WORKING PAPER SERIES

Availability of AI tools and their effect on the auditing process

Jens Robert Schöndube/ Barbara Schöndube-Pirchegger

Working Paper No. 04/2025

OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

FACULTY OF ECONOMICS
AND MANAGEMENT

Impressum (§ 5 TMG)

Herausgeber:
Otto-von-Guericke-Universität Magdeburg
Fakultät für Wirtschaftswissenschaft
Der Dekan

Verantwortlich für diese Ausgabe:
J. R. Schöndube, B. Schöndube-Pirchegger
Otto-von-Guericke-Universität Magdeburg
Fakultät für Wirtschaftswissenschaft
Postfach 4120
39016 Magdeburg
Germany

http://www.fww.ovgu.de/femm

Bezug über den Herausgeber
ISSN 1615-4274

Availability of AI tools and their effect on the auditing process

Preliminary

Jens Robert Schöndube
Institute of Managerial Accounting
Leibniz Universität Hannover
Königsworther Platz 1, 30167 Hannover, Germany
Phone: +49/511/7628131
E-Mail: schoendube@controlling.uni-hannover.de

Barbara Schöndube-Pirchegger
Chair in Accounting and Control
Otto-von-Guericke Universität Magdeburg
Universitätsplatz 2, 39106 Magdeburg, Germany
Phone: 0049/391/6758728
E-Mail: barbara.schoendube@ovgu.de
(corresponding author)

Availability of AI tools and their effect on the auditing process

Abstract

In this paper we model the interaction between an auditor and a client firm. The client firm's manager can either report truthfully or commit fraud. The auditor needs to plan a two stage audit that allows to detect fraud. In the first stage an AI tool is employed that provides a signal about the quality of the client's internal control system (ICS). Classifying the ICS as weak or strong, the signal alters the auditor's expectations regarding the client's fraud probability. In the second stage, the auditor decides about her audit effort conditional on the information provided by the AI. Comparing the AI setting to a benchmark setting without AI use, we find that employing the AI tool reduces the manager's incentives to commit fraud. At the same time it reduces the equilibrium effort provided by the auditor. As a consequence, the probability that actual fraud is detected remains unchanged. We extend our model and allow the AI tool to be customized such that it can either focus on detection of the weak ICS, the strong ICS, or on both equally. We find that the AI specification that minimizes ex ante probability for fraud not necessarily coincides with the specification that minimizes auditing costs. It follows that the auditor in charge of customizing the AI cannot necessarily be expected to do so in a fraud minimizing way.

Keywords: Artificial Intelligence, Auditing, Game Theory, Fraud detection

1. Introduction

Recent advances in AI seem to affect almost all types of professions including the auditing profession. In particular, AI tools are either expected to replace standardized but labor intensive tasks or they do so already.[1] As such, AI is supposed to increase labor efficiency but also changes the job descriptions for many employees.

With regard to the auditing profession, practitioners as well as scientists envision that AI use will not only increase audit efficiency as it relieves auditors from time consuming tasks, but might also help to improve audit effectiveness as it increases audit quality and/or reduces fraud. Empirical evidence that hints in this direction has been presented in a recent study by Law and Shen (2024), who find that AI use reduces the number of misspecified going concern opinions and also achieves more accurate assessments of internal controls in audits. In an earlier contribution Issa et. al. (2016) argue that the availability of AI facilitates the automation of routine tasks in auditing but may also be used as an assistant for non-routine tasks, leaving the decision authority with the auditor. Kokina and Davenport (2017) emphasize that AI may be particularly relevant in auditing with respect to identifying anomalies. On a broader level, AI use might even be considered as a means to restore market participants' trust in financial statements that allegedly suffered as a result of several high profile audit failures.[2]

In this paper, we analyze a game-theoretic interaction between an auditor and a client firm in order to shed some light on the effects of AI use in auditing. In particular, we consider a client firm, whose financial statements need to be audited, and an auditor. The manager of the client firm can set up the financial statements correctly or commit fraud. The auditor, in turn, needs to plan her audit including the use of AI. In line with Eisikovits et.al. (2024) AI is perceived as "a group of statistical machine learning technologies which recognize patterns in large data sets and offer predictions based on those patterns" in what follows.

In our model AI is used to assess the client firm's internal control system (ICS), which can be either strong or weak. Specifically, we assume that the AI detects discrepancies from expected patterns in the data known as accounting anomalies. A large amount of anomalies hints towards a weak ICS. Accordingly, the AI tool predicts the ICS to be weak, if it detects a sufficiently high amount of anomalies and to be strong otherwise.[3] If the ICS is weak, however, it is easier for the client firm's manager to

[1] See, Webb (2020).
[2] See e.g. Blake (2024).
[3] See Kokina and Davenport (2017) and Law and Shen (2004) from above but also, e.g. the German Wirtschaftsprüferkammer (2025).

override the controls and to commit fraud.[4] Once the auditor has observed the ICS classification provided by AI, she decides upon any follow up auditing activities.

We consider two different settings. In the first setting we assume that the classification, or signal, the AI tool provides is informative but imperfect and that its precision is given exogenously.

We find that the additional information provided by the AI results in a decrease in audit effort. This is pretty much in line with the idea that AI replaces some of the auditor's tasks and enhances efficiency.

However, we also find that the probability for actual fraud to remain undetected increases with AI, rather than to decrease. Simultaneously, the probability that a client firm's manager commits fraud decreases. Finally, the two effects, namely the increase in undetected fraud and decrease in the probability that fraud is committed, compensate each other in equilibrium. As a result, the ex ante probability that fraud is not detected remains unaffected by AI use. Accordingly, our model suggests that audit quality, at least if defined as the probability that actual fraud is detected, suffers in the presence of an AI tool. However, AI makes up for this outcome by discouraging fraudulent behavior in the first place.

In the second setting we consider a more sophisticated AI tool. We assume that the auditor can "customize" the tool and specify a critical amount of anomalies that suffices for the ICS to be classified as weak. That way, she can focus on identifying a weak control system or on identifying a strong control system or she can specify the AI to be "neutral" and to focus evenly on both. However, if the tool focuses on detecting the weak system, it will inevitably misspecify a strong system as weak with considerable positive probability and v.v.

In a first step of our analysis, we aim at minimizing the probability for fraudulent behavior. We find that it is either optimal to fully focus on the weak system or on the strong one. It is never optimal to focus on both evenly. If the strong system is more likely than the weak one and additional costs for committing fraud in the presence of a strong system are sufficiently small, it is optimal to focus on the strong system. Otherwise focusing on the weak one minimizes the probability that fraud is committed.

However, while reducing fraud is probably desirable from an investor's or capital market's perspective, the actual choice about the AI specification is made by the auditor. Therefore, we investigate in a second step which AI specification maximizes the auditor's payoff. It turns out that the choice that maximizes the auditor payoff does not always coincide with the one that minimizes expected fraud. It

[4] See Smith et.al. (2000) for a similar argument.

follows that we cannot necessarily expect that AI applied in auditing is used in the most desirable way from a capital market or even general public perspective.

Our paper is closely related and somewhat builds on Smith et al. (2000). They consider a setting in which an auditor can split his effort between two tasks. The first task aims at identifying a possibly weak control system. Given the system has been identified as weak or not, the auditor decides about additional audit effort referred to as substantive testing. While the benchmark case in their setting and ours is similar, the subsequent analysis differs in various respects. Most important, Smith et.al. assume that the auditor spends effort to detect a weak ICS while we assume that an AI tool is present that replaces the auditor. Accordingly, in their model effort costs arise for identifying a weak system and the auditor might decide not to spend any effort in doing so, if e.g. costs are too high. In our paper, we assume that using the AI is costless or, at most, caused an up-front investment expenditure which is sunk. Moreover, Smith et.al. (2000) assume that the auditor either detects a weak system or fails to detect it. The probability to classify a strong system as weak is zero by assumption. In our setting, the AI produces a signal that classifies the ICS as strong or weak, while both classifications can possibly be wrong. In addition, and in order to reflect typical AI attributes, we assume that the auditor can customize the AI tool such that detection probabilities become endogenous. Overall, Smith et.al. (2000) focus on optimal effort allocation and resulting reductions in audit costs. In our paper, the main focus is on whether AI reduces auditor effort and on its effect on incentives for managerial fraud and fraud detection.

Our analysis is also inspired by Kwon (2005). Investigating the effect of accounting conservatism on management incentives, he assumes that an accounting signal is observed that needs to be converted in the financial statements to either state a high or a low result. Depending on the threshold that needs to be exceeded to justify a high report, he classifies the accounting system as liberal, neutral, and conservative. In a somewhat similar approach we assume that the AI tool is customized to either detect a weak and a strong ICS with identical probability (neutral), or to focus either on the weak or the strong system.

The rest of the paper is organized as follows. The next section presents the model. In section three we consider a benchmark case in which no AI tool is available. Section four analyses the setting in which an AI tool is available that detects the client's type of control system with exogenous probability. We extend the model to allow for endogenous AI specification in section five and subsequently present optimal AI specification in order to minimize fraud probability and auditing costs. Section six concludes.

2. Model

We consider a client firm, whose financial statements need to be audited, and an auditor or audit firm. The auditor can use an AI tool that offers insights with regard to the client firm's internal control system.

There are two types of client firms: Firms that have a strong internal control system (ICS) implemented and firms with a weak one. If the ICS is weak, the number and size of irregularities and discrepancies from expected patterns in the financial data, so called accounting anomalies, tends to be higher than with a strong ICS. Therefore, a weak ICS makes it relatively easy for the manager of the client firm to commit and hide fraud, while doing so is costly if the controls are strong. We denote the firm with the strong ICS a type-1 firm (t_1) and the weak ICS firm a type-2 firm (t_2) in what follows. The ex ante probability for a type-1 firm to be present is $\theta \in (0,1)$ and for a type-2 firm it is $1-\theta$. If a type-i firm, with $i = 1,2$, is present, the manager commits fraud with probability γ_i, which is endogenous in the model.

The AI tool analyzes the accounting data provided by the client in order to detect accounting anomalies. Based on the amount of anomalies detected, it provides a signal $\hat{s}_1$ or $\hat{s}_2$ that either classifies the client as type-1 or type-2 firm.

In a first part of our analysis, we assume that the AI identifies the client type correctly with some exogenously given probability. p_1 denotes the probability that the strong ICS firm is identified correctly while p_2 refers to the weak type's identification. The auditor uses the signals to revise his beliefs regarding the client's type and the related fraud risk, and plans audit effort accordingly.

In a second step, we will introduce additional assumptions regarding the AI tool's procedure. Specifically, we allow the auditor to instruct the AI to either focus equally on both types of clients or to place relatively more emphasis on one type than on the other. For instance, the auditor may specify the AI procedure in such a way that it detects one type of client correctly with certainty but misspecifies the other type with considerable probability. Formally, we endogenize the probabilities p_1 and p_2.

The course of the game is as follows. In a first step nature determines whether a type-1 or a type-2 client is present. While the type is unobservable for the auditor, the firm's manager learns the type of ICS and decides whether to commit fraud. If an AI-tool is available, it provides a possibly imperfect signal $\hat{s}_j$, $j = 1,2$, that either indicates that the internal control system of the firm is strong or weak. The auditor decides about audit effort a_j, given the information $\hat{s}_j$ the AI tool has provided. The larger the audit effort, the higher the probability to detect fraud during the auditing process. If no fraud has

been committed, the auditor never detects fraud (no false detection). After completing the audit, an audit opinion is formed and made public.

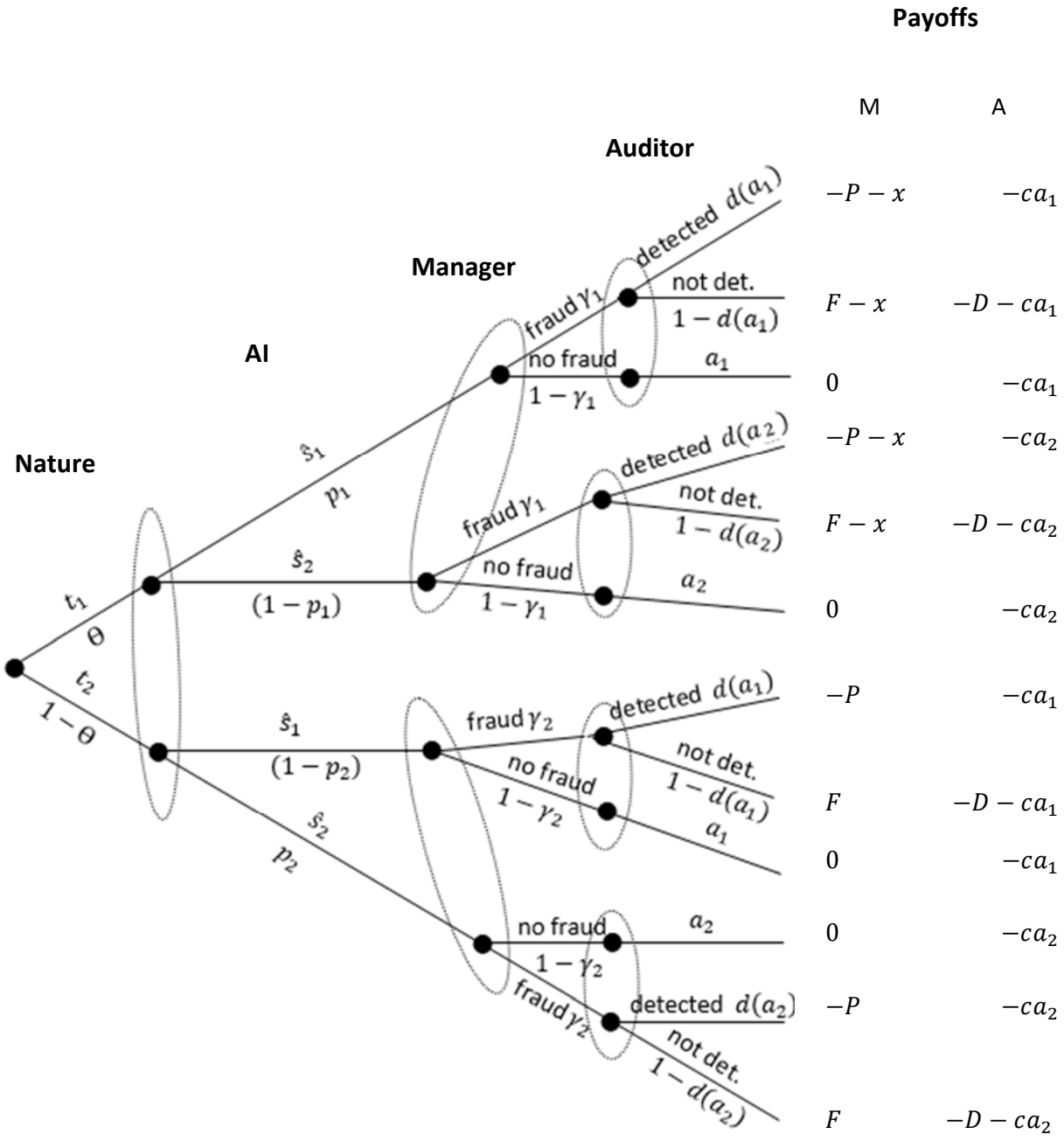

Figure 1: Game tree.

We further assume that an undetected fraud provides the manager with a benefit F. If the manager's fraud is detected, that results in a penalty P. In line with the characterization above, committing fraud in the presence of a strong ICS causes costs of x for the manager. It is costless with a weak ICS.

The auditor collects some fixed fee for the audit, which we neglect in what follows. AI use is assumed to be costless.[5] If she fails to detect fraud, this leads to future costs of D, e.g., due to loss in reputation if the fraud is detected later on by another party.

The auditor performs effort a_j, which increases the probability of detecting fraud. Specifically, $d(a_j) = 1 - \exp[-ba_j]$ is the probability of detecting fraud conditional on fraud having occurred. $b > 0$ is a parameter that reflects the effectivity with which effort increases this probability. Note that the higher a_j, the larger the probability to detect fraud for a given b. If $a_j = 0$, however, $d(0) = 0$ follows and fraud is never detected. Performing effort causes a disutility from working hard equal to ca_j, with $c > 0$. The course of the game is reflected in the game tree depicted in figure 1.

To sustain any equilibrium of our model we assume $x < \min\{F, \frac{F(p_1+p_2-1)}{p_2}, \frac{P(p_1+p_2-1)}{1-p_2}, \frac{P(p_1+p_2-1)^2}{1-p_2}\frac{\theta}{p_2}\}$ and

$$c(F+P) < Db(1-\theta)\min\left\{P, \frac{\theta H}{(1-\theta)(P\theta(p_1+p_2-1)^2 - x(1-p_2)p_2)}, \frac{H}{P(1-\theta)(p_1+p_2-1)^2 - x((p_1+p_2-1)^2\theta - (1-p_1)(1-2p_2) - p_2^2)}\right\},$$

with $H = P((p_1+p_2-1) - x(1-p_2))(P((p_1+p_2-1) + xp_2)$.

These regularity conditions ensure that the manager's fraud probabilities are always between zero and one and that the auditor's effort is non-negative.

3. Benchmark Case: No AI Available

In a first step, we consider a benchmark setting in which no AI tool is available.

With no AI in place, the auditor receives no type-specific information so that her effort (a) is unconditional on any observation. Accordingly, the auditor needs to decide about audit effort based on ex ante expectations regarding the client type. She incurs costs from the audit, which she aims to minimize:

[5] One might assume that acquisition costs arise for the AI tool, which are sunk later on.

$$E(\Pi^A) = -\theta\gamma_1\big(1 - d(a)\big)D - (1-\theta)\gamma_2\big(1 - d(a)\big)D - ca$$

$$= [-\theta\gamma_1 - (1-\theta)\gamma_2]\big(1 - d(a)\big)D - ca \quad (1)$$

As described above, costs arise from reputation damage D if the manager does commit fraud and this remains undetected by the auditor. In addition costs ca arise from auditing effort.

The manager, in contrast, learns the type of ICS present before he decides about committing fraud.

If the ICS is strong (t_1), expected payoff is

$$E(\Pi^M|t_1) = \gamma_1\big[F\big(1 - d(a)\big) - Pd(a) - x\big]. \quad (2)$$

If the ICS is weak (t_2), expected payoff is

$$E(\Pi^M|t_2) = \gamma_2\big[F\big(1 - d(a)\big) - Pd(a)\big]. \quad (3)$$

Note that there is an extra cost of x for committing fraud only if the ICS is strong. Besides, if the manager commits fraud he gets F if he gets away with it and pays a penalty P, if he is caught.

The manager maximizes his payoff by picking his fraud probabilities $\gamma_i,\ i = 1,2$, as best responses to the conjectured auditor effort. The auditor maximizes her payoff (minimizes her cost) by picking auditing effort a optimally. Thus, she maximizes (1), as a best response to the conjectured γ_1 and γ_2.

Proposition 1:

There exists a unique equilibrium in which $\gamma_1^{BM} = 0$, $\gamma_2^{BM} = \frac{c(F+P)}{bDP(1-\theta)}$, and $a^{BM} = \frac{1}{b}\ln\left[\frac{F+P}{P}\right]$ holds.

All proofs are relegated to the appendix.

Proposition 1 states that the equilibrium consists of a mixed strategy played by the manager if the ICS is weak, a pure strategy of no fraud if the ICS is strong, and a pure strategy played by the auditor in picking a.

Intuitively, the manager is willing to randomize between committing fraud and not to do so, only if his expected payoffs are equal. In the absence of fraud, his payoff is zero. In the presence of fraud, the manager's payoffs differ in the type of ICS, as shown in (2) and (3). It follows directly that the manager cannot be indifferent between committing fraud and not doing so in both types simultaneously. We demonstrate in the proof of Proposition 1 that the only equilibrium contains an effort choice a of the

auditor that renders the manager indifferent when facing a weak ICS.[6] This implies that the manager strictly prefers not to commit fraud if the ICS is strong. If the ICS is weak, however, the manager picks the probability for committing fraud, γ_2, such that the auditor's effort choice is indeed optimal.

Corollary 1:

The ex ante probability for fraud to arise in the benchmark setting equals $(1-\theta)\gamma_2^{BM} = \frac{c(F+P)}{bDP}$. The conditional probability of fraud to remain undetected in the presence of fraudulent behavior is $1 - d(a^{BM}) = \exp[-ba^{BM}] = \frac{P}{F+P}$. Accordingly, the ex ante probability for fraud to remain undetected is $(1-\theta)\gamma_2^{BM}\left(1 - d(a^{BM})\right) = \frac{c}{bD}$.

Given the equilibrium amounts of γ_2^{BM} and a^{BM} as derived in Proposition 1, we can easily calculate the probability for fraud to arise and the conditional and ex ante probability that fraud remains undetected by the auditor. Note that the ex ante probability for fraud increases in the manager's benefit from fraud, F, and the cost of auditor effort, c. It decreases in the auditor's detection efficiency b, her reputation costs, D, and the penalty, P. In contrast, the conditional probability for fraud to remain undetected increases in P and decreases in F. Now the ex ante probability for fraud to remain undetected, is independent of managerial benefits and costs from committing fraud. As, P and F affect the probability for fraud to be committed but also the conditional probability that is undetected, in equilibrium both effects cancel each other out, which is important for later reference. Our benchmark result is equivalent to the one in Smith et.al. (2000).[7]

4. AI-tool is available

If the auditor uses the AI-tool, she observes the signal from the tool before she chooses her auditing effort. The signal either states that the ICS is strong (signal $\hat{s}_1$), or that it is weak (signal $\hat{s}_2$). We assume that the AI tool's signals are imperfect but informative[8], implying that the following conditions hold for $p_1 = \Pr(\hat{s}_1|t_1)$ and $p_2 = \Pr(\hat{s}_2|t_2)$:

[6] Assuming, in contrast, that he is indifferent if the ICS is strong would result in a violation of or previous regularity conditions, see the appendix.
[7] See Smith et.al. (2000), Proposition 1.
[8] See also Kwon (2005).

$p_1 \leq 1, p_2 \leq 1$, and $2 > p_1 + p_2 > 1$.

It follows that at least one signal realization is informative about the underlying type of ICS, i.e., $p_j > 0.5$ for $j = 1$ or $= 2$, it even might be perfect. The other one might be below or above 0.5. Both, however, are "on average" informative about the underlying types.

The respective payoff functions in the presence of the AI-tool are as follows. The manager's expected payoffs depending on the observed type of ICS, t_i, are given by

$$E\left(\Pi_{t_1}^{M}\right) = \gamma_1[p_1(F\exp[-ba_1] - P(1 - exp[-ba_1]) + (1 - p_1)(F\exp[-ba_2] - P(1 - exp[-ba_2])] - x, \quad (7)$$

$$E\left(\Pi_{t_2}^{M}\right) = \gamma_2[(1 - p_2)(F\exp[-ba_1] - P(1 - \exp[-ba_1]) + p_2(F\exp[-ba_2] - P(1 - \exp[-ba_2])]. \quad (8)$$

Note that $a_j, j = 1,2$, in (7) and (8) refers to the auditor's effort choice, given she has observed $\hat{s}_j$ from the AI tool. As she picks her effort only after she has observed the output from the AI tool, she maximizes her expected payoff conditional on the signal provided.

$$E(\Pi^A|\hat{s}_1) = -D\exp[-ba_1]\left(\gamma_2 \Pr(t_2|\hat{s}_1) + \gamma_1 Pr(t_1|\hat{s}_1)\right) - ca_1, \quad (9)$$

$$E(\Pi^A|\hat{s}_2) = -D\exp[-ba_2]\left(\gamma_2 \Pr(t_2|\hat{s}_2) + \gamma_1 Pr(t_1|\hat{s}_2)\right) - ca_2, \quad (10)$$

where $\Pr(t_2|\hat{s}_1) = \frac{(1-\theta)(1-p_2)}{(1-\theta)(1-p_2)+\theta p_1}$, $\Pr(t_1|\hat{s}_2) = \frac{\theta(1-p_1)}{\theta(1-p_1)+(1-\theta)p_2}$, $\Pr(t_1|\hat{s}_1) = \frac{\theta p_1}{(1-\theta)(1-p_2)+\theta p_1}$, $\Pr(t_2|\hat{s}_2) = \frac{(1-\theta)p_2}{\theta(1-p_1)+(1-\theta)p_2}$.

In equilibrium $\{\gamma_1^*, \gamma_2^*, a_1^*, a_2^*\}$ the manager plays a mixed strategy for both types of ICSs and the auditor chooses pure-strategy audit efforts a_j conditional on the signal revealed by the AI. More specifically, audit efforts are chosen such that the manager is indifferent between committing fraud and not doing so, no matter whether the ICS is strong or weak. In addition, the manager picks the probabilities of committing fraud γ_1 and γ_2 such that the auditor indeed finds it optimal to pick audit efforts a_1 and a_2 as described before.

The manager is indifferent between committing fraud and refraining from it under both types of ICSs if and only if:

$$-P + \exp[-ba_2](F + P)(1 - p_1) + \exp[-ba_1](F + P)p_1 - x = 0,$$

$$-P + \exp[-ba_1](F + P)(1 - p_2) + \exp[-ba_2](F + P)p_2 = 0.$$

The first-order conditions for the optimal auditor efforts are given by:

$$\frac{dE(\Pi^A|\hat{s}_1)}{da_1} = \frac{bD\exp[-ba_1]\gamma_2(1-p_2)(1-\theta)}{(1-p_2)(1-\theta)+p_1\theta} + \frac{bD\exp[-ba_1]\gamma_1 p_1\theta}{(1-p_2)(1-\theta)+p_1\theta} - c = 0,$$

$$\frac{dE(\Pi^A|\hat{s}_2)}{da_2} = \frac{bD\exp[-ba_2]\gamma_2 p_2(1-\theta)}{p_2(1-\theta)+(1-p_1)\theta} + \frac{bD\exp[-ba_2]\gamma_1(1-p_1)\theta}{p_2(1-\theta)+(1-p_1)\theta} - c = 0.$$

Solving the four equations for $\{\gamma_1, \gamma_2, a_1, a_2\}$ results in:

Lemma 1:

The equilibrium values for fraud and auditor effort are given by:

$$\gamma_1^* = \frac{c(F+P)[P(p_1+p_2-1)^2\theta + (p_2-1)p_2x]}{bD\theta[P(p_1+p_2-1)-(1-p_2)x][P(p_1+p_2-1)+p_2x]},$$

$$\gamma_2^* = \frac{c(F+P)[P(p_1+p_2-1)^2(\theta-1) + (p_1+2p_2-1-2p_1p_2-p_2^2+(p_1+p_2-1)^2\theta)x]}{bDp(\theta-1)[P(p_1+p_2-1)-(1-p_2)x][P(p_1+p_2-1)+p_2x]},$$

$$a_1^* = \frac{1}{b}\ln\left(\frac{(F+P)(p_1+p_2-1)}{P(p_1+p_2-1)+p_2x}\right),$$

$$a_2^* = \frac{1}{b}\ln\left(\frac{(F+P)(p_1+p_2-1)}{P(p_1+p_2-1)-(1-p_2)x}\right).$$

It follows that

$$\begin{aligned} E(\gamma^*) &= \theta\gamma_1^* + (1-\theta)\gamma_2^* \\ &= \frac{c(F+P)(p_1+p_2-1)\big[P(p_1+p_2-1)+x\big(2p_2-1-\theta(p_1+p_2-1)\big)\big]}{Db(P(p_1+p_2-1)+p_2x)\big(P(p_1+p_2-1)-x(1-p_2)\big)}, \end{aligned}$$

$$E(a^*) = \Pr(\hat{s}_1)\frac{1}{b}\ln\left(\frac{(F+P)(p_1+p_2-1)}{P(p_1+p_2-1)+p_2x}\right) + \Pr(\hat{s}_2)\frac{1}{b}\ln\left(\frac{(F+P)(p_1+p_2-1)}{P(p_1+p_2-1)-(1-p_2)x}\right)$$

holds.

Comparing the results from Lemma 1, we observe that $a_2^* > a_1^*$ and $\gamma_2^* > \gamma_1^*$. Intuitively, a type-1 firm has weaker incentives to commit fraud than a type-2 firm as it is more costly to do so. As the auditor

is aware of this fact and the signal received from the AI is informative, she spends more effort on detecting fraud if the AI produces $\hat{s}_2$ then if she observes $\hat{s}_1$.

Comparing the findings from the benchmark setting and the setting with AI we obtain the results stated in Proposition 2.

Proposition 2:

(i) The ex ante probability for fraud to arise is lower in the presence of AI than in its absence.

(ii) The ex ante expected auditor effort is lower in the presence of an AI tool than in its absence.

(iii) The conditional probability for fraud to remain undetected increases with AI if the ICS is strong, is unaffected if the ICS is weak and increases in expectation.

(iv) The ex ante probability for fraud to remain undetected is unaffected by the presence of an AI tool.

The results from Proposition 2 show that (i) an AI tool helps to reduce the ex ante probability for the client firm's manager to commit fraud. At the same time the AI tool's contribution to detect fraud somewhat replaces the auditor's work, such that (ii) auditor effort decreases in equilibrium. However, the "on average" reduction in the auditor's effort also increases the probability that she does not detect fraud if it is present as stated in (iii). It is in that sense that audit quality is reduced in the presence of an AI. In equilibrium the decrease in the probability to commit fraud along with an increased probability that actual fraud is not detected implies (iv), the ex ante probability of fraud to remain undetected is unaffected.

5. Endogenous AI-signal provision

So far we assumed that the conditional probabilities for the AI-generated signals to be correct, namely $p_i = \Pr(\hat{s}_i|t_i)$ with $i = 1,2$, are exogenously given. In what follows we relax this assumption and rather allow the auditor to specify detection probabilities. As we will demonstrate below, the auditor can either decide to equally focus on both types of clients, or to put emphasis on one type at the cost of the other. E.g., the auditor may consider correct detection of the weak ICS particularly important. In that case she can instruct the AI to detect the weak ICS with a large probability, or even with certainty.

Doing so, however, comes at the cost that the strong ICS will be classified as a weak one with larger probability.

5.1. Additional structural assumptions

To endogenize detection probabilities, we model the amount[9] of accounting anomalies present in the financial data as a continuous stochastic variable $\tilde{s}$ and we denote its realization by s. We assume that the distribution of $\tilde{s}$ differs in the type of client such that the AI "picks" the actual amount of anomalies from one out of two distributions. Specifically, we assume that the amount is uniformly distributed on $[\underline{s}, \overline{s}_1]$ if the ICS is strong, $t = t_1$, and it is uniformly distributed on $[\underline{s}_2, \overline{s}]$, if the ICS is weak, $t = t_2$. Assuming that $\underline{s} < \underline{s}_2 < \overline{s}_1 < \overline{s}$ ensures that a firm with a strong (weak) ICS exhibits accounting anomalies in a lower (upper) range. This is in line with our previous assumption that anomalies tend to be higher with a weak ICS. However, as the distributions overlap, it is not necessarily possible to infer the type of client from the observation of s. We further assume that $\overline{s}_1 - \underline{s} = \overline{s} - \underline{s}_2 = \delta$ for simplicity. Hence, the pdf and cdf conditional on t_1 and t_2 are given by:

$$f_1 = f(s|t_1) = \frac{1}{\delta}, F_1 = F(s|t_1) = \frac{s - \underline{s}}{\delta}, \underline{s} \leq s \leq \overline{s}_1$$

$$f_2 = f(s|t_2) = \frac{1}{\delta}, F_2 = F(s|t_2) = \frac{s - \underline{s}_2}{\delta}, \underline{s}_2 \leq s \leq \overline{s}$$

Our distributional assumptions are depicted in figure 2. The dotted line refers to the firm with a strong ICS, t_1, and the dashed one to the firm with a weak one, t_2.

[9] The "amount" of accounting anomalies is representative for a score produced by the AI that incorporates number and size of the anomalies.

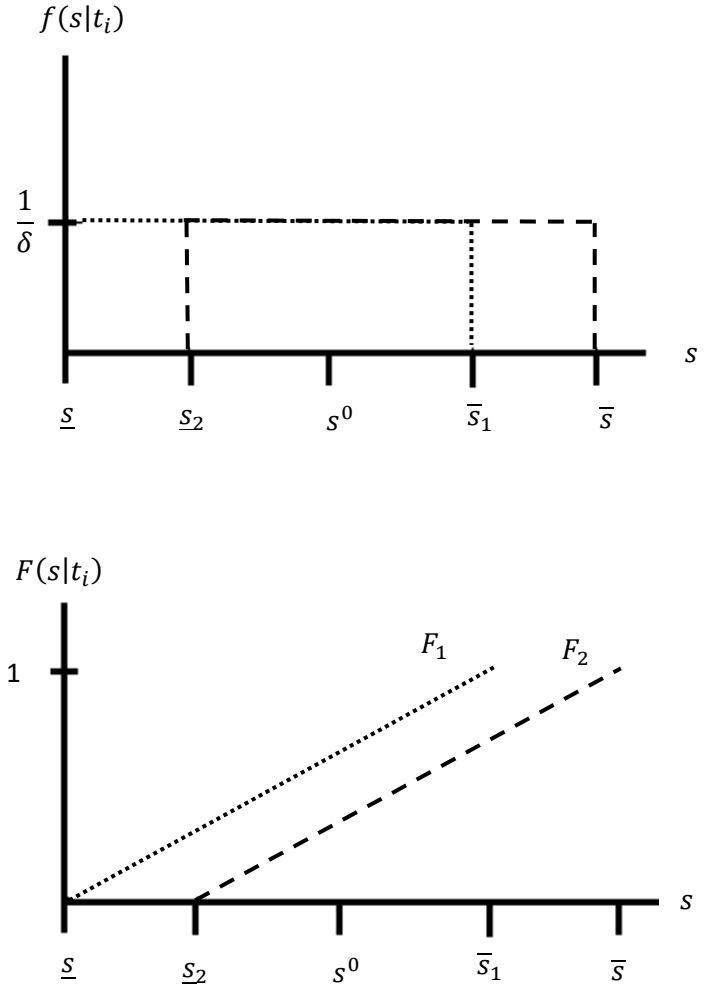

Figure 2: Probability distribution functions and cumulative distribution functions of signal s conditional on client type t_i.

Now the AI detects the amount of anomalies s and classifies the firm as either type-1 or type-2 client, tantamount to a report of either $\hat{s}_1$ or $\hat{s}_2$. For the AI to make this classification, the auditor needs to prescribe a critical s^c such that $\hat{s}_1$ is reported if $s \leq s^c$ and $\hat{s}_2$ gets reported if $s > s^c$ holds.

One way to do this is to specify s^c in such a way that the conditional probability for the report to be correct is identical for both types of clients. We call this a **neutral** AI specification.

It implies that $s^c = s^0$, where s^0 is implicitly defined by $F_1(s^0) = 1 - F_2(s^0)$ such that $s^0 = \frac{\bar{s}+\underline{s}}{2} = \frac{\bar{s}_1+\underline{s}_2}{2}$. Defining $F_1(s^0) = 1 - F_2(s^0) = n$, we can write the conditional probabilities as

$p_1 = \Pr(\hat{s}_1|t_1) = n$ and $p_2 = \Pr(\hat{s}_2|t_2) = n.$

Alternatively, s^c can be shifted in one or other direction such that $s^c = s^0 - \Delta$. The AI now reports $\hat{s}_1$ if $s \leq s^0 - \Delta$ holds and $\hat{s}_2$ otherwise. s^c is naturally restricted such that $\underline{s}_2 \leq s^c \leq \overline{s}_1$ implying that $s^0 - \overline{s}_1 \leq \Delta \leq s^0 - \underline{s}_2$.[10]

As $\Delta \neq 0$ implies that the probability of correctly identifying one type of firm increases while the one for the other type decreases, we refer to this as a **non-neutral** AI specification. The resulting detection probabilities can be written as follows.

$$p_1 = \Pr(\hat{s}_1|t_1) = F_1(s^0 - \Delta) = \underbrace{F_1(s^0)}_{n} - \underbrace{\frac{\Delta}{\delta}}_{m} = n - m \qquad \text{and}$$

$$p_2 = \Pr(\hat{s}_2|t_2) = 1 - F_2(s^0 - \Delta) = \underbrace{1 - F_2(s^0)}_{n} + \underbrace{\frac{\Delta}{\delta}}_{m} = n + m.$$

If the AI is specified such that s^c is at its upper bound, $\overline{s}_1$, this implies that

$p_1 = F_1(s^0) - \frac{s^0 - \overline{s}_1}{\delta} = 1$, or, equivalently, $m = m^- = \frac{s^0 - \overline{s}_1}{\delta}$,

so that a type-1 firm is always correctly identified, while

$$p_2 = F_2(s^0) + \frac{s^0 - \overline{s}_1}{\delta} < 1.$$

In contrast, if s^c is at its lower bound, $\underline{s}_2$, it follows that

$p_2 = 1 - F_2(s^0) + \frac{s^0 - \underline{s}_2}{\delta} = 1$, or, equivalently, $m = m^+ = \frac{s^0 - \underline{s}_2}{\delta}$,

which means that a type-2 firm is always correctly identified, while $p_1 < 1$ results. (From the lower and upper bound for Δ, it follows directly that $m^- \leq m \leq m^+$ must hold.)

Thus, the higher m, the more precise signal $\hat{s}_2$ becomes in identifying a type-2 firm (weak ICS), but at the same time the less precise signal $\hat{s}_1$ becomes in identifying a type-1 firm (strong ICS), and, vice versa. Accordingly, if the strong ICS is identified with certainty, $m = m^-$,

$p_1 = n - m^- = 1 \Leftrightarrow m^- = n - 1$ results,

while, if the weak one is identified with certainty, the principal picks $m = m^+$ such that

$p_2 = n + m^+ = 1 \Leftrightarrow m^+ = 1 - n$ holds.

[10] Doing so ensures that the AI never classifies the firm as a type that arises with zero probability given the observed s.

5.2. Optimal AI specification

With the additional structure in place, we can now proceed to identify optimal AI specifications. We consider two scenarios below. In the first scenario, we focus on the AI specification that minimizes the expected fraud level. Such a specification might be considered as the one that is desirable from a capital market or investors' perspective. In the second, we acknowledge, however, that the specification of the AI is most likely a choice of the auditor. Therefore, we analyse the auditor's incentives to pick a particular AI specification.

5.2.1. AI specification to minimize expected fraud level

Inserting $p_1 = n - m$ and $p_2 = n + m$ we get the expected level of fraud depending on m:

$$E(\gamma^*(m)) = \frac{4c(F+P)\left(n-\frac{1}{2}\right)\left[P\left(n-\frac{1}{2}\right)+x\left(m+(1-\theta)n+\frac{1}{2}(\theta-1)\right)\right]}{Db\left(P(2n-1)-x(1-n-m)\right)\left(P(2n-1)+x(n+m)\right)}.$$

Minimizing $E(\gamma^*(m))$ with respect to m and solving the optimality condition $\frac{dE(\gamma^*)}{dm} = 0$, we obtain two candidates for local extrema of $E(\gamma^*)$:

$$m_1 = \frac{\sqrt{h_1 h_2} + (1-2n)\left(P + x(1-\theta)\right)}{2x},$$

$$m_2 = \frac{-\sqrt{h_1 h_2} + (1-2n)\left(P + x(1-\theta)\right)}{2x},$$

with $h_1 = h + x$, $h_2 = h - x$ and $h = (2n-1)(P + x\theta) > 0$.

For further analysis we define two critical values for x:

$$\bar{x} = \frac{-P(2\theta-1)(2n-1)}{\theta(2n-1)-1},$$

$$\bar{\bar{x}} = \frac{-P(2n-1)}{\theta(2n-1)-1},$$

with $\bar{\bar{x}} > \bar{x}$.

Lemma 2:

(i) If $x < \bar{\bar{x}}$, the only stationary point of $E\big(\gamma^*(m)\big)$ within the feasible range is a local maximum. Accordingly, the minimum expected fraud is obtained at either the left or the right boundary of m.

(ii) If $x \geq \bar{\bar{x}}$, $\frac{dE(\gamma^*)}{dm} < 0$ for all (feasible) m. It follows that the expected fraud level is minimized at the upper bound of m.

In order to figure out whether expected fraud for $x < \bar{\bar{x}}$ is minimized at m^+ or m^-, recall from above that $m^+ = 1 - n$ and $m^- = n - 1$. Inserting these expressions into $\Delta E(\gamma^*) = E(\gamma^*|m^+) - E(\gamma^*|m^-)$ results in

$$\Delta E(\gamma^*) = \frac{2c(F+P)x^2(1-n)[P(2\theta-1)(2n-1) + x(\theta(2n-1)-1)]}{bDP(P(2n-1)+x)(P+x)\big(P(2n-1)+2x(n-1)\big)}.$$

Interestingly, Lemma 2 shows that it is never optimal to pick a neutral specification of the AI when the goal is to minimize expected fraud probability. Moreover, from Lemma 2 (ii) we observe that it is optimal to focus on the weak ICS, whenever the cost of committing fraud in the presence of the strong system is sufficiently high, that is $x \geq \bar{\bar{x}}$ holds. If x is below $\bar{\bar{x}}$, we observe from (i) that it might either be optimal to focus on the strong system or on the weak one. Which alternative is optimal once again critically depends on the cost of fraud, x, but also on the probability of types, θ. This is stated in Proposition 3.

Proposition 3:

In order to minimize the expected fraud probability $E(\gamma^*(m))$, m^* needs to be chosen as follows.

$$m^* = \begin{cases} m^- = n-1 & \text{if } \theta > \frac{1}{2} \text{ and } x < \bar{x} \\ m^+ = 1-n & else \end{cases}$$

Proposition 3 states that $m^* = m^-$ is optimal under specific conditions. Otherwise $m^* = m^+$ is optimal.

Recall that $m^* = m^-$ implies that the AI tool is specified in such a way that it detects the strong system with certainty. This is achieved at the cost of reducing the probability to identify the weak system correctly. According to Proposition 3, doing so is only optimal if the strong system is a) more likely than

the weak one and b) the additional cost of committing fraud under a strong system is sufficiently low. Both aspects combined imply that the auditor not only expects a strong ICS to be present but also that management fraud comes with it. It is therefore optimal to specify the AI accordingly.

In the absence of either a) or b), in contrast, it is optimal to specify the AI to detect the weak ICS with certainty and accepting that a strong ICS is classified as weak with positive probability.

Intuitively, if the extra cost of fraud x is sufficiently high, the auditor expects the probability of fraud to be present much smaller if the ICS is strong than if it is weak. It is therefore important to identify the weak system with certainty to ensure a proper audit, whenever the signal indicates a large audit risk. Moreover, classifying a low-risk client erroneously as the high-risk type, results in overly high audit effort for the type 1 client. Anticipating the possibility of an intense audit as a result of misclassification, the manager of a strong ICS firm is further discouraged from committing fraud. It follows that focusing on the weak type is the optimal strategy in order to minimize expected fraud. However, even if the extra cost x is not particularly high, it suffices that the weak ICS is more likely than the strong one, to render $m^* = m^+$.

5.2.2. AI specification to maximize auditor utility

From the previous section it turns out that in order to minimize the probability for fraud it is optimal in many scenarios to focus on the weak ICS. This implies that the AI produces the signal $\hat{s}_2$ not only if the ICS is weak, but also with positive probability if it is strong. Observation of $\hat{s}_2$, as described above, triggers a large audit effort and in turn reduces incentives for fraud. Such an effort intensive approach, however, might not be in the best interest of the auditor. Rather than to minimize fraud probability, the auditor is interested in maximizing the expected payoff from the audit which, in equilibrium, equals

$$E(\Pi^{A*}) = -c[\Pr(\hat{s}_1) \cdot a_1^* + \Pr(\hat{s}_2) \cdot a_2^*] - D \cdot \Pr(\text{Fraud undetected with AI}). \quad (13)$$

From Proposition 2 (iii) we already know that $\Pr(\text{Fraud undetected with AI})$ is constant and therefore unaffected by any specification of m (it is equal to $\frac{c}{bD}$). The first term in (13) is the auditor's loss from conducting effort.

Thus, the auditor picks m in order to minimize his effort costs

$$\text{EC}_{\text{effort}} = c[\Pr(\hat{s}_1) \cdot a_1^* + \Pr(\hat{s}_2) \cdot a_2^*] .$$

Note that $\Pr(\hat{s}_2) = 1 - \Pr(\hat{s}_1)$ and that observation of a signal $\hat{s}_i$ always triggers the corresponding effort choice a_i^* from the auditor, i.e., $\Pr(\hat{s}_i) = \Pr(a = a_i^*)$. The equilibrium values of $\Pr(\hat{s}_i)$ and a_i^* have been derived in Proposition 2 and depend on m a via p_1 and p_2. Using these equilibrium values, expected effort costs can be written as

$$\mathrm{EC}_{\text{effort}} = \frac{c}{b}[\omega_1^* \ln(y_1^*) + (1 - \omega_1^*)\ln(y_2^*)]$$

with

$a_1^* = \frac{1}{b}\ln(y_1^*)$ and $y_1^* = \frac{(F+P)(2n-1)}{P(2n-1)+x(m+n)}$; $a_2^* = \frac{1}{b}\ln(y_2^*)$ and $y_2^* = \frac{(F+P)(2n-1)}{P(2n-1)+x(m+n-1)}$;

$\omega_1^* = \Pr(\hat{s}_1) = \Pr(a = a_1^*) = 1 - m - n - \theta(1 - 2n)$.

Thus, the auditor's optimization problem with regard to m is

$$\min_m \mathrm{EC}_{\text{effort}} = \frac{c}{b}[\omega_1^* \ln(y_1^*) + (1 - \omega_1^*)\ln(y_2^*)]$$

(OP)

$$\text{subject to } m^- \leq m \leq m^+.$$

Minimizing the expected effort costs, the auditor faces a trade-off between the levels of effort and their probabilities of occurrence. If she increases m by a marginal unit, both efforts, a_1^* and a_2^*, decrease; with the reduction being stronger for effort a_2^*. At the same time, an increase in m decreases the probability that a_1^* occurs, ω_1^*, and accordingly increases the probability that a_2^* occurs, $1 - \omega_1^*$. Thus, the optimal choice of m trades off the reduction in both efforts against the increase in the probability of a_2^* (and the corresponding decrease in the probability of a_1^*), i.e., the effort for which the reduction caused by an increase in m is stronger.

To get the intuition for the above results, recall that an increase in m is equivalent to specifying the AI in such a way that the probability of correctly identifying a weak ICS increases. The probability of correctly identifying a strong ICS in turn decreases. In terms of our representation in section 4 this implies that the critical value s^c decreases and $\hat{s}_2$ ($\hat{s}_1$) is reported for a larger (smaller) range of realizations. Accordingly, the unconditional signal probability $\Pr(\hat{s}_2)$ increases and $\Pr(\hat{s}_1)$ decreases.

With regard to the auditor's effort, we observe that it decreases if the focus shifts towards the weak ICS, no matter which signal is observed. Intuitively the increase in m renders $\hat{s}_1$ a more reliable signal, indicating strongly that the strong ICS is indeed present. At the extreme, it even holds that $\lim_{m \to m^+} \Pr(t_1|\hat{s}_1) = 1$. If the auditor is confident that the ICS is indeed strong, however, she reduces her effort a_1^*.

With respect to $\hat{s}_2$ the effect is somewhat reversed: Shifting the AI focus towards the weak system implies that not only weak ICSs are detected but also strong ones are erroneously classified as weak. From the auditor's perspective, this renders the signal $\hat{s}_2$ less credible. Formally, this is reflected in the fact that $\Pr(t_1|\hat{s}_2)$ increases in m. As a consequence, the auditor cannot be so sure anymore that a firm's ICS is really weak and requires intense audit after $\hat{s}_2$ is reported. Thus, she reduces a_2^* in equilibrium.

As the optimization program (OP) is different from minimizing expected fraud (see Proposition 3), the optimal solution for m in general does not coincide with the fraud-minimizing level of m.

We state this in Proposition 4 and provide a numerical example below.

Proposition 4:

The level of m that minimizes the expected probability of fraud does not necessarily coincide with the level of m that minimizes the auditor's cost. Accordingly, the auditor may pick m differently from the socially desired level.

To demonstrate the above result, we consider the following example with the parameter values

$$F = P = D = 10;\ \theta = 0.8;\ n = 0.7;\ x = 3.6;\ c = 2;\ b = 1.$$

As becomes evident from Figure 3, EC_{effort} has no local minimum; the optimal solution is a corner solution as in the case of minimizing expected fraud (Figure 4). However, whereas expected fraud is minimized at $m = m^+$, expected auditor effort cost are minimized at $m = m^-$, see Table 1.

	$m^+ = 1 - n = 0.3$	$m^- = n - 1 = -0.3$
γ_1^*	0.2105	0.2078
γ_2^*	0.8547	0.8696
a_1^*	0.0513	0.3857
a_2^*	0.6931	1.4697
ω_1^*	0.32	0.92
EC_{effort}	0.9755	0.9448*
$E(\gamma^*)$	0.3394*	0.3402

Table 1: Equilibrium values for the numerical example

In the example, we have $\bar{x} = 3.53$ and thus $x > \overline{x}$. It follows directly from Proposition 3 that $m = m^{+}$.

In minimizing expected auditor effort, in contrast, the corresponding optimal AI specification requires maximizing the precision of signal $\hat{s}_1$, i.e., $m = m^{-}$. Note that switching from a maximum precise signal $\hat{s}_2$ to a maximum precise signal $\hat{s}_1$ increases effort levels a_1^* and a_2^* considerably. However, this increase in effort levels is more than offset by the sharp decline in the probability of a_2^*, $1 - \omega_1^*$, from 0.68 to 0.08.

Given this potential conflict of interest, it might be desirable to implement appropriate incentives or regulatory measures to induce the auditor to align the AI specification with the objective of fraud minimization. One possible measure is to require the auditor to disclose the AI specification process. However, the problem is that an external party is unlikely to possess sufficient (firm-specific) information to assess whether the auditor has specified the AI in the "right" way.

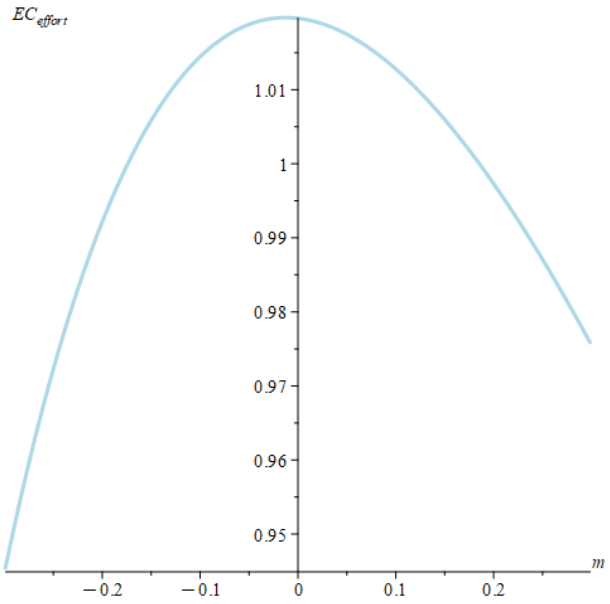

Figure 3: Expected effort costs depending on m

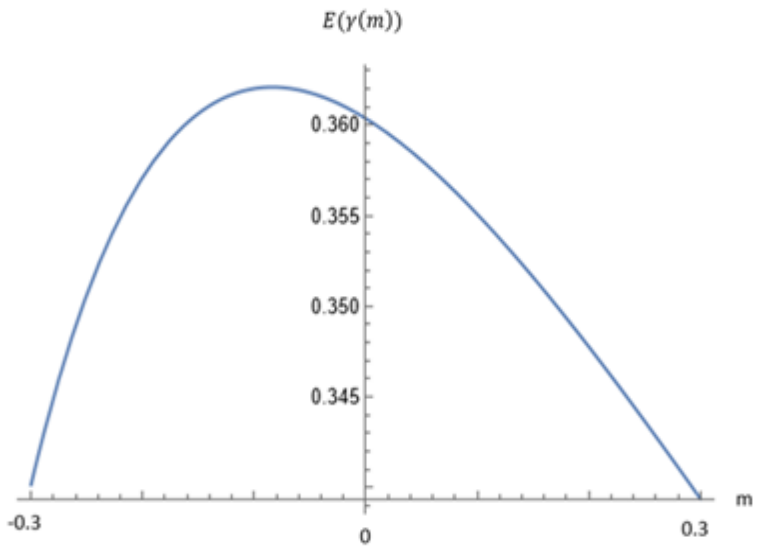

Figure 4: Expected probability of fraud depending on

6. Conclusion

In this paper we consider a game-theoretic interaction between an auditor and a client firm. We assume that the auditor uses an AI tool as part of the auditing process. Specifically, the auditor uses AI to detect possible weaknesses in the ICS of the client firm and in turn revises her expectations regarding incentives for managerial fraud. Based on the results provided by the AI, the auditor plans her personal audit effort. We assume that the information provided by the AI is informative and that the auditor has the option to customize the AI according to her needs.

While there is considerable hope that AI tools help to improve audit efficiency and audit effectiveness, our findings substantiate this hope only partially (at best). We find that part of the auditor's effort can be replaced by AI, which can be interpreted as an increase in efficiency. We also find that AI use reduces ex ante incentives for managerial fraud. This is probably beneficial as we can assume that any fraud, whether detected or not, is costly for firms or for capital markets. However, the reduction in auditor effort in our model goes along with a reduced probability of detecting actual fraud. It is in that sense, that audit effectiveness, or audit quality, decreases rather than to increase. Moreover, AI is unable to decrease the ex ante probability that fraud remains undetected, such that reliability of audit

opinions is not improved. Besides, we state that the auditor cannot be expected to customize the AI in the best interest of stakeholders but rather maximizes her own payoff.

While we believe that our model provides some insights into potential effects of AI use, we also realize that we neglect some potentially critical aspects in our parsimonious model. In particular we allow the auditor to "customize" or "bias" the AI tool according to her needs, but we do not allow the client firm to "customize" the data provided in order to mislead the AI and to keep it from detecting a weak ICS. Moreover, we do not address in which way the AI tool detects accounting anomalies and what is considered a normal pattern as opposed to abnormal patterns. Thus, problems regarding the training data, that are frequently considered important, are ignored in our study.

Appendix

Proof of Proposition 1:

According to our regularity conditions, $c(F+P) < bDP(1-\theta)$ holds. Consider a mixed-strategy for the manager where at least one type randomizes with respect to committing fraud.

The first order condition for optimal auditor effort derived from (1) equals:

$$\frac{dE(\Pi^A)}{da} = (-\theta\gamma_1 - (1-\theta)\gamma_2)bD\exp[-ba] - c = 0. \qquad (4)$$

If both manager types randomize, they must be indifferent between committing fraud and not to do so:

$$F\exp[-ba] - P(1-\exp[-ba]) - x = 0 \qquad (5)$$

$$F\exp[-ba] - P(1-\exp[-ba]) = 0. \qquad (6)$$

Note that (5) and (6) cannot hold simultaneously. If (6) holds, the LHS of (5) is negative.

We therefore distinguish two cases:

Case 1: We assume that (6) holds in equilibrium, implying that (5) is strictly negative. In that case the manager randomizes over committing fraud if the ICS is weak and optimally refrains from committing fraud when the ICS is strong.

Solving (4) and (6) for a and γ_2, given that $\gamma_1 = 0$, we obtain

$\gamma_2^{BM} = \frac{c(F+P)}{(1-\theta)bDP}$ and $a^{BM} = \frac{1}{b}\ln\left(\frac{F+P}{P}\right)$.

Case 2: We assume that (5) holds in equilibrium implying that (6) is strictly positive. In that case the manager randomizes over committing fraud if the ICS is strong and optimally commits fraud with certainty when the ICS is weak.

Solving (4) and (5) for a and γ_1, given that $\gamma_2 = 1$, we obtain

$\gamma_1^{BM} = \frac{\frac{c(F+P)}{bD(P+x)} - (1-\theta)}{\theta}$ and $a^{BM} = \frac{1}{b}\ln\left(\frac{F+P}{P+x}\right)$.

Note, however, that we assumed that $c(F+P) < bDP(1-\theta)$ holds, implying that $\frac{c(F+P)}{bDP} < 1-\theta$.

As $\frac{c(F+P)}{bD(P+x)} < \frac{c(F+P)}{bDP} < 1-\theta$ this implies that $\gamma_1^{BM} < 0$ which is impossible.

Accordingly, the equilibrium derived in case 1 is the only feasible solution.

■

Proof of Proposition 2:

(i) $\Delta_\gamma = E(\gamma^*) - E(\gamma^{BM}) = -\frac{c(F+P)x[P\theta(p_1+p_2-1)^2 - xp_2(1-p_2)]}{DbP(P(p_1+p_2-1)+p_2x)(P(p_1+p_2-1)-x(1-p_2))} \leq 0.$

(ii) Define $a^{BM} = \frac{1}{b}\ln(y^{BM})$ with $y^{BM} = \frac{F+P}{P}$

and $a_1^* = \frac{1}{b}\ln(y_1^*)$ with $y_1^* = \frac{(F+P)(p_1+p_2-1)}{P(p_1+p_2-1)+p_2x}$,

$a_2^* = \frac{1}{b}\ln(y_2^*)$ with $y_2^* = \frac{(F+P)(p_1+p_2-1)}{P(p_1+p_2-1)-x(1-p_2)}$.

Then the expected auditor effort under AI is

$$E(a^*) = \Pr(\hat{s}_1)\frac{1}{b}\ln(y_1^*) + \Pr(\hat{s}_2)\frac{1}{b}\ln(y_2^*)$$

From Jensen's inequality it follows that $\frac{1}{b}\ln\big(E(y^*)\big) > \frac{1}{b}E(\ln(y^*)) = E(a^*)$.

Since $a^{BM} = \frac{1}{b}\ln(y^{BM}) \geq \frac{1}{b}\ln\big(E(y^*)\big)$, it follows $a^{BM} > E(a^*)$.

(iii) $\Pr(\text{undetected fraud with AI}|\text{fraud is committed in } t_1) = (p_1\exp[-ba_1] + (1-p_1)\exp[-ba_2]) = \frac{P+x}{P+F}$.

As no fraud is committed with the strong ICS in the benchmark setting, the conditional probability for it to be undetected is zero. As $\frac{P+x}{P+F} > 0$ it increases with AI.

$\Pr(\text{undetected fraud with AI}|\text{fraud is committed in } t_2) = (p_2 \exp[-ba_2] + (1 - p_2)\exp[-ba_2]) = \frac{P}{P+F}$, which is equivalent to $\exp[-ba^{BM}]$.

It follows directly that $\theta \frac{P+x}{P+F} + (1-\theta)\frac{P}{P+F} > (1-\theta)\frac{P}{P+F}$.

■

(iv) $\Pr(\text{undetected fraud with AI}) = \theta\gamma_1(p_1 \exp[-ba_1] + (1-p_1)\exp[-ba_2]) + (1-\theta)\gamma_2(p_2 \exp[-ba_2] + (1-p_2)\exp[-ba_1]) = \frac{c}{bD}$.

■

Proof of Lemma 2:

Note that our previous assumption of signal informativeness, specifically assuming that $p_1 + p_2 > 1$, now implies that $2n > 1 \Leftrightarrow n > 0.5$.

(i) Assume $x < \bar{\bar{x}} \Leftrightarrow h_2 > 0$.

From

$$\frac{dE(\gamma^*)}{dm} = \frac{-8c(F+P)\left(n-\frac{1}{2}\right)x^2[x\left(n^2\left(\frac{1}{2}-\theta\right)+n\left((1-\theta)m+\theta-\frac{1}{2}\right)+\frac{1}{4}+\frac{m^2}{2}+\frac{m(\theta-1)}{2}-\frac{\theta}{4}\right)+P(n(1-2\theta)+m+\theta-\frac{1}{2})(n-\frac{1}{2})]}{Db\left(P(2n-1)-x(1-n-m)\right)^2\left(P(2n-1)+x(n+m)\right)^2} = 0,$$

we derive the solutions for m_1 and m_2 as given above.

From $\frac{d^2E(\gamma^*)}{dm^2}(m_1) = -\frac{1024(F+P)c\left(n-\frac{1}{2}\right)\frac{x^2}{4}h_1h_2\left(\frac{\sqrt{h_1h_2})}{2}+(P+x\theta)\left(n-\frac{1}{2}\right)\right)}{Db\left(\sqrt{h_1h_2}+h_2\right)^3\left(\sqrt{h_1h_2}+h_1\right)^3} < 0$, we conclude that m_1 is a local maximizer of $E(\gamma^*)$.

Notice that $m_2 < 0$. We now show that m_2 violates the condition $n - m_2 \leq 1$:

$$n - m_2 = \frac{\sqrt{h_1h_2} + P(2n-1) + x(\theta-1) + 2xn(2-\theta)}{2x} \leq 1$$

$$\Leftrightarrow$$

$$\sqrt{h_1 h_2} + P(2n-1) + x\big(\theta - 1 + 2(2n - n\theta - 1)\big) \leq 0. \qquad (11)$$

By assumption $h_2 = (2n-1)(P + x\theta) - x = P(2n-1) + x(-1 - \theta + 2\theta n) > 0$. We now show that $h_2 > 0$ implies $P(2n-1) + x\big(\theta - 1 + 2(2n - n\theta - 1)\big) > 0$ and in turn (11) is violated. To demonstrate this, we show that $P(2n-1) + x\big(\theta - 1 + 2(2n - n\theta - 1)\big) > h_2$:

$$P(2n-1) + x\big(\theta - 1 + 2(2n - n\theta - 1)\big) > P(2n-1) + x(-1 - \theta + 2\theta n)$$

$$\Leftrightarrow$$

$$x\big(\theta - 1 + 2(2n - n\theta - 1)\big) > x(-1 - \theta + 2\theta n)$$

$$\Leftrightarrow$$

$$\theta - 1 + 2(2n - n\theta - 1) > -1 - \theta + 2\theta n$$

$$\Leftrightarrow$$

$$4n - 2 > 4\theta n - 2\theta$$

$$\Leftrightarrow$$

$$4n - 2 > \theta(4n - 2),$$

which is a true statement. It follows that (11) is violated and m_1 is the only stationary point in the feasible range.

(ii) Assume $\geq \bar{\bar{x}} \Leftrightarrow h_2 \leq 0$.

Notice that $\frac{dE(\gamma^*)}{dm}$ can also be written as

$$\frac{dE(\gamma^*)}{dm} = \frac{-8c(F+P)\left(n - \frac{1}{2}\right)x^2\left[\frac{x}{2}m^2 + \mu m + v\right]}{Db\big(P(2n-1) - x(1-n-m)\big)^2\big(P(2n-1) + x(n+m)\big)^2},$$

with $\mu = \frac{(2n-1)\big(P + x(1-\theta)\big)}{2} > 0,$

$$v = \frac{1}{2}(2\theta - 1)(2P + x)n(1-n) + \frac{1}{4}\big(P(1-2\theta) + x(1-\theta)\big),$$

such that the two solutions of $\frac{dE(\gamma^*)}{dm} = 0$ can be written as

$$m_1 = \frac{\sqrt{\mu^2 - 2xv} - \mu}{x}, m_2 = \frac{-\sqrt{\mu^2 - 2xv} - \mu}{x},$$

where $\mu^2 - 2xv = \frac{h_1 h_2}{4}$.

$\mu^2 - 2xv < 0$ ($\Leftrightarrow h_2 < 0$) implies that the term $\frac{x}{2}m^2 + \mu m + v$ is always strictly positive, which, in turn, implies that $\frac{dE(\gamma^*)}{dm} < 0$. Thus, the fraud-minimizing value of m is at the highest possible value of it: $m^* = m^+ = 1 - n$.

Assume now $h_2 = 0$. In this case we have a double null, $m_1 = m_2 = \frac{(1-2n)(P+x(1-\theta))}{2x} < 0$. The term $\frac{x}{2}m^2 + \mu m + v$ is zero for $m = \frac{(1-2n)(P+x(1-\theta))}{2x}$ and strictly positive otherwise (upward-opening parabola). However, $m = \frac{(1-2n)(P+x(1-\theta))}{2x}$ violates the condition $n - m \leq 1$. Thus, $\frac{dE(\gamma^*)}{dm} < 0$ in the feasible range, which implies again that the fraud-minimizing value of m is at the highest possible value of it: $m^* = m^+ = 1 - n$.

■

Proof of Proposition 3:

Note that

$$\Delta E(\gamma^*) = E(\gamma^*|m^+) - E(\gamma^*|m^-) = \frac{2c(F+P)x^2(1-n)[P(2\theta-1)(2n-1)+x(\theta(2n-1)-1)]}{bDP(P(2n-1)+x)(P+x)(P(2n-1)+2x(n-1))}. \quad (12)$$

(12) is positive if and only if $\theta > \frac{1}{2}$ and $x < \frac{-P(2\theta-1)(2n-1)}{\theta(2n-1)-1} = \bar{x}$, implying that $m^* = m^- = n - 1$ in these cases.

In all other cases, (12) is negative. Since, according to Lemma 2, the optimal value for m for $x < \bar{\bar{x}}$ lies at the right (m^+) or left (m^-) boundary of m, and for $x \geq \bar{\bar{x}}$ it is always at the right boundary, it follows that m^*=$m^+ = 1 - n$ in all remaining cases.

■

Literature

Blake, O. (2024): The Eroding Trust in Audits: Confronting a Crisis of Confidence, Retrieved November 10, 2025, https://earmarkcpe.com/the-eroding-trust-in-audits-confronting-a-crisis-of-confidence/.

Eisikovits, N., Johnson, W., and Markelevich, A. (2024): Should Accountants be Afraid of AI? Risks and Opportunities of Incorporating Artificial Intelligence into Accounting and Auditing, Working Paper.

Issa, H., Sun, T., and Vasarhelyi, M. A. (2016). Research Ideas for Artificial Intelligence in Auditing: The Formalization of Audit and Workforce Supplementation, *Journal of Emerging Technologies in Accounting*, 13, 1–20.

Kokina, J., and Davenport, T. H. (2017). The Emergence of Artificial Intelligence: How Automation is Changing Auditing, *Journal of Emerging Technologies in Accounting*, 14, 115–122.

Kwon, Y.K.(2005): Accounting Conservatism and Managerial Incentives, *Management Science*, 51, 1626-1632.

Law, K.K., and Shen, M. (2024): How does Artificial Intelligence Shape Audit Firms?, Management Science, Acticle in Advance, pp.1-26, ISSN 1526-5501.

Smith, J.R., Tiras, S.L., and Vichitlekarn, S.S. (2000): The Interaction between Internal Control Assessment and Substantive Testing in Audits for Fraud, *Contemporary Accounting Research*, 17, 327-356.

Webb, M. (2020): The impact of Artificial Intelligence on the Labor Market, Working Paper, Stanford University

Wirtschaftsprüferkammer (2025): KI Fragen und Antworten zum Einsatz von künstlicher Intelligenz in der WP-Praxis, Retrieved November 12, 2025, https://www.wpk.de/fileadmin/documents/Wissen/KI/WPK_Fragen_Antworten_Einsatz_KI_WP-Praxis_21-07-2025.pdf.

Otto von Guericke University Magdeburg
Faculty of Economics and Management
P.O. Box 4120 | 39016 Magdeburg | Germany

Tel.: +49 (0) 3 91 / 67-1 85 84
Fax: +49 (0) 3 91 / 67-1 21 20

www.fww.ovgu.de/femm

ISSN 1615-4274