware of the aggregate ratio of outsider to insider capital and can make their investment decision accordingly. However if this information could be kept secret – for example if insiders secretly divest (as in Enron), raising the ratio of outsider to insider financing – outside shareholders become more vulnerable.

9. Related Literature

Our paper is related to many strands of the theoretical and empirical literature. It contributes to a body of work on "order without law" – though there may be a legal

mechanism in the background its role is limited where non-verifiability leads to incomplete contracting. Given this framework we inquire into the possibility of sustaining "order" – here, honest behavior by firms – in a repeated game with imperfect public information. We also examine how and when private intermediation in the form of informed auditing could help. Dixit [2003] is quite closely related to our work. But while Dixit considers random one-period encounters between anonymous buyers and sellers, who may each be intrinsically honest or opportunistic, our relationships are more persistent, terminated only by the detection of cheating. Further, no one is intrinsically honest, but only firms have an incentive to cheat, so that ours is a one-sided prisoner's dilemma. Also, while Dixit's outcomes are determinate, ours are stochastic: so, despite perfect observability of payoffs not just by the cheated investor but by the general public, not even the cheated party can tell for sure if it has really been cheated. In Dixit's model, any one who is cheated knows this when he observes his payoffs – the problem is how to convey the information to the cheat's potential future partners.

Other cheating literature relevant to our analysis includes Greif's theoretical models and case studies of the multilateral punishment of cheats by Maghribi trader-coalitions and the contrast with bilateral punishments. The public signal as well as private intermediation facilitate multilateral punishment of cheating firms by all investors (not just the ones cheated) and multilateral punishment for a deviant auditor.

Another strand of the literature our paper relates to is work on imperfect markets for capital or credit. For example, in Banerjee and Newman [1993], the initial wealth distribution in the population determines occupational choice - only those whose wealth exceeds a certain floor can become entrepreneurs. The underlying causes are imperfect

capital markets (as entrepreneurs may renege on loans) and a minimum size requirement for making the enterprise operational. In our paper we combine ideas of indivisibilities in enterprise size and imperfect capital markets with the moral hazard that entrepreneurs (insiders in our model) face with regard to their outside shareholders.

Our paper also relates to some corporate finance literature. In Gale and Hellwig [1985], firms have the potential of cheating their investors because the state of nature is costlessly observable only by the firm. In that paper, investors (often big entities like banks) *can* verify the state of nature, but verification is costly. The paper recommends a debt contract (which being state-invariant, leaves the firm with less scope for cheating by exploiting its private knowledge of the state) with the proviso that declarations of bankruptcy should be followed by costly state verification by the lender. Beyond one period, however, problems of renegotiation-proofness arise. While we focus on equity contracts, in principle a very similar analysis is applicable to a debt contract as well. In that event, firms could exploit their asymmetric information about the state of nature to declare bankruptcy and give the investors a cash flow consistent with bad luck. However, the same firms could, unless caught by the public signal, again be refinanced by investors who believe that the firm genuinely experienced bad luck. If the public signal is sufficiently imprecise, investors know beforehand that the firms face too strong a moral hazard, and do not finance the industry.

We now come to the empirical literature. Some of this supports (a) our result that a high ratio of outsider to insider financing intensifies moral hazard (Joh, [2003], Lemmon and Lins, [2003]); (b) our prediction that firms would be widely held only in countries with better shareholder protection in the form of external auditing or a transparent public

environment (La Porta et al [1999], Shleifer and Wolfenzon [2002]), and (c) our implication that investor protection (like auditing), by making it possible to credibly raise share-capital, determines whether firms go public or remain family enterprises (La Porta et al [1997]). Studies of the effect of the provision of non audit services on auditing firms' tendency to qualify a report are also relevant (Wines [1994], Barkess and Simnett [1994], Craswell, Stokes and Laughton [2002]).

Appendices

The Zero Savings Assumption

Somewhat extreme assumptions with regard to savings are in fact common in the literature where they serve a variety of purposes. For example, Galor and Zeira, [1993], assume – without any attempt at justification – that the young generation in a two-period overlapping- generations model doesn't consume at all; Bernanke and Gertler, [1989], assume that one class of agents, in another overlapping generations model, consumes only after retirement. A certain degree of licence has been traditionally permitted in the literature with regard to the savings assumption.

This is not to say that no justification can be offered for the no-savings assumption in our model. One possible justification is as follows. In models like ours, risk-neutrality

is a standard postulate. With a constant rate of time-preference, the intertemporal utility function can then be written as

$$U = \sum \delta^t c_t$$

where c_t is consumption in the t-th period. The net gain in utility from a one-period postponement of a unit of t-th period consumption is then

$$\delta^t[-1 + \delta(1 + r_t)]$$

where r_t is the rate of return to capital in the t-th period. With risk-neutrality, savings are no longer needed to perform a consumption-smoothing function. They now reflect only the discrepancy, if any, between the rate of time-preference and the rate of return on capital. When these are independent of the level of consumption, savings will have a bang-bang character. If there are no constraints on the consumption of capital and time preference exceeds the rate of return, all wealth will be dissipated in the first period. On the other hand, if the rate of return is higher, all income will be saved and consumption postponed for ever. Savings will be precisely zero if (1) capital is not consumable (again a standard assumption, see Bernanke and Gertler, [1989]) and (2) time preference is higher than the rate of return.[10] In our model, the highest rate of return is H: a sufficient condition for zero savings therefore is $H < (1 - \delta)/\delta$. It can be readily checked that such a restriction is not inconsistent with any of our results.

One could of course question the origin of what wealth there is. Where did it come from if there are no savings? Here, we must resort to 'manna from heaven' assumptions. All wealth could be land, as in some banana republic where the consumption good is too perishable to be stored. Alternatively, in an industrial economy, wealth could be machinery, that the country has received through foreign aid or as war reparations. Our

essential purpose of course is a focus on the problem of cheating independently of the level or distribution of wealth; and all we need for this purpose is that the zero-savings assumption should be self-consistent, not that it should be realistic.

The Optimality of Multiple Collusions

An auditor who makes an offer of collusion to a single client can extract a maximum bribe amounting to his client's maximum gain from cheating

$$X = \frac{DSL}{H} - q(1)P_h + \frac{q(1)\delta V}{1-\delta}$$

But his expected income is less than X by the expected value of the fees paid by all of his clients that would be lost in the event of exposure - which is $\frac{q(1)N\delta V}{1-\delta}$. Denote this by q(1)Y where Y is the discounted value of audit fees from all N clients. If, however, the auditor were to offer to collude with all his clients, and this offer is accepted by all, his potential income from bribes would be multiplied by N, while the risk of exposure would increase from q(1) to q(N). (We have proved in the text that an offer acceptable to one client is acceptable to all). His income expectation from multiple collusion is

$$NX - q(N)Y$$

His income from single collusion is

$$X-q(1)Y.$$

A *sufficient* (but not necessary) condition for him to prefer multiple collusion is

$$Nq(1) \geq q(N).$$

This condition is sufficient because if it holds, multiple collusion yields more profits for the auditor than N times the profits from single collusion (and N cannot be less than one). An interpretation of the sufficient condition is that the risk of exposure should not increase

more than additively as the number of firms increases. While this seems highly plausible, multiple collusion can be optimal even under weaker conditions.

Brishti Guha

Department of Economics

Princeton University.

References

Banerjee, A. V. and A. F. Newman, "Occupational choice and the process of development", *Journal of Political Economy*, CI (1993), 274-298.

Barkess, L and R. Simnett , "The pricing of other services by auditors : independence and pricing issues" , *Accounting and Business Research*, XXIV (1994), 91-108.

Bernanke, B and M. Gertler, "Agency costs, net worth and Business Fluctuations", *American Economic Review*, LXXIX (1989), 14-31.

Campbell, T. S. and W. A. Kracaw, "Information Production, Market Signaling and the Theory of Financial Intermediation", *Journal of Finance*, XXXV (1980), 863-882.

Craswell, A., D. Stokes and J. Laughton, "Auditor Independence and Fee Dependence" , *Journal of Accounting and Economics*, XXXIII (2002), 253-75.

Diamond, D, "Monitoring and Reputation: the Choice Between Bank Loans and Directly Placed Debt" , *Journal of Political Economy*, XCIX (1991), 689-721.

Diamond, D. and R. Rajan, "Liquidity risk, liquidity creation and financial fragility", *Journal of Political Economy*, CIX (2001), 287-328.

Dixit, A. K., " On Modes of Economic Governance" , *Econometrica*, LXXI (2003), 449-481.

Fuller, K. P, "The impact of informed trading on dividend signaling : a theoretical and empirical examination" , *Journal of Corporate Finance*, IX (2002), 385-407.

Gale, D. and M. Hellwig, "Incentive-Compatible Debt Contracts : the One-Period Problem", *Review of Economic Studies* , LII (1985), 647-663.

Galor, O and J. Zeira, "Income Distribution and Macroeconomics" , *Review of Economic Studies*, LX (1993), 35-52.

Green, E and R Porter, "Noncooperative collusion under imperfect price information", *Econometrica*, LII (1984), 87-100.

Greif, A., "Contract Enforceability and Economic Institutions in Early Trade: the Maghribi Traders' Coalition", *American Economic Review*, LXXXIII (1991), 525-48.

—, "Contracting, Enforcement and Efficiency: Economics Beyond the Law", *Annual World Bank Conference on Development Economics*, (1996), 239-265.

Joh, Sung Wook, "Corporate Governance and Firm Profitability : Evidence from Korea before the Crisis", *Journal of Financial Economics*, LXVIII (2003), 287-322.

John, K. and L. Lang, "Insider Trading around Dividend Announcements: Theory and Evidence", *Journal of Finance*, XLVI (1991), 1361-89.

Kim, O. and R. E. Verrachia, "Pre-announcement and event-period private information", *Journal of Accounting and Economics*, XXIV (1997), 395-419.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer and Robert Vishny, "Legal Determinants of External Finance", *Journal of Finance*, LII (1997), 1131-1150.

La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer, "Corporate Ownership Around the World", *Journal of Finance*, LIV (1999), 471-517.

Lal, Deepak and Hla Myint, *The Political Economy of Poverty, Equity and Growth,* Clarendon Press, Oxford (1996).

Lemmon, Michael and Karl Lins, "Ownership Structure, corporate governance and firm value : evidence from the East Asian financial crisis", *Journal of Finance*, LVIII (2003), 1445-1468.

Lundholm, R,"Public Signals and the Equilibrium Allocation of Private Information" , *Journal of Accounting Research*, XXIX (1991), 322-49.

Murphy, Kevin, Andrei Shleifer and Robert Vishny, "Income Distribution, Market Size and Industrialization", *Quarterly Journal of Economics*, CIV(1989), 537-564.

Ramey, G. and J. Watson, "Contractual Intermediaries", *Journal of Law,Economics and Organization,* XVIII (2002), 362-384.

Shleifer, Andrei and Daniel Wolfenzon, "Investor Protection and Equity Markets", *Journal of Financial Economics*, LXVI (2002), 3-27.

Wines, G, "Auditor independence, audit qualifications and the provision of non-audit services : A Note" , *Accounting and Finance*, XXXIV (1994), 75-86.

Notes

[1] However, investors will take care that their investment in the firm is not so much that it tempts cheating.

[2] Empirical work links the extent to which firms go public, as opposed to operating primarily as family-owned firms, to the extent of shareholder protection available. This becomes relevant to our work as we show later that credible external auditing can create a market and facilitate raising capital in the share market.

[3] Lal and Myint [1996] provide a good discussion of this.

[4] We do not of course assume that there is only one auditor to the industry: N is simply the number of firms hiring a particular auditor, not the total number in the industry

⁵· We show in the appendix that this is generally the optimal course for an auditor who proposes to collude.

⁶· The intuition for this is as follows. Firms stand to lose more in the event of dishonesty in model 1, provided they are caught by the public signal. Therefore, they are honest for a greater range of public signal accuracy**.**

⁷· This could be the subject of another paper.

⁸· This conclusion might be modified if we assumed that auditors can detect cheating only inaccurately. In that case, they might be helped by higher q, with the public signal complementing their efforts. But the effects mentioned above would still be present.

⁹· D > H is incompatible with the existence of the firm since the entrepreneur would then prefer to become an outside investor.

¹⁰· No individual can dissave by trading capital for output, since, if one wishes to dissave, so will everyone else – so that the potential dissaver cannot find anyone to trade with.