Mitigating the Dilution Effect of Non-diagnostic Information on Auditors' Judgments Using a Frequency Response Mode

Aasmund Eilifsen
Norwegian School of Economics

Natalia Kochetova
Saint Mary's University

William F. Messier, Jr.
Norwegian School of Economics

SAINT MARY'S UNIVERSITY
SOBEY
SCHOOL OF BUSINESS

- Motivation/Theory
- Method
- Findings
- Implications

AGENDA

Motivation/Theory

SAINT MARY'S UNIVERSITY
SOBEY
SCHOOL OF BUSINESS

Non-diagnostic or irrelevant data: it is everywhere!

- Substantial amount of research in various content areas (psychology, law, and marketing) that shows that individual judgments are affected by irrelevant ("non-diagnostic") information or evidence.
- Basic findings of this line of research: the presence of non-diagnostic evidence leads to a *dilution effect*; that is, individuals make less extreme (more regressive) decisions than those in the presence of diagnostic evidence only.
- Attention to irrelevant information has the potential to significantly limit the potential value from incorporating Big Data into the audit process (Brown-Liburd et al. 2015).

Motivation/Theory

SAINT MARY'S UNIVERSITY
SOBEY
SCHOOL OF BUSINESS

What is dilution effect?

- The information generated by Big Data is largely ambiguous, unstructured, voluminous, and represents a mix of relevant/diagnostic and irrelevant/non-diagnostic - all of these characteristics affect auditor judgments negatively.
- Several major studies in auditing addressing the issue of dilution effect of non-diagnostic evidence on auditor judgement: Hackenbrack (1992), Hoffman and Patton (1997), Glover (1997), Shelton (2008).
- In summary, auditors, similar to other humans, are unable to discount irrelevant/non-diagnostic information in making probabilistic judgements and in other JDM tasks.

Motivation/Theory

What is dilution effect?

- Hackenbrack [1992] assessed how much a company's exposure to fraudulent reporting changed when presented with *a mixture of diagnostic and non-diagnostic evidence*: the auditors' fraud risk assessments became less extreme in the presence of non-diagnostic evidence.
- Hoffman and Patton [1997] and Glover [1997] examined whether accountability and time pressure eliminated or mitigated the dilution effect.
- Hoffman and Patton [1997] report, “auditors' judgments exhibited the dilution effect both when they were held accountable and when they were not (p. 228).”

Motivation/Theory

What is dilution effect?

- Glover [1997]: accountability had no effect on the dilution effect; however, time pressure reduced the dilution effect, although it did not eliminate it.
- Shelton [2008]: audit managers and partners are less susceptible to the dilution effect than senior auditors.
- Assuming perceptual approach of dilution effect as in prior auditing studies, we continue to ask:
  - *How can dilution effect in auditor judgment be ameliorated?*

Motivation/Theory

What is dilution effect?

- Detecting financial reporting fraud continues to be a priority (PCAOB 2018).
- To improve auditors' fraud judgments, firms increasingly rely on Big Data and data analytics (FRC 2017).
- *Can dilution of fraud risk assessments can be reduced using a frequency mode in situations where diagnostic and non-diagnostic or irrelevant information supplements the output from a fraudulent client profile analytics?*

Motivation/Theory

What is frequency argument?

- Kochetova-Kozloski, Messier, and Eilifsen (KME) (2011): statistical reasoning within a Bayesian framework can be improved, especially in low base rate events (i.e., fraud): the auditors' fraud judgments using a frequency response mode, as compared to a probability response mode, are *closer* to the Bayesian benchmark.
- Gigerenzer and his colleagues (e.g., Gigerenzer, Hoffrage, and Kleinbolting 1991; Gigerenzer and Hoffrage 1995) and others (Cosmides and Tooby 1994, 1996): if people are asked to estimate the *probability* of a single event, the question does not connect to probability theory in their minds, whereas the *frequency* of such an event does (Gigerenzer and Goldstein 1996; Gigerenzer 2004).

Motivation/Theory

SAINT MARY'S UNIVERSITY
SOBEY
SCHOOL OF BUSINESS

What is frequency argument?

- Bayesian computations are *cognitively simpler* when information is encoded in a frequency format rather than in a probability format.
- The estimation of the likelihood of a single event and the judgment of frequency are *cognitively different processes* (Cosmides and Tooby 1994, 1996; Gigerenzer et al. 1991). Based on KME's findings, H1:
- *H1: Auditors demonstrate a lower dilution effect when they receive case information and make required judgments in a frequency response mode as compared to a probability response mode.*

Motivation/Theory

SAINT MARY'S UNIVERSITY
SOBEY
SCHOOL OF BUSINESS

Types of non-diagnostic evidence

- As in Hackenbrack (1992), three types : favorable, unfavorable, and neutral. In the fraud-risk setting:
- *Favorable* non-diagnostic evidence would be information that does not relate directly to possible fraud but may be viewed as positive by the auditor.
- *Unfavorable* non-diagnostic evidence describes negative client information that is not directly related to the presence of client fraud but might be viewed by the auditor as negative.
- *Neutral* non-diagnostic evidence includes information that is neither positive nor negative and evaluated as unrelated to the presence of client fraud by the auditor.

Motivation/Theory

Types of non-diagnostic evidence

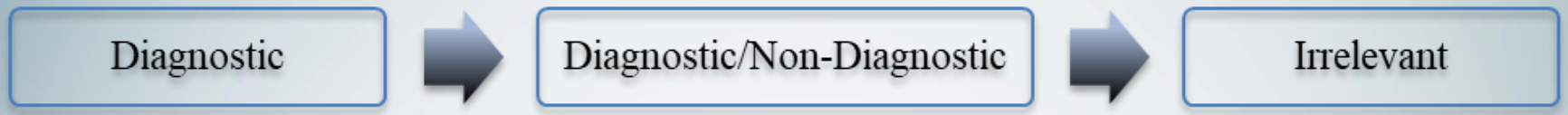
- Hackenbrack's (1992) H: non-neutral (favorable and unfavorable combined) non-diagnostic evidence has a higher dilutive capacity than neutral non-diagnostic evidence:
  - non-neutral, non-diagnostic evidence is more *salient* and auditors will devote more attention to such evidence (e.g., Tversky 1977;
  - Hackenbrack (1992): mixed results across the two versions of the task (increasing versus decreasing fraud risk);
  - Hoffman and Paton (1997) distinguish between favorable and unfavorable non-diagnostic information but find no differences in their dilutive effect.
- Literature in psychology: neutral non-diagnostic evidence is more likely to be ignored than non-neutral (e.g., LaBella and Koehler 2004).

Motivation/Theory

Types of non-diagnostic evidence

- *RQ*: *In a frequency response mode, do auditors exhibit the dilution effect differentially across the different types of non-diagnostic/irrelevant evidence?*

Continuum of Evidence Relevance/Diagnosticity



- *Diagnostic*: information that is clearly relevant to the specific fraud event; i.e., it is a robust "red flag" indicating increased likelihood of fraud; e.g. fraud risk factors identified by Bell and Carcello (2000) (and those clearly rated by our experts).
- Diagnostic/non-diagnostic: e.g. there are many fraud-related factors in auditing standards that auditors believe to be diagnostic - but which are not (e.g., see Hogan et al. 2008; Trompeter et al. 2014; Bell and Carcello 2000).
- *Irrelevant:* has not predictive ability or association with event being judged .

Method: 2 Experiments

Participants

- Norwegian auditors in NHH MRR program
- A mix of senior auditors, staff or associates, and managers
- Some had a master's degree, while all had a bachelor's degree
- All participants either had or were in the process of obtaining a professional designation
- The majority of the participants worked for a Big 4 firm at the time of the experiment
- Experiment 2 participants were, on average, more experienced than Experiment 1
- Paper and pencil vs. *Qulatrics* administration

Method: 2 Experiments

Design

- *Experiment 1:*
  - 2 (Response Mode) x 3 (Type of Non-diagnostic Evidence) x 2 (Order) between-participants
  - Response Mode (RM) at two levels: frequency response mode vs. probability response mode;
  - Type of Non-diagnostic Evidence (TYPE-EV) at three levels: neutral, favorable, and unfavorable; and
  - Order (ORDER) of the non-diagnostic evidence cues at two levels.

Method: 2 Experiments

SAINT MARY'S UNIVERSITY
SOBEY
SCHOOL OF BUSINESS

Design

- *Experiment 2:*
  - 2 (Response Mode) x 3 (Type of Irrelevant Evidence) Response Mode (RM) at two levels: frequency response mode vs. probability response mode;
  - Type of Irrelevant Evidence (TYPE-EV) at two levels: favorable and unfavorable; and
  - Order (ORDER) of the non-diagnostic evidence cues was randomized in *Qulatrics*

Method

Procedure: Experiment 1

- Expert panel evaluated 41 fraud risk factors: see Appendix A.
- We selected 3 diagnostic factors and three each of neutral, favorable, and unfavorable non-diagnostic factors: see Table 1 for selected factors (cues).
- Same case materials as KME except: presented 3 pieces of diagnostic evidence and then 3 pieces of either neutral, favorable, and unfavorable non-diagnostic factors.
- This approach follows a belief revision procedure followed by LaBella and Koehler [2004].
- Auditors were asked to rate the fraud risk factors in the same manner as the expert managers.
- Participants were asked a series of demographic questions.

Method

Procedure: Experiment 2

- Used Hoffman and Patton (1997) irrelevant cues: 3 favorable and 3 unfavorable.
- Same case materials as KME except: presented 3 pieces of diagnostic evidence and then 3 pieces of either favorable, or unfavorable irrelevant cues.
- Otherwise similar to Experiment 1.

- *Note: an* alternative approach would have been to "bundle" diagnostic and non-diagnostic cues (Fanning et al. 2015; Lambert and Peytcheva 2017) vs. our "step-by-step," sequential, approach.

Method

Dependent Variables

- Replication of KME: *F-DEV* = |Auditor's Fraud Response – Fraud Bayesian Response|.
- Tests of H1 and RQ:
- *F-ABSREV* = |Auditor's Fraud Response: diagnostic evidence only – Auditor's Fraud Response: added non-diagnostic evidence| and
- *F-REV* = Auditor's Fraud Response: diagnostic evidence only – Auditor's Fraud Response: added non-diagnostic evidence.
- Ps agreement with Expert Panel in Experiment 1: a reasonable level of agreement but we also conducted sensitivity analyses.

Findings

SAINT MARY'S UNIVERSITY
SOBEY
SCHOOL OF BUSINESS

Replication of KME

- The Bayesian benchmarks for the frequency and probability response modes are 0.0776 and 0.0767, respectively (KME [2011, p. 846]).
- *Experiment 1:* For the low base rate (1%), the absolute deviations from the Bayesian benchmark are *smaller in the frequency response mode* (marginal mean =0.262) than in the probability response mode (marginal mean =0.423) (F = 7.504, p =0.004, one-tailed, not tabled).
- Note: this mean is significantly different from zero (t=7.190, p=.000, two-tailed), i.e. the participants *still show significant base rate neglect*. This result is also consistent with KME [2011, p. 853].
- Did the same for Experiment 2.

Findings

Tests of H1

- *Experiment 1*: Tables 2 and 3 - main analyses (n=174); Table 4- sensitivity analyses on reduce sample (n=108)
- *Experiment 2*: Table 6 (n=110)
- In both experiments, the use of a frequency response mode only reduced the dilution effect in the presence of favorable non-diagnostic evidence using both specifications of the DV.
- In Experiment 2, we observe “opposite-to-dilution” effect in the cells with unfavorable irrelevant evidence

Findings

Experiment 1: Tests of H1 on Full Sample (*n*=174)

Panel A: Analysis of Variance (*n*=174); Signed Revision (*F-REV*) as a Dependent Variable

Source	SS	df	*F*	*p*–value
Intercept	.559	1	18.683	.000
RM	.172	1	5.728	.018
TYPE-EV	.893	2	14.918	.000
RM x *TYPE-EV*	.195	2	3.249	.041
Error	5.030	168		

Panel B: Analysis of Variance (*n*=174); Absolute Revision (*F-ABSREV*) as a Dependent Variable

Source	SS	df	*F*	*p*–value
Intercept	1.810	1	71.584	.000
RM	.260	1	10.266	.002
TYPE-EV	.191	2	3.780	.025
RM x *TYPE-EV*	.234	2	4.625	.011
Error	4.248	168		

Findings

Experiment 2: Tests of H1

Panel A: Analysis of variance (*n*=110); Signed Revision (*F-REV*) as a Dependent Variable

Source	SS	df	*F*	*p*–value
Intercept	.010	1	12.662	.001
RM	.000	1	.176	.675
TYPE-EV	.006	1	7.922	.006
RM x *TYPE-EV*	.004	1	4.676	.033
Error	.085	106		

Panel B: Analysis of variance (*n*=110); Absolute Revision (*F-ABSREV*) as a Dependent Variable

Source	SS	df	*F*	*p*–value
Intercept	.027	1	37.115	.000
RM	.001	1	.697	.406
TYPE-EV	.000	1	.374	.542
RM x *TYPE-EV*	.001	1	.995	.321
Error	.076	106		

Findings

Tests of RQ

- RQ: In a frequency response mode, do auditors exhibit the dilution effect differentially across the different types of non-diagnostic evidence?
- *Experiment 1*: signed revisions as a DV(*F-REV*), *TYPE-EV* is significant (p=.026):
  - Statistically significantly different regressive (dilutive) effect of non-diagnostic evidence between conditions with favorable and unfavorable cues, and between neutral and unfavorable cues.
  - Affected by the direction of revision (*F-ABSREV*)- sensitivity to DV specification
- *Experiment 2*: *F-REV* as DV, *TYPE-EV* is significant (p=.001):
  - Dilutive effect of irrelevant evidence is larger for cell with unfavorable cues than for the cell with favorable cues.

Findings

SAINT MARY'S UNIVERSITY
SOBEY
SCHOOL OF BUSINESS

Tests of RQ

- RQ Conclusion: Indeed, auditors are still susceptible to dilution effect in frequency response mode, and differentially so across the different types of non-diagnostic/irrelevant evidence.
- Frequency response mode with irrelevant unfavorable cues appears to increase fraud risk assessments and produce opposite-to-dilution effect ("over-reaction" to negative irrelevant information)

Implications

What does it all mean?

- A simple approach to representing probability information as frequencies to auditors may mitigate a bias in risk assessment that has been shown to be extremely robust to various settings (Hackenbrack [1992]; Hoffman and Patton [1997]; Glover [1997]).
- Our results indicate that while the use of a frequency response mode reduced the dilution effect, this finding is driven by the auditors' responses to cases where non-diagnostic evidence is favorable.
- This is an important finding since clients who are committing fraud are likely to present favorable (non-diagnostic) explanations/evidence to an auditor's inquiry about fraud.

Implications

SAINT MARY'S UNIVERSITY
SOBEY
SCHOOL OF BUSINESS

Where do we go from here?

- Future research should investigate *why* the dilution effect appeared to be unaffected by response mode when non-diagnostic/irrelevant cues were neutral or unfavorable.
- Why did auditors exhibit "opposite-to-dilution" effect in response to unfavorable cues in frequency response mode?
  - Excessive sensitivity to negative information? (Bhattacharjee et al. 2012)
  - Conversational approach to dilution? (Tetlock an Boettger 1989)
- Cue bundling/aggregation design vs. sequential approach: does response mode matter?
- Impact of time pressure, experience, other factors on dilution effect in frequency mode


THANK YOU
THANK YOU
THANK YOU
THANK YOU
THANK YOU
THANK YOU
THANK YOU
THANK YOU
SOBEY
SCHOOL OF BUSINESS
SAINT MARY'S UNIVERSITY
SINCE 1802
One University. One World. Yours.