Using Voluntary Disclosure Intensity to Infer the Precision of Management's Vision

Miles B. Gietzmann
Faculty of Finance
Cass Business School

Adam J. Ostaszewski
Department of Mathematics
London School of Economics

version January 2009 CDAM-Din.tex

CDAM Research Report LSE-CDAM-2009-2

Abstract

Shin (2006) has argued that in order to understand the equilibrium patterns of corporate disclosure, it is necessary for researchers to work within an asset pricing model framework in which corporate disclosures are endogenously determined. Furthermore, he argues that without such a framework optimal disclosure strategies may seem counter-intuitive. With this in mind, we generalize the Dye (1985) and Penno (1997) upper tailed disclosure models, so that management's strategic disclosure behaviour can be shown to result in an optimal observable disclosure intensity. We show why a higher equilibrium disclosure intensity may need to be interpreted as implying management have less precise forecasts of future firm value (referred to in the title precision of management's vision). The derived results call into question the specification of empirical studies which test whether firms with higher disclosure intensity will face a lower cost of capital. Working within a generalized Dye-Penno framework this research shows why in equilibrium the converse case applies.

1 Introduction

Companies recognize that implementation of a news-disclosure strategy will affect market value. Simultaneously investors infer that observed disclosure patterns are driven by company type: that is, investor responses (in terms of trading behaviour, and therefore stock price) are guided by beliefs as to the company's type. In this respect some theoretical disclosure models are not readily amenable to empirical investigation, because the key parameters upon which equilibrium beliefs are based are not empirically observable. This research offers an equilibrium model of *market* response to news-disclosure in a form readily amenable to empirical research design. Specifically, this research establishes why in equilibrium investors may be assumed to act as if they base beliefs upon the observed disclosure intensity of a company. As the starting point for the theoretical modelling we draw upon Dye's disclosure model and his theorem (see below for details) that in equilibrium, when managers are ex-ante informationally partially-endowed, they will only voluntarily disclose news that has been perfectly revealed to them if it is sufficiently good – above an optimal (equilibrium) cutoff. This is succinctly described as the adoption of an upper-tailed disclosure strategy.

Until now the Dye framework has not been readily amenable to empirical study. One of the difficulties is that an underlying parameter (probability of receiving information which we describe as the information endowment) is a latent variable that may vary between companies. Dye's model posits a distribution of company value dependent in part upon an exogenously given information endowment faced by management. We generalize the setting and specify **endogenously** how management will optimally choose their information endowment. Working within an optimized framework it then becomes possible to show how the optimal disclosure strategy of a company implies an observable disclosure intensity which can be used by outside investigators (econometricians) to infer the underlying parameters of the company which determine equilibrium valuation in the market. Many empirical disclosure models are concerned with the link between cost of capital and disclosure so, if the current research is to have relevance to those studies, it is necessary to show how in this new model setting risk should be priced. With this aim in mind, one important form of risk is incorporated into the Dye model: noisy rather than certain signals of company value – using the idea of Penno (1997) – as detailed below. The model developed here shows how investors can infer managerial signal risk from observed disclosure intensity. We recall

Penno's original contribution: an 'impossibility' result, that for specific assumed functional forms (in particular the distributions of underlying values and noises) the intensity of disclosure is invariant to signal risk (in which case inferences would not be possible). It is shown here that more generally this latter result effectively rests upon the underlying investor risk aversion (describable via a requirement of *first-order stochastic dominance* to be exhibited by the distributions in terms of the underlying parameters). Thus, for a wide class of distributions that includes the traditional log-normality assumption for returns, inferences based upon observed disclosure intensity can be made, because disclosure intensity and signal risk are monotonically related.

Before commencing with the formal model a short discussion of the Easley and O'Hara (2004) theoretical model, which is increasingly being used by disclosure empiricists, is now presented in order to clarify the significant differences in focus as between their model and the one presented here.

The comparison is best achieved by considering the two companion papers: Easley, O'Hara and Paperman (1998) and Easley, Kiefer and O'Hara (2002) which operationalize the theoretical model. These papers develop a multi-day microstructure model of trading between an uninformed, Bayesian, risk-neutral, competitive market maker and two types of trader (informed or uninformed) with unobservable type. On each day information, either no news, good news or bad news, arrives randomly at the start of the day and this is seen only by the traders. The traders can then trade several times in that day. The market maker has a prior distribution over their information set which enables her to update her beliefs given any trade of the day. She can therefore set ex-ante (prior to trade) bid and ask prices to 'immunize' herself (albeit only in expectation) against risk. The opening bid-ask spread (i.e. the bid-ask spread at the *start* of the day) is assumed to be proportional to the probability of a trade being information based. This probability at the start of the day, which has a very appealing formula, has become widely known as the PIN.

In the empirical study the two papers show that there exists a positive dependence of the bid-ask spread on the PIN, which is consistent with the assumptions just outlined. This is taken to be evidence that PIN may be used to explain asset returns. Asset pricing researchers have since increasingly investigated the linkage between microstructure, accounting and asset pricing.

However, one concern with PIN market micro-structure models is that

they are being increasingly used arguably out of context – for example, in corporate governance, where the primary asymmetry is between management and investors.

It is perhaps worth stressing that PIN models are based upon asymmetries of information that exist between traders. As such these models do not look at the traditional asymmetry between the two classes: management and investors and so do not help understand how the enduring existence of this later form of asymmetry drives disclosure practice. For this reason, the current research turns to consider the Dye (1985) model because its focus is upon exactly how this later asymmetry drives disclosure policy.

The paper is organized as follows. In section 2 the generalized Dye-Penno model of voluntary disclosure is derived using the first lower partial moment (fLPM) approach. In section 3 we develop the relationship between the endogenized value of the partial information endowment parameter p and the intensity of disclosure, denoted τ. In Section 4 the derived link between disclosure intensity τ and information risk, measured by the signal noise variance (denoted by σ_Y), is presented. Section 5 studies how this link depends upon the distributional assumptions underlying the model. In section 6 we discuss empirical implementation and compare and contrast our formulation to two recently introduced alternative approaches to empirical implementation. Concluding comments are presented in section 7.

2 The Generalized Dye-Penno model

The Dye (1985) equilibrium model was developed to explain the seeming contradiction between on the one hand the early disclosure unravelling theory of Grossman and Hart (1980) and on the other hand claimed empirical observation of companies choosing not to disclose information. Critically the new modelling assumption introduced by Dye was that managers were only informed about the underlying state of nature (company value) *probabilistically*. This introduced a new tension, not present in earlier models, since on observing non-disclosure, investors needed to apply caution before assuming non-disclosure was driven by bad news. In the Dye setting, the absence of news could also be explained by management simply not having been informed (as a realization of the probability law). The principal result in Dye (1985) was to establish that the optimal management disclosure strategy was an upper tailed strategy under which management disclosed only if observed

news was sufficiently good (above an equilibrium cutoff).

While the Dye model clearly contributed to understanding why non disclosure could happen in equilibrium, the model had a number of restrictive assumptions that reduced its empirical applicability. In particular, when managers were informed, they were perfectly informed as to the value of the company. Thus if the valuation news was above the critical cutoff, managers would disclose this value and the challenge for investors to value the company would no longer exist, since the manager's information was under assumption perfect[1].

Penno (1997) introduced the most obvious generalization of the Dye model. He amended the Dye model by relaxing the assumption of random observation of a true value and instead allowed for management to be informed probabilistically with a noisy signal as to the future valuation (state of nature). Thus in the Penno setting, on observation of a management disclosure, investors needed to form an opinion as to the underlying noise process faced by management before they could rationally process the disclosure. Intuitively, if investors believed that management's *good news* information (that lead to a disclosure) was very noisy they would be less inclined to increase their valuation for the company as compared to the case were management's information was less noisy. Thus in this more realistic setting investors are not assumed to take disclosures at face value; instead they also estimate the precision of management's signal of value. That is, when valuing a company, investors are required not only to estimate the likelihood of non disclosure, but also when there is disclosure, what precision should be assumed for that disclosure[2].

The particular appealing features of the Penno model were that: not only did it introduce this greater realism for the investor valuation problem, but also showed that the optimal disclosure strategy was simply a slightly adjusted form of the Dye upper tailed disclosure strategy and the specific cutoff value of the upper tail strategy shifted down.

The Penno model thus seemed a prime candidate for empirical investigation. However, before attempting to implement an empirical procedure, at issue now was how to model investors' inferences as to the precision of managements information on valuation. This stage is a key modelling step

[1] Following Dye it is assumed that when managers disclose value they always do so truthfully.

[2] This is analogous to the risk return tradeoff faced by portfolio investment strategists.

and admits a range of possibilities. Motivated here by actual disclosure data, we propose that the observed intensity of disclosure can be used by investors to infer the precision of managements information. To proceed along this modelling route next requires one to show how disclosure intensity is imbedded in an equilibrium Dye-Penno model. However, unfortunately the final proposition of Penno (1997) is starkly negative about such meaningful imbedding of disclosure intensity. In particular Penno (page 280) concludes that "contrary to the popular notion that higher quality information is accompanied by more voluntary disclosure, the paper has demonstrated that this notion is, in general, not true." Penno derives this result using a particular assumption about the way the noisy signal received by management combines with the underlying uncertainty of company valuation. A key contribution of this research is to provide researchers with a means by which to assess how restrictive the original Penno modelling was. This is achieved by close consideration of how the predictions of the equilibrium disclosure model change as the general distributional assumptions about the underlying information sets vary.

Following from the above discussion this section is organized as follows. Subsection 2.1 reviews the underlying Dye upper tailed disclosure calculus. It is shown how the disclosure strategy is driven by the properties of what we call the hemi-mean function, which is obtained by varying a parameter in the well-established first lower partial moment used in financial risk management. In subsection 2.2 it is shown how to endogenize management's choice of information-endowment parameter so that when ones refers to the probability that a manager is informed, it is an equilibrium choice rather than an exogenous model assumption. Subsection 2.3 generalizes the Penno model to an arbitrary distributional setting consistent with the preceding subsections. The contribution of this subsection is to show that despite the arbitrary setting investor's inference of company valuation can be interpreted as though a Kalman filtration model applied; this provides a simple robust method for developing intuition on how investors incorporate noisy management disclosure into equilibrium investment valuation decisions. A brief subsection 2.4 connects in a simple fashion the (observable) disclosure intensity with the optimized information-endowment parameter providing a basis for empirical research design. Standard notions of stochastic dominance are recalled in section 3 to identify distributions for which the Penno impossibility fails.

2.1 Valuation under non-disclosure with Dye's disclosure calculus

In the Dye model there is a rational (equilibrium) reason why management may not disclose information voluntarily (a relaxation of the unravelling paradigm). This necessitates a procedure (to be developed below) enabling investors to value the company at other than the minimum (assuming bad news) when they observe non-disclosure. We point out that the equation identifying the disclosure cutoff is in fact a no-arbitrage condition. As there is asymmetry of information, the manager has an alternative valuation of the company under certain circumstances, which depends on the rules of trade applied to managers.

When analyzing information flows the Dye disclosure model assumes three distinctive dates: ex-ante, interim and terminal dates. In the model a random variable X relating to company valuation has density $f(x)$ and associated distribution function $F(x)$ with an ex-ante expected value m_X. Realizations of the random variables are observed by management at the interim date with a probability q. Management's decision whether or not to disclose an observed realization of company value x is a voluntary (strategic) decision. Dye (1985) establishes that under continuity of f there will exist a unique value $x = \gamma$ at which management will be indifferent between disclosure or non-disclosure. Here γ will be called the *Dye cutoff*. The indifference point is characterized by equality between a credibly disclosed value γ and the valuation formed by investors when they see non-disclosure (ND), or more formally from $E[X|ND(\gamma)]$ the computed expected value of the company conditioned by the absence of disclosures below the value γ. That is, the indifference is described by the equation:

$$\gamma = E[X|ND(\gamma)]. \tag{1}$$

A particularly clear intuition for the equilibrium conditions is provided by Jung & Kwon (1988) which we now adapt. When investors value the company ex-ante they need to assign probabilities to the following three events; no information received by management (which occurs with probability $p = 1 - q$), information is received by management but not disclosed (with probability $(1-p)F(\gamma)$), information is received by management and is disclosed (with probability $(1-p)(1-F(\gamma))$). Thus for (1) to hold for γ requires the expected

payoff from non disclosure to equal the payoff from disclosure:

$$\begin{aligned}[p+(1-p)F(\gamma)](\gamma-m_X) &= (1-p)(1-F(\gamma))(\gamma-m_X)\\ \frac{p}{(1-p)}(m_X-\gamma) &= \int_{x\leq\gamma}(m_X-\gamma)dF(x)\end{aligned}$$

using integration by parts gives

$$\begin{aligned}\frac{p}{q}(m_X-\gamma) &= \int_{x\leq\gamma}F(x)dx\\ &= \int_{x\leq\gamma}(\gamma-x)dF(x)\equiv H_X(\gamma) \qquad (2)\end{aligned}$$

where $H_X(\gamma)$ is the lower first partial moment well known in risk management[3]. As this function is central to the Dye calculus in our analysis we explicitly name it the *hemi-mean*. Intuitively one can see why a construct from risk management arises since typically in financial risk management one is concerned with protecting oneself from an expected payoff in the lower tail of a distribution for instance following bad events. Similarly here the ex-ante valuation of the company has to take account of the valuation implications of a manager not making a disclosure (which occurs for all observed values $x<\gamma$). The appeal of this form lies in the separation of the odds ratio p/q which characterizes management information technology on one side and on the other a convex function H_X containing all the information[4] on the distribution of X.

It is important to note that equation (2), when written in the alternative form

$$p(\gamma-m_X)+qH_X(\gamma)=0, \qquad (3)$$

yields

$$(p+qF(\gamma))(\gamma-m_X)+q[H_X(\gamma)+F(\gamma)(m_X-\gamma)]=0, \qquad (4)$$

expressing a non-arbitrage condition relative to the information structure of the model, as we now show. That is, given a risk-neutral distribution F_X of future company value, the company is fairly priced initially at m_X — since with probability $p+qF(\gamma)$ its value adjusts by $\gamma-m_X$ as it falls to γ (in the

[3]See for example McNeil,Frey and Embrechts (2005).

[4]The Characterization Theorem in appendix A provides a precise statement of this.

absence of disclosure) and (in the presence of disclosure) with probability q its value rises on average by:

$$\int_{u\geq\gamma}(u-m_X)dF(u)=\int_{u\leq\gamma}(m_X-u)dF(u)=(m_X-\gamma)F(\gamma)+\int_{u\leq\gamma}F(u)du, \tag{5}$$

appealing to some integration by parts. This no-arbitrage condition[5] provides one of the central distinguishing feature of the Dye model, differentiating it from other disclosure valuation models such as Verrecchia (1990).

The Dye equation has an interesting 'reduced form' interpretation: the term $(m_X-\gamma)$ measures the downgrade in company value, so that $p(m_X-\gamma)$ represents the expected downgrade, conditional on the manager receiving no information. Of course, γ is a risk-shield (since values below γ if seen, remain are unreported), so, given the risk-shield enjoyed by the investors, the term $m_X-\gamma$ is the extent of the downgrade (loss).

Now, by (2),

$$p(m_X-\gamma)=qH_X(\gamma)$$

and so the term on the right can be interpreted as the balancing expected upgrade (at the equilibrium γ), conditional on the manager receiving information. Inspection of the left-hand side of (5) shows the upgrade as an upper partial moment, but from the expected upgrade term $qH_X(\gamma)$ we have identified a more compact element of the right-hand side of (5), namely the lower partial moment $H_X(\gamma)$, as the equivalent measure of upgrade.

The expression $p(m_X-\gamma(p))$, in which we have stressed the dependence of γ on p, has a further interpretation which makes our theory tractable. With probability p the manager will know that no new information is available on the company's future value. Conditional on this absence of information the

[5]The mathematics of equilibrium existence for the Dye model can be interpreted as the *expected advantage* over inferiors, as computed in the integral $H_X(.)$, rather than the *mean advantage over inferiors,* defined to be $H_X(.)/F(.)$ by Begstrom and Bagnoli (2005). For a survey of these concepts see Bergstrom and Bagnoli (2005). In view of its special role, we term the lower partial momemt function more briefly the *hemi-mean* function, and thus prefer the notation H switching round Begstrom and Bagnoli's notation (they use H to denote the upper partial moment function). The mathematics of equilibrium existence is thus much easier in the context of expected advantage, as too is the comparative statics of the equilibrium location, which can draw freely on the log-concavity features developed for the mean-advantage context of Bergstrom and Bagnoli (2005). We will see that in fact a significantly weaker notion suffices, namely ρ-concavity, with $\rho=-1$, as defined by Caplin and Nalebuff (1991a) – see Section 3 below.

manager could, if permitted, buy the stock at the interim market price γ and liquidate the stock at the terminal date. The expected terminal value is m_X given the absence of information. Thus ex-ante the manager holds an option with expected value (under the investors' risk neutral measure) equal to

$$p(m_X - \gamma). \tag{6}$$

If the manager can receive a share of this value in remunerations, then the expression above becomes the manager's objective function. This assumes that the manager's trade remains unobserved by the investors, as would be the case in the Kyle (1985) one-shot market model. In sequential market models (with dates in between the interim and terminal dates) the manager's trading could become observable; the revised managerial opportunity set necessitates that optimal managerial behaviour uses a mixed strategy of buying and selling in order to optimally preserve the manager's private information. One expects that the revised valuation of the manager's option to trade is a convex function of p, say $v(p)$, with zero value at the endpoints $p = 0$ and $p = 1$,just as is the case with $p(m_X - \gamma(p))$, for which see Ostaszewski and Gietzmann (2008). Our theory applies to such valuations; parsimoniously we work with the $p(m_X - \gamma(p))$, as it turns out to be a very tractable model.

The Dye equation identifies how γ should be selected by requiring a balance of expected upgrades with expected downgrades, but not how to select p. In subsection 2.4 below we cast the problem of selecting p as a trade-off between the downgrade term (risk shielding) and the upgrade term (value enhancement), and explain the trade-off in the formal setting of utility maximization. The trade-off exists because of the *countervailing effects*. The objective of risk-shielding under non-disclosure prescribes the selection of a cutoff γ (risk-shield) as large, and as close, as possible to μ to ensure that the value of the company does not fall below γ. To see the trade-off with enhanced valuation, consider that the extreme case $\gamma = \mu$ which implies $q = 0$ with $q = 0$ the manager will never see any news — including 'good' news.(and there will never be any disclosures) that is the risk shield ensures company value does not fall below γ but in the limit as $\gamma \to \mu$ the higher the risk shield the smaller the probability that management will be informed of good news x were $x > \mu$. Analogously in the other extreme case of $q = 1$ there will be no risk shielding and $\gamma = 0$ offering no risk-shielding as full 'unraveling' occurs since this is the Grossman and Hart (1980) setting.

2.2 Endogenous managerial information endowment

It is in this section where we differ from the Dye paradigm, in that the p (or q) which was an exogenous datum for Dye, becomes a choice variable for the manager. There are a number of routes for deducing the endogenous solution of p, the technical optimization via elementary calculus approach being one that is obvious and feasible, but not very insightful. It is suggested here that the most intuitive route commences (for the reasons explained immediately above) by regarding the manager as having to make a trade-off between risk shielding and enhanced valuation, achieved by assuming that the manager makes the trade-off in text-book fashion by reference to a general utility function. Let $U(x, y)$ describe managerial preference over the risk-shielding loss, assessed as $x = \mu - \gamma$ when setting the cutoff at γ, and value enhancement, assessed as $y = H(\gamma)$. We shall see in Section 3 another advantage of this approach: it makes explicit the two lotteries (over x and y) which determine the basis for stochastic dominance comparisons. We now find that under these assumptions the utility function is uniquely determined and has the following (standard) CES format: $U(x, y) = (x^{-1} + y^{-1})^{-1}$, and we therefore refer to it as the *implied utility* in order to stress that it is not imposed but derived from the underlying payoff structure (6). To prove this, suppose the manager chooses among the points in the opportunity set defined by:

$$\{(x, y) : 0 \leq x \leq \mu, \qquad y = H(\mu - x)\}.$$

employing the general utility function $U(x, y)$. Now in reduced form the Dye cutoff condition (3) for expected indifference between non-disclosure and disclosure is:

$$px = qy = (1 - p)y \qquad \text{or} \qquad p(x + y) = y, \tag{7}$$

which yields $p = y/(x + y)$. So the manager's maximization objective is now $p \cdot x$, hence eliminating p via (7)

$$U(x, y) = p \cdot x = \frac{xy}{x + y} = (x^{-1} + y^{-1})^{-1},$$

– compare Caplin and Nalebuff (1991a), who consider generalized averages such as this harmonic one. Note also that the contour $U = c$ is a pair of rectangular hyperbolas with centre of symmetry at $x = y = c$, and is expressible as

$$(x - c)(y - c) = c^2.$$

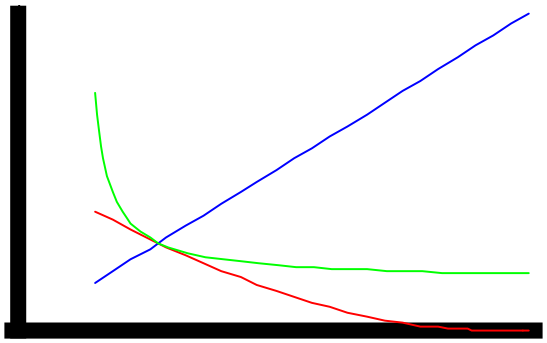

Figure 1. The arbitrage line (blue), the opportunity curve (red), and the tangential utility contour (green).

Letting

$$\lambda = \frac{p}{q}$$

be defined as the odds ratio, the common tangency (illustrated in Figure 1) of the utility contour and opportunity curve implies what we shall describe repeatedly as *the optimal odds equation* as follows.

Since $y = h(x) = H(\mu - x)$ implies $h'(x) = -F(\mu - x)$, and as

$$\frac{U_x}{U_y} = \frac{y^2}{x^2},$$

the common tangency condition implies that

$$\frac{dh}{dx} = -\frac{U_x}{U_y},$$

and, from Dye's equation $px = qy$, we deduce[6]

$$F(\gamma) = F(\mu - x) = \frac{y^2}{x^2} = \frac{p^2}{q^2} = \lambda^2. \tag{8}$$

[6]The optimal odds equation implies that $\lambda < 1$, i.e. $p < q$, or, finally, $p < 1/2$.

2.3 The Penno extension with noisy signals

The Penno (1997) extension to the Dye model assumes that managers do not see a realization of company value x, instead they see a noisy observation (noisy signal) of the underlying value, or state, x. Thus, letting T denote management's observed noisy signal, it is assumed that a noise variable Y with a general distribution, which is assumed for simplicity to be *independent* of X, enters the observation process and the manager now formally observes the *garbled* random variable $T = T(X, Y)$, for some given function $T(.,.)$. The manager thus computes an (updated) state estimate S, namely the expected value of the company given the observed noisy signal T, i.e. the conditional expected value of X given T. This is denote as:

$$\mu_X(T) := E[X|T(X,Y)],$$

where

$$\mu_X(t) := E[X|T(X,Y) = t]$$

is the *regression function*. That is $S = \mu_X(T)$ is the manager's estimate of X and is best in the sense of least squares. Note that

$$E[S] = E[E[X|T(X,Y)]] = E[X],$$

by the law of iterated expectation. (or, law of total expectation).

In order to understand how the introduction of noise modifies investors inferential process it is helpful to review what happens in a specific simple setting. The well-known case (cf. Kyle (1985)) of X and Y normal (and independent) with $T = X + Y$ and $m_Y = 0$ is given by

$$S = \mu_X(T) = m_X + \kappa(T - m_X), \text{ where } \kappa = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_Y^2}, \text{ and so} \quad (9)$$

$$\sigma_S^2 = \kappa^2 \sigma_T^2 = \frac{\sigma_X^4}{\sigma_X^2 + \sigma_Y^2} = \kappa \sigma_X^2. \quad (10)$$

Thus, to borrow from the *language* of Kalman filtering, κ is the optimal Kalman gain coefficient used to update the ex-ante prediction m_X, given the observed residual $T - m_X$. (Here again optimal means 'minimizing mean-square error'.) It is important to notice that in the presence of more σ_Y^2 noise the coefficient is smaller and so the weight given to the residual correction term is smaller. In the extreme, with very diffuse information the predictor

is close to m_X (hardly any updating / error correction) and so, with such very little updating, the derived σ_S^2 is very low.

Having seen how the relative magnitude of σ_Y^2 noise influences investors updating the next step is to consider the revised form of the optimal disclosure strategy. Supposing that the manager discloses according to a cutoff t for the observed signal T, following Penno's approach one should identify the indifference point for the value of t (in the noisy signal domain) by the equation

$$E[X|T(X,Y) = t] = E[X|ND(t)],$$

which we can write more compactly as

$$\mu_X(t) = E[X|ND(t)]. \tag{11}$$

That is, comparing (1) and (11), the effect of imposing noisy observation on management is for the observed (with certainty) value x to be replaced by its mean value $\mu_X(t)$, when the observation t is noisy. Thus a disclosure by the manager is now of the *estimated value* for the state S. Given the disclosure an *updated distribution* of future values for the company (in the Kyle setting) is again a normal distribution, but with its mean the updated state estimate and with a *reduced* updated variance $\kappa\sigma_Y^2$ which may be regarded as having been 'risk-adjusted'.

Having concentrated on the simple Kyle setting in order to develop intuition consider now the general setting. In general the best linear predictor[7] is:

$$\ell(X|T) = m_X + \frac{cov(X,T)}{\sigma_T^2}(T - m_T). \tag{12}$$

Combining this with the lower partial moment reformulation one can see how the risk-adjustment occurs in a general setting. First commence by considering $\mu_X^{-1}(.)$, the inverse of $\mu_X(.)$, which exists provided we assume that μ_X is *strictly* increasing. We will denote the inverse by L. Thus L is like T a mapping from an underlying value X to a noisy signal T. To find the cutoff γ, first find the (cumulative) distribution function of the random variable S; that is given by:

$$F_S(s) = F_T(L(s)),$$

in view of

$$S \leq s \text{ iff } \mu_X(T) \leq s \text{ iff } T \leq L(s).$$

[7] See for instance Roman (2004).

Now find the solution x_S to the canonical relationship (2) with X replaced by S, as follows:

$$\frac{p}{q}(m_S - \gamma) = H_S(\gamma), \qquad \text{where} \qquad H_S(\gamma) = \int_{z \leq \gamma} F_S(z)dz, \qquad \text{with} \qquad m_S = m_X,$$

Thus the disclosure cutoff may be calculated in three steps. Firstly, construct the cumulative distribution function of the (noisy) signal $F_T(.)$ from the distributions F_X and F_Y. Secondly, form the estimator distribution $F_S(.) = F_T(L(.))$ for the estimator $S = \mu_X(T)$. Finally, compute the Dye cutoff γ_S from the estimator distribution. In summary the manager announces[8] the estimate S provided a signal T has been received for which the estimator satisfies $S > \gamma_S$.

Thus, in the modified Penno setting of optimal disclosure in which management receives a noisy signal, the optimal strategy is of the same canonical form as the Dye cutoff, but with an appropriate change of variables.

2.4 Endogenous optimal disclosure intensity

This subsection is dedicated to considering whether, given the observation of a company's disclosure intensity τ, defined as

$$\tau = q(1 - F(\gamma)),$$

investors can make rational inferences concerning the amount of noise (σ_Y^2) associated with management's disclosures. One could proceed directly by trying to identify the formula linking the two variables. However there is a less cumbersome route. Since σ_Y^2 uniquely determines the cutoff γ (for given fixed σ_X^2), one can instead just consider how τ is related to γ as this is sufficient to establish the existence of a closed form functional relationship.

It follows from the definition of τ, by simple arithmetic[9], that

$$\tau = 1 - \lambda \text{ iff } F(\gamma) = \lambda^2.$$

But, the right hand-side condition is exactly the optimization condition derived in the discussion of subsection 2.2 concerning management's optimal choice of information endowment $\widehat{q} = 1 - \widehat{p}$. This can be summarized as:

[8] The cutoff $\hat{t}$ used by the manager given the observed noisy signal T is computed by the manager via $\hat{t} = \mu_X^{-1}(x_S)$.

[9] $1 - F(\gamma) = \frac{q+p}{q}\frac{q-p}{q} = \frac{q^2-p^2}{q^2}$.

Optimal Intensity Theorem. *The odds-ratio* $\lambda = \widehat{p}/\widehat{q}$ *and the intensity of disclosure* τ *sum to unity, i.e.*

$$\tau + \lambda = 1,$$

iff the value of p *is selected optimally as in Subsection 2.2 above, i.e.* $p = \hat{p}$*, or, equivalently* $\tau = \hat{\tau}$*. In this case the corresponding Dye cutoff, denoted* $\hat{\gamma}$*, and the odds ratio* $\hat{\lambda}$ *are related according to the rule*

$$\hat{\lambda} = \sqrt{F(\hat{\gamma})}. \tag{13}$$

We stress that the simplicity of this formula is evidence of the tractability of the valuation (6).

3 Monotonicity between disclosure intensity τ and signal noise σ_Y

We know from Penno's closing proposition of his (1997) paper, that for the special case of normally distributed underlying parameter and independently normally distributed 'additive' noise (with $\sigma_T^2 = \sigma_Y^2 + \sigma_X^2$) the disclosure intensity τ is unrelated to signal noise σ_Y. At issue then is how general is the possibility of such non-dependence. In general when a distribution F_T is parametrized by a scalar σ this has traditionally been interpreted as reflecting the relative 'riskiness' and consideration of the family of distributions $F_T(x, \sigma)$ has been a topic of central concern for portfolio management research for many years. A key construct for making comparisons within a given class of distribution is an investigation of various forms of 'stochastic dominance' properties of the distributions. See H. Levy (1992) for an overview.

This section presents two types of results. First in subsection 1 a set of general results which define the class of distributions defined by stochastic dominance criteria which admit monotonicity between disclosure intensity and signal noise σ_Y. It is argued that the class of distributions that admit monotonicity are in a meaningful sense reasonable, in particular since the class identified (wider than the log-concave ones) are such as are commonly assumed in the related literature on stochastic dominance and investment opportunities. After the general discussion of subsection one the analysis turns to consider two specific distributions. First the normal distribution

which does not admit monotonicity is investigated and then the log normality which does. Put simply this explains the final Penno proposition. He assumed a restrictive class of distribution (normal distributions for both true value and the 'additive' noise) that does not satisfy a mild extension of traditional stochastic dominance criteria. Once one moves to a distributional class satisfying the extended traditional criteria (and working, say with the log-concave distributions) there exists a formal link between disclosure intensity and signal noise, whereupon it is valid to assume in an empirical study that investors may draw inferences regarding managerial signal noise from observed disclosure intensity.

3.1 Stochastic dominance, noise and disclosure lotteries

For clarity we commence the discussion under the assumption we are in a Dye setting and then extend the analysis to the noisy observation setting of Penno. We begin by recalling from the end of subsection 2.3 that the manager may be regarded as facing trade-offs resulting from an equilibrium choice of the cutoff γ. (That is, although q is actually selected by the manager, we here think of the corresponding $\gamma = \gamma(q)$, as being selected.) We repeat for convenience: given a γ, if that γ were selected, the manager trades off two *countervailing effects*: risk-shielding versus enhanced valuation. The objective of risk-shielding under non-disclosure prescribes the selection of a cutoff γ (risk-shield) as large, and as close, as possible to μ to ensure that the value of the company does not fall below γ. To see the trade-off with enhanced valuation, consider that the extreme case $\gamma = \mu$ implies $q = 0$, which signifies that the manager will never see any news — including 'good' news.(and there will never be any disclosures). Likewise, enhanced valuation as represented by the expected value of the company under disclosure, which would be maximized by having $q = 1$. But $q = 1$ offers no risk-shielding – full 'unraveling' occurs. We recall here that the enhanced value may be interpreted as $H(\gamma)$ as explained in subsection 2.1

These considerations lead us to studying the relationship between two functions: $H(\gamma)/(\mu-\gamma)$, i.e. a *gain-to-loss ratio*, and $\sqrt{F(\gamma)}$, which identifies the *optimized odds* when the cutoff γcorresponds to an optimal selection of information-endowment by the manager. Up until now we have been working as if σ, the standard deviation of the distribution of company value, F did

not vary. We now make explicit the possibility of such variation and note this complicates our earlier explanation of 'countervailing effects'.

In the Penno disclosure setting, if there are two companies (with $\sigma_1 < \sigma_2$) and each manager announces the company's value has been estimated to be identically X^{est} in both cases, investors will prefer the first company with lower σ by risk aversion. To understand how investors' attitude to risk aversion will determine relative valuations in the disclosure setting we employ *first order stochastic dominance* (FSD, for short). From the above discussion of countervailing effects we see that the equilibrium conditions are determined by pairs of lotteries: one lottery representing risk-shielding the other value enhancement. The lottery pair is characterized by the two variables: the *cutoff* γ and the signal *noise* σ. Our comparison of pairs of lotteries is one where risk aversion matters, as it does in the standard mean-variance portfolio analysis. Hence the two functions earlier identified must explicitly display dependence on σ :

$$\begin{aligned} \Pi(\gamma,\sigma) &= \sqrt{F(\gamma,\sigma)}, \qquad \text{(optimized odds)} \\ \Lambda(\gamma,\sigma) &:= \frac{H(\gamma,\sigma)}{\mu-\gamma} \qquad \text{(gain-to-loss).} \end{aligned}$$

Recall that $\lambda = \Lambda$ is the Dye no-arbitrage equation (4) and $\lambda = \Pi$ is the optimized odds equation (13) which endogenizes the manager's information endowment $\hat{p}$,and both equations must hold simultaneously at an equilibrium.

Because both of these functions are increasing, we may regard them as inducing lottery pairs, that is they may be viewed as distributions (after rescaling Λ) of two random variables, respectively Z_Π and Z_Λ, that define a pair of lotteries in both of which success (interpreted as disclosure of a value above γ) corresponds to the events $Z_\Lambda > \gamma$ or resp. $Z_\Pi > \gamma$. The interpretation in terms of disclosure is valid when γ is selected as the manager's cutoff disclosure strategy (which necessarily requires that γ satisfies $\Pi = \Lambda$). Of course, the word 'failure' corresponds to the company being downgraded to a value of γ following non-disclosure.

Trade-offs between single lotteries (equivalently, between the success cut-off and the spread) are traditional analyzed using FSD. Recall that in Markowitz' portfolio theory rankings are made between vectors (μ, σ) on the basis of the natural order in respect of μ and its inverse in respect of σ. ("More μ to less is preferred, and less σ to more is preferred.")

Here the added complication is we have pairs of lotteries. So it is natural to extend FSD when ranking Π lotteries and similarly with Λ lotteries – see

(i) and (ii) in the Definition below. In general there need not be any further opportunity to compare lotteries. But suppose, for a fixed σ for some reason the Π lotteries offer greater chance of success than do Λ lotteries for all γ large enough (above some threshold determined by σ) as in (iii) below.

Under these circumstances, given an initial lottery pair with $\sigma = \sigma_0$ and with γ at the corresponding threshold γ_0 suppose that σ, γ are increased above these initial values, we consider what comparisons may be made between lotteries defined by parameter pairs (γ, σ). Specifically, consider a succession of steps of simultaneously reducing σ (starting from $\sigma = \sigma_V > \sigma_0$, say) and increasing γ (starting from γ_0, say) so as to hold the chance of failure in the Π lottery constant. To reflect consistent risk-aversion one would expect at each step a compensation in the Λ lottery (in terms of an increased chance of success, i.e. lower Λ value). Likewise, for a succession of steps in which Λ is held fixed would require compensation in the Π lottery (in terms of an increased chance of success, i.e. lower Π value).

As we can see in the illustration of Figure 2 this passage between the lotteries of triangle PML in the (Π, Λ)-plane presents a clockwise orientation for the triangle. That figure is consistent with our narrative if and only the directions of increasing σ, γ contours are as indicated, that is to say the transformation $(\gamma, \sigma) \mapsto (\Pi, \Lambda)$ is orientation preserving. That is what the

definitions below formalize.

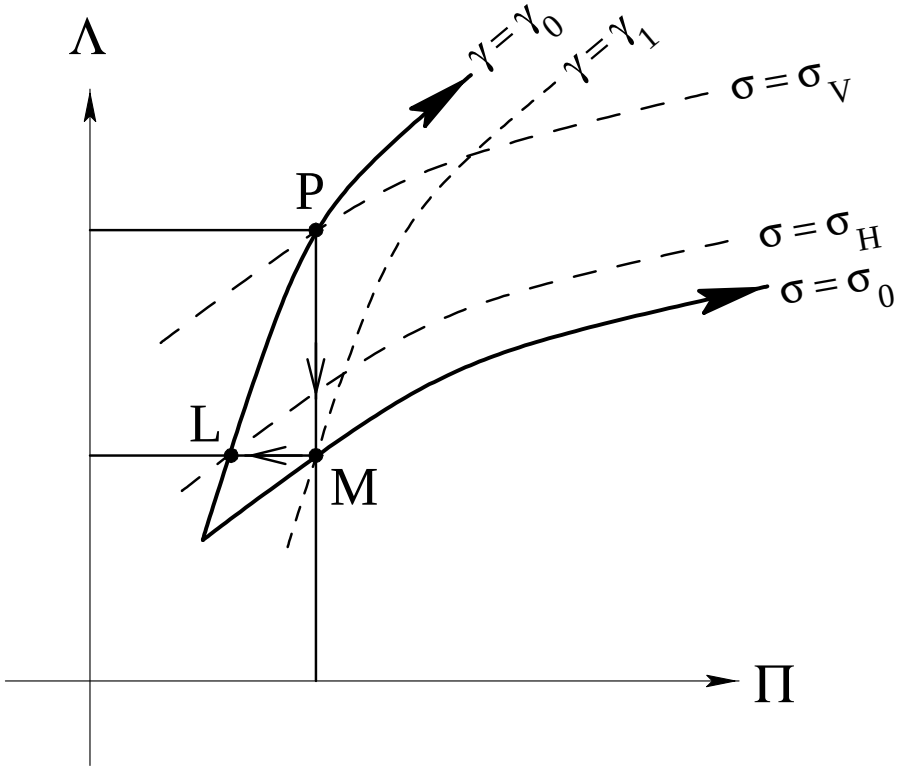

Figure 2. Π and Λ contours. The orientation on the points P, M, L signifies preferences of risk-averse agents over corresponding lotteries for appropriate γ, σ values as indicated..

Lottery definition. Let $\pi(\gamma, \sigma)$ and $\lambda(\gamma, \sigma)$ be two distributions parametrized by $\sigma > 0$. Suppose that for some μ we have:

(i) $\pi(\gamma, \sigma_1) < \pi(\gamma, \sigma_2)$, for all $\gamma < \mu$, and $0 < \sigma_1 < \sigma_2$,

(ii) $\lambda(\gamma, \sigma_1) < \lambda(\gamma, \sigma_2)$, for all $\gamma < \mu$, and $0 < \sigma_1 < \sigma_2$,

(iii) for each σ, there is a unique $\hat{\gamma}(\sigma)$, such that $\pi(\gamma, \sigma) < \lambda(\gamma, \sigma)$, for all γ with $\hat{\gamma}(\sigma) < \gamma < \mu$.

We say that the two families of distributions $\{\pi(\gamma, \sigma), \lambda(\gamma, \sigma)\}$ exhibit *joint stochastic dominance*, if the mapping $(\gamma, \sigma) \to (\pi, \lambda)$ is orientation preserving, equivalently its Jacobian determinant is (strictly) positive, i.e.

$$\frac{\partial(\pi, \lambda)}{\partial(\gamma, \sigma)} := \begin{vmatrix} \pi_\gamma & \pi_\sigma \\ \lambda_\gamma & \lambda_\sigma \end{vmatrix} > 0.$$

Strong First Degree Stochastic Dominance: We say that the single family of distributions $F(x, \sigma)$ has *strong first degree stochastic dominance* if $F(x, \sigma_1) \leq F(x, \sigma_2)$ whenever $0 < \sigma_1 < \sigma_2$ (i.e. first degree stochastic dominance obtains), and in addition the transformation $(\gamma, \sigma) \rightarrow (\pi, \lambda)$ is orientation preserving, or equivalently, the Jacobian

$$\frac{d(\pi, \lambda)}{d(\gamma, \sigma)} = \begin{bmatrix} \pi_\gamma & \pi_\sigma \\ \lambda_\gamma & \lambda_\sigma \end{bmatrix}$$

has positive determinant, where π, λ are defined from F as $\pi = \sqrt{F}$ and $\lambda = H/(\mu - \gamma)$ with $H = \int F(z, \sigma)dz$, the hemi-mean function.

Monotonicity Theorem for disclosure intensity τ and signal noise σ
In any region of a model in which the state estimator distributions F_S are ***strongly dominant*** *relative to σ_S, the intensity of disclosure τ_S is decreasing in σ_S (and so increasing in σ_Y given a fixed σ_X).*

The intuition for this result is as follows. The higher the signal noise σ_Y the more an investor has to be compensated for the risk that a given disclosure is imprecise (in the limit becoming spurious), hence management need to compensate investors by increasing the probability (intensity) of disclosure. To summarize then if the distributional assumption for the valuation random variable admits *strong dominance* then investors can infer that the management of companies with higher observed disclosure intensity τ are facing greater signal noise σ_Y.

In the following two subsections we then consider specific distributions, first the class of scale and location preserving distributions which include the normal which do not satisfy the monotonicity theorem and then the log normal distribution which does.

3.2 The normal distribution and the class of location and scale distributions

We begin by considering a location and scale class of models which includes the Penno (1997) normal model setting. This class of models have many favourable features, in particular a straightforward representation of the Dye cutoff. Unfortunately, an important disadvantage when pursuing inferences from voluntary disclosure behaviour (disclosure intensity) in this case is a lack of dependence on the signal noise σ_Y

With inter-company comparisons in mind, we begin with a standardization exercise.

Standardization Theorem. *Let $\Phi(x)$ be an arbitrary zero-mean, unit-variance, cumulative distribution defined on $\mathbb{R}$. For the following location and scale family of distributions $\Phi(\frac{x-\mu}{\sigma})$, with mean μ and variance σ^2, the Dye cutoff $\gamma(\mu, \sigma, \lambda)$ satisfies*

$$\gamma(\mu, \sigma, \lambda) = \mu - \sigma\xi(\lambda), \text{ where } \lambda = \frac{p}{q}$$

and where

$$\xi(\lambda) = -\gamma(0, 1, \lambda)$$

is the cutoff, when standardizing to zero mean and unit variance, and is a function only of the odds-ratio. The standardized cutoff $\xi(\lambda)$ is a convex and decreasing function of λ and satisfies

$$\lambda = H_\Phi(-\xi)/\xi,$$

where $H_\Phi(x) = \int_{-\infty}^{x} \Phi(t)dt$ is the corresponding lower partial moment.

The first consequence of the standardization result is a simple but highly significant result.

Corollary 1. *For the location and scale family above and for a fixed odds-ratio $\lambda = p/q$, the cutoff for the observed noisy signal $\gamma(\mu, \sigma, \lambda) = \mu - \sigma\xi(\lambda)$ recedes away from the mean at a constant rate $\xi(\lambda)$, as the precision is reduced (noise σ_Y is increased).*

The Theorem and Corollary taken together greatly aid development of intuition as follows. The Corollary identifies the statics of the cutoff for changes in the location and scale parameters. Thus, working within the Dye framework (i.e. when endowed, the manager has perfect information – receives a clean/non-noisy signal), if the companies in a given industry have value distributions in a common scale and location family, Corollary 1 demonstrates that, following non-disclosure, the value of the company is reduced to a cutoff value that distinguishes between companies with the same mean μ (expected return), but different variances σ^2. Holding μ constant, companies with higher variance are subject to greater downgrade in value.

Next in the step from the clean signal X of the Dye model to the noisy signal T of the Penno model, there is in general also a change of distribution: from H_X to H_S, where $S = \mu_X(T)$ is the regression (best estimate) of the

underlying value X given T. But, it may happen that S has the same probability law as X, apart from a change in scale and location, in which case it is usual to say that the laws of S and X are of the same *type.* In this case, provided $m_X = m_T$ (unbiased signal), the cutoff as identified in Corollary 1 varies, merely by an adjustment from $\sigma = \sigma_X$ to $\sigma = \sigma_T$.

Also using our notation, in Penno's model we have $T(X, Y) = X + Y$ (as in Kyle) and so the regression function is affine as in (12), and hence H_S is indeed in the scale and location family of X. Thus in Penno's model it is *as though the noisy signal T was clean albeit with an adjusted variance (or, more properly, as though the noisy signal was the estimator).*

Furthermore, the Standardization Theorem evidently includes as a special case the standard normal distribution and thus elucidates the following observation in Penno (1997). Using our notation, Penno asserts that "the threshold γ_T increases as information quality (precision) ρ_Y increases" (i.e. as $\rho^{-1} := \sigma_Y^2$ decreases). Indeed, this is clear from substituting σ_T for σ, where $\sigma_T^2 = \sigma_X^2 + \sigma_Y^2$, since X, Y are assumed independent.

Finally, it is possible to state a more general form of the final Penno proposition of (1997).

Corollary 2. *For the location and scale family above, the manager maximizes the value of*

$$p(\mu - \gamma_T(\mu, \sigma, \lambda)) = p\sigma_T\gamma(0, 1, \lambda) \text{ with } \lambda = p/(1-p).$$

over p by a choice of $p = \widehat{p}$ that is independent of σ_T.

Having shown immediately how tractable the Penno type model is, the next result identifies the limitation in Penno's setting: the voluntary disclosure intensity is constant across companies, precisely because the regressor $S = \mu_X(T)$ has a distribution in the same scale and location family as X and with identical locations (means) of S and T.

Invariance Theorem. (cf. Penno, Proposition 1). *If X and S have distribution/law of the same type (their distributions are equivalent under a change of Location and Scale), then the theoretical intensity of voluntary disclosure is invariant under changes in scale, equivalently in precision. Specifically it takes the form*

$$\hat{\tau} := (1 - \hat{p})(1 - \Phi(\xi(\widehat{\lambda})) = 1 - \sqrt{\Phi(\xi(\widehat{\lambda}))},$$

where $\xi(\lambda)$ *satisfies*

$$\lambda = H_\Phi(-\xi)/\xi,$$

and $H_\Phi(x) = \int_{-\infty}^{x} \Phi(t)dt$ *is the lower partial moment of an arbitrary distribution* Φ *with mean zero and unit variance.*

Thus in a normal model setting, such as that in Penno (1997), the optimal level $\hat{p}$ of information endowment is independent of the noise parameter σ_Y although the cutoff $\hat{\gamma}(\mu, \sigma_T)$ still varies (affinely) with σ_Y.[10]

Arguing from intuition, one may expect that $\hat{\tau}$ is proportional to $\hat{q}$ with a constant of proportionality approximately $(1 - \Phi(m_X))$, on the grounds that the cutoff γ is not far below the mean. It is also natural to expect the optimally selected $\hat{q}(\rho)$ to be decreasing in ρ, on the grounds that a better estimate when information is more error-prone may be gained from a larger number of observations. Indeed, Penno (1997) posits functional forms for this behaviour. Thus one expects that inferior companies (those with lower ρ) to have higher $\hat{\tau}$. This, as we cite in the introduction, is reported as a paradox, "contrary to the popular notion that higher-quality information is accompanied by more voluntary disclosure, the paper has demonstrated that this notion is, in general, not true." (p. 280). In fact as we have shown this result is in accord with Shin's comments (see abstract) on how some disclosure strategies may seem unintuitive unless disclosure decisions are taken in equilibrium. This analysis shows clearly why the more risky companies, those where management face more noisy signals, in equilibrium have a higher disclosure intensity.

3.3 Log-normal models

In this subsection we work with a distributional assumption that both satisfies strong first degree dominance and is the traditional modelling assumption for securities (company equity value), namely the log-normal distribution. The model is defined by specifying the underlying company value X as

$$X = m_X e^{U - \frac{1}{2}\sigma_U^2},$$

[10] Furthermore, if there is a cost attached to the precision $\rho_Y = 1/\sigma_Y^2$ there is scope for variation in $\hat{p}$ against the cost $C(\rho)$. Thus, one may derive the relationship between $\hat{p}$ or $\hat{q}$ and the optimized value of $\bar{\rho}_Y$ – much as one derives indirect profit against (optimized) output in microeconomics. This insight may be used to study Penno's proposed dependence of $q(\rho)$ on ρ – see Penno's equation (A1) – by comparison with $\hat{q}(\rho)$.

with U a normal zero-mean random variable with variance σ_U^2. The noisy signal is modelled multiplicatively as $T = T(X, Y) = XY$ with

$$Y = e^{V - \frac{1}{2}\sigma_V^2},$$

a random variable independent of X, so that V is independent of U and is normal zero-mean with variance σ_V^2. Hence,

$$T = m_X e^{W - \frac{1}{2}\sigma_W^2}$$

where $W = U + V$ is normal zero-mean with variance

$$\sigma_W^2 = \sigma_U^2 + \sigma_V^2.$$

Thus the observed signal T is again a log-normal variable, but with an adjusted variance. It is straightforward to show that the estimator, $S = \mu_X(T)$, is also a log-normal variable, in fact a simple power function transform of T. This leads to:

Intensity Invertibility Theorem *In the log-normal model the value of the optimal level $\hat{p}$ of information endowment (as well as the cutoff) are identifiable from the intensity τ.*

The log-normal model is reasonably tractable as an analytic tool, since all relevant random variables are from the same family. (It is a log-scale model rather than a location and scale family, because of the non-linearity of the logarithm.) A key result is the Regressor Functional Form proposition below. However we first identify the usual disclosure functions within the log-normal setting.

The clean (Dye) signal cutoff for X (had that been the observed signal) is given by

$$\hat{x} = m_X \cdot \widehat{\gamma}, \tag{14}$$

where $\widehat{\gamma} = \widehat{\gamma}(\lambda, \sigma_U)$ is the solution to the equation

$$\lambda(1 - \widehat{\gamma}) = H_{\mathrm{LN}}(\widehat{\gamma}, \sigma_U), \tag{15}$$

where H_{LN} denotes the hemi-mean function for the log-normal and is given by:

$$H_{\mathrm{LN}}(\gamma, \sigma) = z \cdot \Phi_{\mathrm{N}}\left(\frac{\log(\gamma) + \frac{1}{2}\sigma^2}{\sigma}\right) - \Phi_{\mathrm{N}}\left(\frac{\log(\gamma) - \frac{1}{2}\sigma^2}{\sigma}\right).$$

To aid intuition recall (2) and note that the hemi-mean on its right-hand side may be interpreted as the valuation of a call-option struck at the money on the mean m_X, hence the familiar appearance of the Black-Scholes formula, here without the factor m_X which cancels against the same factor on the left-hand side $\lambda(m_X - m_X \cdot \widehat{\gamma}) = m_X\lambda(1-\widehat{\gamma})$ where $\lambda = \frac{p}{q}$.

It is straightforward to show that the conditional expectation estimator (leading readily to the regression function) is given by

$$X^{\text{est}} = S = m_X \exp\left(\kappa W - \frac{1}{2}\kappa^2\sigma_W^2\right) = m_X \exp\left(\kappa W - \frac{1}{2}\kappa\sigma_V^2\right).$$

So the cutoff for the estimator X^{est} is given by

$$\hat{x}^{\text{est}} = m_X \cdot \widehat{\gamma}^{\text{est}},$$

where $\widehat{\gamma}^{\text{est}} = \widehat{\gamma}(\lambda, \kappa\sigma_W)$ is the solution to the equation

$$\lambda(1-\gamma) = H_{\text{LN}}(\gamma, \kappa\sigma_W).$$

Note that there is a double adjustment here: the variance σ_U^2 is replaced by the observed signal variance σ_W^2 and is then downgraded by the factor κ :

$$\kappa = \frac{p_V/p_U}{1 + p_V/p_U} < 1.$$

In summary we have:

Regressor Functional Form. *If $T = XY$ and X, Y are independent log-normally distributed random variables, then the regression function is concave and follows a power law:*

$$\mu_X(t) = E[X|T=t] = e^{\frac{1}{2}\kappa(1-\kappa)\sigma_W^2} m_X \left(t/m_X\right)^{\kappa},$$

where $X = m_X e^{U - \frac{1}{2}\sigma_U^2}$ and $Y = e^{V - \frac{1}{2}\sigma_V^2}$ (with U, V independent normal zero-mean variates) and

$$\kappa = \frac{\sigma_U^2}{\sigma_U^2 + \sigma_V^2} = \frac{p_U}{p_U + p_V}, \text{ and } \sigma_W^2 = \sigma_U^2 + \sigma_V^2.$$

Given a power format, the value of κ is easily guessed, whereupon the constant of proportionality may be computed using the fact that $E[\mu_X(T)] =$

m_X (by the 'law of total probability') – see below. Thus the inverse regression function $L(x)$ is convex and is given by:

$$L(x) = m_X e^{-\frac{1}{2}(1-\kappa)\sigma_W^2}(x/m_X)^{1/\kappa}.$$

Corollary. *The random variable $X^{\text{est}} = S = E[X|T]$ has mean m_X and is log-normal with representation:*

$$m_X \exp\left(\kappa W - \frac{1}{2}\kappa^2\sigma_W^2\right) = m_X \exp\left(\kappa W - \frac{1}{2}\kappa\sigma_U^2\right),$$

where W is zero-mean normal with variance

$$\sigma_W^2 = \sigma_U^2 + \sigma_V^2.$$

Log-normal disclosure intensity: *The disclosure intensity in the log-normal model is given by:*

$$\hat{\tau} = \hat{q}(\kappa\sigma_W)\left(1 - \Phi_{\text{LN}}(\hat{z}^{\text{est}}, \kappa\sigma_W)\right) = 1 - \hat{p}(\kappa\sigma_W)/\hat{q}(\kappa\sigma_W),$$

$$\textit{where } \Phi_{\text{LN}}(\gamma, \sigma) = \Phi_{\text{N}}\left(\frac{\log(\gamma) + \frac{1}{2}\sigma^2}{\sigma}\right),$$

and Φ_{N} is the standard normal distribution.

An illustrative example graph of $\hat{\tau}$ against precision as measured by κ, as defined in equation (10), is offered in Figure 3.

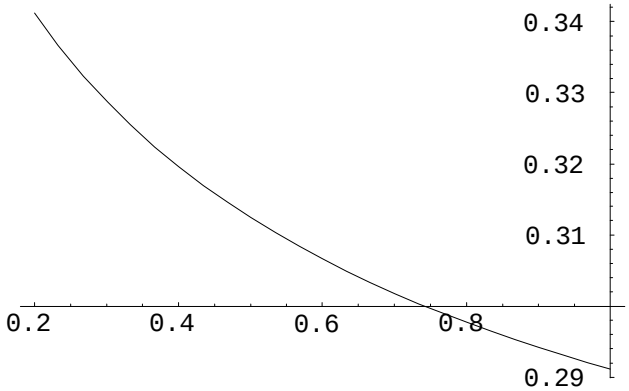

Figure 3. The theoretical intensity $\hat{\tau}$ of voluntary disclosure as a function of κ.

4 Empirical analysis with disclosure intensity

A recent paper in this area is the work of Cousin and de Launois (2006). In their work they consider traditional competing models of conditional volatility; the GARCH specification and a Markov Switching two state market model. The innovative feature they introduce is that they argue that information arrival affects stock return volatility. That is, from our perspective they are arguing that news intensity affects conditional volatility; so, for instance, in the GARCH framework the specification of conditional variance is given by:

$$\sigma_{i,t}^2 = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2 + \lambda_i N_{i,t} \tag{16}$$

where the new term $N_{i,t}$ is a proxy[11] for the number of company i specific news events announced to the stock market per interval t. Their main objective is to compare and contrast the performance of this adjusted GARCH model to a two state Markov Switching Regression (MSR) model where now the disclosure intensity determines the probability that a company under consideration is either[12] in a low or high volatility regime.

What is of particular interest for us is that they assume disclosure intensity is an important empirical explanatory variable for conditional volatility as modelled theoretically in our framework. In the GARCH framework their empirical findings are consistent with our theoretical predictions in that the conditional volatility is increasing in disclosure intensity and in the MSR framework the probability of being in the high volatility state is increasing in disclosure intensity. Thus their empirical tests appear to be broadly in line with our theoretical predictions. However, before coming to this conclusion we believe it is important to raise an important note of caution. Of critical importance is how Cousin and de Launois measure disclosure intensity. As Table 1 makes clear they simply record the frequencies of Factiva disclosures by category. However, if one just records all the raw empirical disclosure intensities for companies this does not capture the essential features of our generalised Dye-Penno model for the following reason. The theoretical model is of *voluntary* disclosures, that is the model concerns only those news wires corresponding to management receiving information about future events that

[11] They measure the variable by identifying the frequency of a subset of firm news releases on Factiva.

[12] To be more precise the disclosure intensity in part determines whether the state regime dummy variable $D_{i,t}$ is above or below a threshold that the firm is in the high volatility regime.

affect their *voluntary* ability to issue the news wires and thus indicate value *above* the Dye cutoff. Companies in addition are required under regulatory provisions to make *mandatory* disclosures. Thus the raw data on disclosure intensities is a mix of disclosure 'types', whereas the theory only speaks to the 'above Dye cutoff' voluntary disclosures. Thus, when working with raw disclosure intensity data an essential step is to implement an estimation procedure for separating out the voluntary Dye type disclosures.

With this empirical issue in mind one procedure could be to exploit the distributional assumptions of the model and then the Dye cutoff can be shown to be close to the mean (just below) and one can use this to validate an empirical approach which measures *dimensionless relative intensity,* i.e. excess relative to the mean in proportion to standard deviation. Looking at disclosure intensities above the mean rate ('high rates') abstracts away from mandatory good news disclosures that happen on a regular basis. Thus restricting attention only to high intensity disclosure periods, we need to distinguish between those that approximate to good news (voluntary disclosures) and those that approximate to bad news (mandatory disclosures), typically driven by regulations put in place to protect investors from delay of bad news disclosure. In order to identify which are good news and which are bad news disclosures when there is no standard "message space" for voluntary disclosures, it is suggested here that one could identify good news disclosures as those that give rise to an increase in analysts' consensus forecasts and so exclude those that give rise to a decline in analysts' consensus forecasts for the company.

In contrast recent research by Rogers, Schrand and Verrecchia (2008) (RSV) use an EGARCH model which allows them to estimate the conditional variance when modelled as one of two functions depending on the sign of the return shock. The intuition behind this asymmetric modelling assumption is that "bad news" seems to have a more pronounced effect on conditional volatility than has "good news". For many companies there is a strong negative correlation between the current stock returns and future volatility. The tendency for conditional volatility to decline when returns rise (following good news) and rise when returns fall (following bad news) is typically referred to in behavioural finance as the *leverage effect.* RSV propose that when companies follow a strategy of reporting good news and withholding bad news this can be described as 'strategic disclosure'. In a setting where good news is taken at face value, bad news below the cutoff threshold has to be inferred by investors, and it is this difference in the in-

ferential process that leads to the asymmetric responses in the market. To see this in the limiting case of full disclosure remove the leverage (asymmetric) effect whereupon current changes in valuation (impounded in returns) would always be associated with recent news arrival rather than the need for investors to make inferences following non-disclosure. Rather than look at actual disclosures, RSV instead develop two hypotheses about the leverage effect. The first is that the leverage effect is stronger for companies about which there is less private information; that feature is assumed to increase the threshold level of disclosure (implying a lower disclosure intensity). The second is that the leverage effect will be weaker when increased litigation risk affects a company's propensity to adopt a 'strategic disclosure' strategy. RSV report interesting results; however, our research on disclosure intensity suggests an alternative empirical implementation. Specifically, they use the variable PUBINFO as a measure of private information. That measure captures the extent to which information is likely not to be private, because on their analysis, if company returns move together then, ceteris paribus, homogeneity subsists in that sector of industry; so there is less private information when results of company operations are similar. Thus, they do not actually measure disclosure intensities. Accordingly, on the view that our model may have wider empirical applicability than the special two-case scenario investigated by RSV, we suggest that an EGARCH model variant of the standard GARCH model, redesigned so as to refer to disclosure intensities in (16), may also be worth investigating.

5 Conclusion

We have shown that in equilibrium the managers of companies facing higher signal noise will rationally increase their disclosure intensity. That is, working back from observed disclosure intensity, investors can infer that ceteris paribus high intensity disclosing companies are more risky, as management's truthfully disclosed estimates have larger standard deviations associated with them (for instance, because the managers are subject to greater noise in their operating environments). This theory, based on generalizations of the established Dye-Penno models, suggests both new empirical testing procedures and also critically a different direction in assumed causation. The theory shows why one should not base empirical hypotheses on an a priori assumption that 'better' companies make more voluntary disclosures, since we have

shown that it is in fact the companies with the most poorly informed management (facing highest noise) which will in equilibrium disclose with the greatest intensity.

The research is subject to a number of caveats. We abide by the assumptions of the Dye model in regard to truthful disclosure and the inability of credible disclosure of absence of information. The model is essentially a single period project model in which success in one period does not influence successes in later periods. That is, multi-period project dependence (and related disclosure) is not modelled. This is clearly a topic for future research. Furthermore mangers here make disclosures according to their own optimal cutoff rather than mimicking a different manager type; any other behaviour would require an alternative model.

We note that the model is robust to changes in the valuation model (6) to other (differentiable) concave valuations (see subsection 2.1): a small perturbation of that value function is reflected in small perturbations elsewhere in our analysis.

6 References

Bergstrom,T.C. and M. Bagnoli, (2005), "*Log-concave Probability and it's Applications",* Economic Theory, 26, 445-469.

Caplin, A. and B. Nalebuff, (1991), "*Aggregation and social choice: a mean voter theorem",* Econometrica 59 (1) 1-24.

Cousin, J.-G. and T. de Launois, (2006), "*News Intensity and Conditional Volatility on the French Stock Market", Finance,*Presses Universitaires de France,VOL 27; PART 1, pages 7-60.

Dye, R.A., (1985), "*Disclosure of Nonproprietary Information*", Journal of Accounting Research, 23, 123-145.

Easley, D.; S. Hvidkjaer and M. O'Hara, (2002), "*Is Information Risk a Determinant of Asset Returns*" Journal of Finance, Vol. 57 Issue 5, 2185-2221.

Easley and M. O'Hara, (2004), "*Information and the cost of capital*"Journal of Finance, Vol. 59 Issue 4, 1553-1583.

Easley, D.; O'Hara, M.; Paperman, (1998), "*Financial Analysts and Information-Based Trade*", Journal of Financial Markets; Vol. 1, p175-201.

Grossman, S., and O. Hart, (1980) "*Disclosure Laws and Take-over bids",* Journal of Finance, 35, 323-34.

Jung, W. and Y. Kwon, (1988), "*Disclosures when the market is unsure of information endowment of managers",* Journal of Accounting Research, 26, 146 - 153.
Kyle, A. (1985), "*Continuous auctions and insider trading*", Econometrica 53, 1315 - 1335.
McNeil, A.J, Frey, R., and Embrechts, P., "*Quantitative Risk Management*", Princeton University Press, (2005)
Penno, M.C., (1997), "*Information Quality and Voluntary Disclosure*", The Accounting Review, 72, 275-284.
Rogers, J.L., C.M. Schrand and R.E. Verrecchia, (2008) "*The impact of strategic disclosure on returns and return volatility: Evidence from the 'leverage effect',* working paper Wharton School.
Roman, S. "*Introduction to Mathematical Finance: From Risk Management to Options*" Springer (2004).
Shin, H.S., (2006), *Disclosure Risk and Price Drift,* Journal of Accounting Research, Vol. 44 Issue 2, 351-379,
Verrecchia, R.E., (1990), *Information Quality and Discretionary Disclosure,* Journal of Accounting & Economics, Vol. 12, Issue 4, p365-380.

7 Appendix: Newswire vs. disclosure intensity (empirics)

Here we suggest how to model an empirical link between disclosure intensity and the intensity with which companies issue newswires.

Having established that the disclosure intensity is a theoretically valid construct upon which to base empirical study, a number of cautionary remarks need to be stressed before actual empirical implementation is attempted. If one just recorded all the raw empirical disclosure intensities for firms this would not capture the essential features of the disclosure model for the following reason. The theoretical model is of *voluntary* disclosures, that is its logical connection is only with newswires corresponding to management receiving information about future events which affect their *voluntary* ability to issue the newswires (and so to indicate value *above* the Dye cutoff). In addition firms are required under regulatory provisions to make *mandatory* disclosures. Thus the raw data on disclosures is a mix of several disclosure 'types', whereas the theory only speaks to the 'above Dye cutoff' voluntary

disclosures. Thus when working with raw disclosure intensity data an essential step is to implement an estimation procedure for separating out the voluntary Dye type disclosures.

With this empirical issue in mind it is straightforward to check that under various distributional assumptions the Dye cutoff is close to the mean (just below) and use this to validate an empirical approach which measures *dimensionless relative intensity,* i.e. excess relative to the mean in proportion to standard deviation. We shall derive such a quantity below.

First, turning to the issue of *mandatory* disclosures, an important element of the empirical investigation is how to separate out when a disclosure is voluntary (as in the Dye model) or alternatively mandatory. Since we will be looking at disclosure intensities above the mean rate ('high rates') we will be abstracting away from mandatory good news disclosures that happens on a regular basis. Thus restricting attention only to high intensity disclosure periods, we need to distinguish between those that approximate to *good news* (voluntary disclosures) and those that approximate to *bad news* (mandatory disclosures), typically driven by regulations put in place to protect investors from delay of bad news disclosure.

In order to identify which are good news and which are bad news disclosures, when there is no standard "message space" for voluntary disclosures, it is suggested here that we identify good news disclosures as those that give rise to an *increase in analysts' consensus forecasts* and so exclude those that give rise to a decline in analysts' consensus forecasts for the firm.

Our basis for modelling is the assumption that the number of a firm's newswire releases in any unit period of time takes the form

$$N = N_V + N_M,$$

where the two independent random variables are N_V, relating to voluntary disclosure, and N_M, relating to mandatory disclosure are Poisson random variables. For each, we assume that the Poisson rate of disclosure is dependent on the information-endowment of management. The state of endowment is modelled as

$$\omega = (e, t, z)$$

where: $e \in \{u, i\}$ is the manager's endowment type (uninformed/informed), $t = x \cdot y$ is a realization of $T(X, Y) = XY$, as in subsection 2.3 above, and z is a realization of an independently distributed random variable Z, which represents those aspects of the firm which are governed by mandatory

disclosure requirements. It is assumed in both cases that when the Poisson rate of disclosure, θ, is non-constant, then it is affinely related to observed signals. Thus the disclosure intensity follows one of two state-dependent regimes (according as disclosure does or does not occur) as shown below:

$$\theta_V = \begin{cases} \alpha t + \beta, & \text{if } e = i \ \& \ t \geq \gamma, \ \text{(Disclosure)} \\ \alpha\gamma + \beta, & \text{if } e = i \ \& \ t < \gamma, \ \text{or if } e = u. \ \text{(Non-Disclosure)} \end{cases}$$

Thus θ_V follows the same constant regime in the non-disclosure region in the outcome space $\{u, i\} \times \mathbb{R}_+$ of the voluntary random variable. We propose to treat the mandatory variable in a similar fashion: we presume that there is a lower-threshold (fall in value) for the variable Z, namely ζ (which precipitates intensive disclosure activity), whose value we model exogenously. Thus θ_M follows one of two (state-dependent) regimes:

$$\theta_M = \begin{cases} a - bz & \text{if } z < \zeta, \\ a - b\zeta = \delta & \text{if } z \geq \zeta, \end{cases}$$

with $a, b > 0$. For both random variables, we thus use the probability assignments

$$P[N = n] = e^{-\theta}\frac{\theta^n}{n!},$$

conditional on the state which determines θexplicitly (as above). So, conditional on θ being constant (depending on regime), we have

$$E[N] = \sum_n n e^{-\theta}\frac{\theta^n}{n!} = \theta.$$

We refer to the region where $e = i$, $t \geq \gamma, z \geq \zeta$ as the Good News region.

We note that the unconditional mean of N is simply:

$$E[N] = E[N_V] + E[N_M],$$

and furthermore point out that expectations can be computed by conditioning on the various regimes. For example,

$$E[N_V] = E[E[N_V|ND] + E[N_V|D]],$$

and similarly for N_M.

Now, if one conditions on good news GN, the Poisson rate of $N = N_V + N_M$ is $\alpha t + \beta + \delta$. So, using bars to denote such conditional expectations, we have

$$\begin{aligned} \bar{N} &= \alpha \bar{T} + \beta + \delta, \\ \bar{\sigma} &= \alpha \bar{\sigma}_T. \end{aligned}$$

Thus, for any given realization N, we compute the score statistic N^* to be

$$N^* = \frac{N - \bar{N}}{\bar{\sigma}} = \frac{T - \bar{T}}{\bar{\sigma}_T},$$

so that the score statistic N^* (in the Good News region) is independent of the signalling parameters α, β, δ. We call this the 'good news count' statistic of the firm and suggest that this be the basis for empirical analysis.