Client risk calibration in PCAOB audits: an analytical procedures panel risk assignment protocol

Edward J. Lusk*

The State University of New York (SUNY) at Plattsburgh,
101 Broad St., Plattsburgh,
NY 12901, USA
and
Department of Statistics,
The Wharton School,
University of Pennsylvania,
3730 Walnut St.,
Philadelphia, PA, 19104-6340, USA
Fax: +01-518-564-3183
Fax: +01-215-898-6803
Email: luskej@plattsburgh.edu
Email: lusk@wharton.upenn.edu
*Corresponding author

Michael Halperin

The Wharton School,
University of Pennsylvania,
3420 Walnut St.,
Philadelphia, PA, 19104-6340, USA
Fax: +01-215-898-9434
Email: halperin@wharton.upenn.edu

Abstract: Analytical procedures are recommended by the AICPA to calibrate the risk level of the audit. In this paper, we detail an Analytical Procedures technique to aid in assigning an initial risk level to the audit client. Following the AICPA *analytical procedures* guide, we offer a panel analytical procedures model that uses *a priori* accrued firms to form benchmarking triage for the main effect: risk level. After testing for the expected fixed effects of the firm benchmarks and also for the explicit input variable power, we show that the panel risk assignment protocol (PRAP) only needs a set of high risk firms as the client benchmarking group. We offer six other sources of risk-related information that may inform the audit regarding client relative risk at the planning stage of the audit. Finally, the final scoring of the risk level is produced by a VBA-Excel decision support system that is available free as a download without restrictions on its use.

Keywords: longitudinal benchmarking; profiling client risk.

Reference to this paper should be made as follows: Lusk, E.J. and Halperin, M. (2016) ‘Client risk calibration in PCAOB audits: an analytical procedures panel risk assignment protocol’, *Int. J. Auditing Technology*, Vol. 3, No. 1, pp.1-21.

Biographical notes: Edward J. Lusk is Professor of Accounting, the State University of New York (SUNY), School of Business and Economics, Plattsburgh, NY USA and Emeritus: The Department of Statistics, The Wharton School, The University of Pennsylvania, Philadelphia, PA USA. From 2001 to 2006 he held the Chair in Business Administration at the Otto-von-Guericke University, Magdeburg, Germany. He is also an Academic Consultant to the Public Accounting Firms of: *McSoley McCoy, LLP*, Burlington, VT USA and *Deloitte Touche, LLP* Frankfurt [Main], Germany.

Michael Halperin (retired) had been the Director of the Lippincott Library, the Library of the Wharton School, of the University of Pennsylvania since 1984. He had served as liaison for the Departments of Accounting, Finance and Statistics of the Wharton School. Previously, he worked in the Temple University and Drexel University Libraries.

1 The analytical procedures context for risk calibration of the audit

Analytical procedures (AP) were re-introduced *by implication* as part of Audit Standard 2 (AS2, http://pcaobus.org/Standards/Auditing/Pages/Auditing_Standard_2.aspx) of the Public Company Accounting Oversight Board [PCAOB: Sarbanes-Oxley: Pub. L. 107-204, 116 Stat/745 (2002)] which is the Public Accounting LLP licensing partner of the Security and Exchange Commission (SEC). Perhaps a better phrasing would be that the emphasis on the importance of AP was *resuscitated* as they have always been part of the execution of certification audits. Specifically, regarding the statistical domain of AP, Kinney (1978) tested the possible use of ARIMA models [including transfer-function arrangements] profiled against various benchmarks: OLS linear-regression and naïve models in the martingale class. These ARIMA models were previewed for the first time in the AP context – essentially as forward projection tools to provide testing information at the substantive phase. At the other end of the analytics spectrum is the work of Bettauer (1975) a partner at Price Waterhouse LLP. He examined various non-modelling information sources that offer AP inferential possibilities. Both of these early works correctly point out that analysis of past information at the outset of an audit can make projections, the nature of which, can impact the amount of testing at the substantive phase which is directly related therefore to the risk level of the audit. Kinney (1978, p.48) notes:

> "Analytical review can be a relatively inexpensive means of increasing auditor confidence in the validity of reported balances through the assessment of the 'reasonableness' of balances in view of all known circumstances. The basis for analytical review can range from information scanning and comparison by an experienced auditor to the application of various statistical models."

More currently, AP are correctly identified by the PCAOB, SEC and the American Institute of Certified Public Accountants (AICPA) as an important way for the auditor to create audit evidence, the evaluation of which will rationalise or justify the opinions that are required in certification audits.

The AICPA *Audit Guide (2012): Analytical Procedures*, a deliverable of the *Clarity Project* of 2004 (http://www.aicpa.org/InterestAreas/FRC/AuditAttest/Pages/ImprovingClarityASBStandards.aspx) supported by the Auditing Standards Board of the

AICPA, cites the following definition of AP as offered by the *International Auditing and Assurance Standards Board* [IAASB: para *04*;p.436: AU-C:ISA 520 (in conjunction with *International Standards on Auditing* [ISA] 200)]:

> "For the purposes of the ISAs, the term 'analytical procedures' means evaluations of financial information through analysis of plausible relationships among both financial and non-financial data. Analytical procedures also encompass such investigation as is necessary of identified fluctuations or relationships that are inconsistent with other relevant information or that differ from expected values by a significant amount."

Additionally, the AICPA in their AP guide provides a very valuable case study: *On the Go Stores*™ where the following four dimensions offered in chapter 1 of their audit guide: *trend analysis*, *ratio analysis*, *reasonableness testing (testing account balances relative to expectation)* and *regression analysis* are expertly detailed and their use clearly illustrated in the creation of AP information sets for the certification audit. One appreciates the detailed explanations offered by the AICPA in their audit guide as this is a critical aspect of informing *all* the individuals that are charged with oversight of particular audits and so speaks to the quality of the oversight review. In this regard we will endeavour to follow the AICPA illustrative modelling format for the AP model that we offer for consideration.

2 Risk 'issues' and dimensions related to the use of AP

It is critical to understand, *and we wish to emphasise*, that there is a dynamic structural relationship between:

1 the nature and the use of the AP in the audit

2 the risk profile of the audit client (AC)

3 the risk born by the auditor relative to justifying the two opinions required by the PCAOB through the SEC.

Most simply stated:

- The nature of audit evidence created through the use of AP is certainly the judgemental purview of the auditor, as stated clearly in the AICPA audit guidelines: Generally Accepted Auditing Standards (GAAS): in particular General Standard 3 and Standards of Field Work: 1, 2 and 3. The planning of the collection of audit evidence in executing these GAAS protocols depends upon the perceived risk of the AC which is the risk or, more appropriately the jeopardy, faced by the auditor relative to the certification opinions required of the auditor by the AICPA, PCAOB and SEC.

This argues clearly that the *first* question to be addressed by the auditor at the planning stage of the audit is: *What is the risk level of this auditee?* In turn, the elaboration of *risk* is formed around the following question:

- Given the client's policies, procedures and protocols as they pertain to the AIS system of internal control over financial reporting: *What is the possible chance, due to error or fraud that material errors will be found in the clients' financial statements?*

This thematic connection between risk and the audit actions is detailed by Imoniana et al. (2012). One of the important experimental results of their study of students, many of whom were experienced in audit matters and all of whom worked with one of the big four LLPs, is that *professional judgement* and *materiality* play a central role in executing the audit (see also O'Donnell and Perkins, 2011). As Materiality is often client specific and so is benchmarked by the value of the clients' accounts particularly at the substantive phase this leaves professional judgement as one of the other drivers of the audit. Therefore early-warning of possible risk issues, if you will, a Bayes-conditioning of the audit context, can provide important AP-design imperatives so that the auditor can adequately plan the audit. Also recognise that this is the underlying underpinning of the AP-protocol offered by ISA (2009, p.520): analytic procedures as it pertains to the substantive phase:

> §5. When designing and performing substantive analytical procedures, either alone or in combination with tests of details, as substantive procedures in accordance with ISA 330,3 the auditor shall: (Ref: Para. A4–A5)
>
> (a) Determine the suitability of particular substantive analytical procedures for given assertions, taking account of the assessed risks of material misstatement and tests of details, if any, for these assertions; (Ref: Para. A6–A11)

Finally, this conditioning focus is consistent with the work of Wheeler and Pany (1990) and Fukukawa et al. (2011) who also emphasise that AP are critically important in calibrating the risk-level of the AC so that the auditor can use this risk-assessment to intelligently design the testing procedures for collecting audit evidence.

This is the point of departure of our study. We are encouraged by and appreciate the AICPA initiative in the *clarity project* of effecting a detailed AP 'work-up' of the *On the Go Stores* case as reported in AICPA (2012): *Analytical Procedures*. Such detailed case illustrations are the best way to aid *all* the individuals in the audit oversight-chain to understand, embrace, and use all the technical dimensions possible in executing the AP phases of the certification audit. In this spirit, we will continue the *statistical-thread* opened by the AICPA by examining an AP technique called: *panel data analytics* that may be used to initially calibrate the risk level of the audit. Following, we will:

1. Detail an AP modelling protocol called: panel risk assignment protocol (PRAP) for benchmarking the AC so as to create a relative risk profile at the Planning stage of the audit using a longitudinal profile of the data. This Panel analysis embodies, elaborates, and extends the four components: *trend analysis*, *ratio analysis*, *reasonableness testing* and *regression analysis* as previewed in the *On the Go Stores* case. Also the panel is a direct and relevant extension of Kinney's ARIMA modelling protocol in that it extends the ARIMA(0, 1, 1): [also the simple exponential smoothing model – itself an extension of the moving average models(k)], and the ARIMA(0, 2, 2): (also the two-parameter Holt model) by adding regression-impact variables in the filtering process. The longitudinal extension is consistent with the ISA (2009) discussion of AP. They note:

> Suitability of Particular Analytical Procedures for Given Assertions (Ref: Para. 5(a))
>
> A6. Substantive analytical procedures are generally more applicable to large volumes of transactions that tend to be predictable over time. The application of planned analytical procedures is based on the expectation that relationships among data exist and continue in the absence of known conditions to the contrary. However, the suitability of a particular analytical procedure will depend upon the auditor's assessment of how effective it will be in detecting a misstatement that, individually or when aggregated with other misstatements, may cause the financial statements to be materially misstated.

2 Illustrate and provide developmental as well as holdback testing validation of the application of the PRAP.

Consider now the operational nature of the panel analysis and the information needed to make the parameterisations of this model form.

3 An AP initial risk-benchmarking protocol using a panel dataset analysis

The logic of the APs risk assessment protocol is simple; the auditor wishes to profile or benchmark *the client under audit* (AC) against two sets of market traded firms: one accrual group is a collection of n 'high' risk firms: $\{F_i^{hr}, i: 1, 2, \ldots, n\}$, and the other accrual group is a collection of m 'low' risk firms: $\{F_j^{lr}, j: 1, 2, \ldots, m\}$. The simple conditional risk assignment protocol is:

- *If* the AC is *more likely* to be, in nature, a member or aligned with the collection of firms in the high risk (HR) accrual set: $AC \in \{F_i^{hr}\}$ *than* to be aligned with the collection of firms in the low risk (LR) accrual set, $AC \in \{F_j^{lr}\}$, *then* the initial risk level of the AC is assessed as high and *visa-versa*.

In this regard, there are indeed many such benchmarking accruals groups that could form the accrual selection menu for the auditor. For example, we have examined and recommend for consideration:

1 *Intra-group screening protocols:* specifically, for either the SIC or the NAICS groups for a selected performance variable, V_k, of interest, $\{F_j^{lr}\}$ and $\{F_i^{hr}\}$ are blocked into *one* of the SIC or NAICS groups such that: $\{F_j^{lr}\}$ are firms that relative to V_k are:

   a lower than the median

   b lower than the 25% percentile

   c lower than the 10% percentile.

   While $\{F_i^{hr}\}$ are a collection of firms that relative to V_k are:

   a higher than the median

   b higher than the 75% percentile

   c higher than the 90% percentile.

Typically, the firms used to populate: $\{F_j^{lr}\}$ and $\{F_i^{hr}\}$ are randomly selected from these blocked partitions – to wit one selects firms within (intra) partitions a, b or c.

b *Inter-group screening protocol:* this is the same as the *intra*-grouping except that the firms are not blocked as to a particular SIC or NAICS collective nor are the random selections restricted to intra-quantile selections. So for example, $\{F_i^{hr}\}$ could be drawn from a relevant NAICS group for the V_k partition a; whereas $\{F_j^{lr}\}$ could be a collection from some other SIC or NAICS group for the V_k partition b.

c *Acquisition and target groups:* in this accrual protocol for a blocked or un-blocked SIC or NAICS accrual set: $\{F_i^{hr}\}$ are firms randomly selected from the collection of firms that were rated as *M&A targets* while $\{F_j^{lr}\}$ are so selected from the collection of firms that were rated as *M&A acquirers*. CapitalCube™ (http://www.capitalcube.com), for example, has just such a rating system.

d *Rated or scored firm groupings:* there are a number of organisations that rate or score firms:

- CapitalCube™ mentioned above
- Bloomberg™ (http://www.bloomberg.com/markets/world)
- Morningstar™ (http://www.morningstar.com/)
- Cable News Network, Inc (CNN™, http://money.cnn.com/data/markets/sandp/).

These are proprietary organisations. In the public domain, we strongly recommend the SEC COSO-scoring where details are given to rationalise the fact that certain traded organisations have been identified as having *PCAOB Section 404 Weaknesses:* http://www.sec.gov/spotlight/soxcomp.htm. The summary of SEC weakness criteria is found at https://www.sec.gov/info/smallbus/404guide.pdf.

In considering these myriad numbers of possibilities, we elected to use an inter-group SIC scoring using the SEC weakness designations (*discussion point 4 above*), as the accrual protocol as follows:

1 *HR benchmarking group:* for 2008, we first accrued those firms reported by the SEC as having *weaknesses* in *internal control over financial reporting* – the so called 404/302:COSO-flag. There were 195 firms in this group. We then took this Weakness group and collected from *AuditAnalytics*™ (through *WRDS*™) the five previous years [2003 to 2007] of reported *market cap*: i.e., market cap was our version of V_k as the panel response measure. We then selected as the most extreme case: those firms in the weakness group that were *lower* than the median split for their SIC cohort for *each* of the five years 2003 to 2007 relative to market cap. Further, we eliminated any firms that did not fill in the panel variable set for: V_k and the explicit input variable set noted below. These eliminations then tacitly avoid survival bias issues in the HR Panel group. This yielded a sample of n = 11 HR firms: $\{F_{11}^{Weakness}\}$ – the SEC-weakness flagged group.

2 *LR benchmarking group*: for 2008, we took a random sample accrual of 25 firms traded on the NYSE or the NASDAQ that:

a were *not* identified by the SEC as having weaknesses in *internal control over financial reporting* for *any* year from 2003 to 2012 inclusive

b for the entire set of Panel firms did not replicate more than twice a four digit SIC code.

We then took this *LR* group and collected from *AuditAnalytics*™ the five previous years (2003 to 2007) of their reported market cap. We then selected as the most extreme case those firms in the LR group that were *higher* than the median split for their SIC cohort for *each* of the five years 2003 to 2007 relative to market cap. Further, we eliminated any firms that did not fill in the Panel variable set for: V_k and the explicit input variable set noted below. These eliminations then tacitly avoid survival bias issues in the LR Panel group. This yielded a sample of n = 9 LR firms. $\{F_9^{Leaders}\}$ –the sample from the NYSE or NASDAQ of industry leaders.

As for the *explicit input variables* mentioned above, we selected the following five 'impact' variables from the standard text in ratio analysis (Fraser and Ormiston, 2013):

1 *CR* (liquidity): current ratio [current assets/current liabilities]

2 *ROA* (profitability): return on assets [net income/total assets]

3 *ROE* (aggregated profitability): return on equity [net income/stockholder equity]

4 *FLI* (return leverage): financial leverage index [ROE/adjusted ROA]

5 *GR* (operating leverage): gearing ratio [total assets/total liabilities]

6 *D/E* (debt leverage): [total liabilities/stockholder equity].

This impact-variable set fits well with the client impact cycles identified by the empirical study of Imoniana et al. (2012, *Frame 3*, p.293).

Finally, for purposes of robustness relative to the *response variable: market cap*, we collected *EPS: including and excluding extra-ordinary items* as additional response variables.

Panel summary: effect consideration tested: risk level; effect groups: $\{F_{11}^{Weakness}\}$ HR firms and $\{F_9^{Leaders}\}$ LR firms (see the Appendix); response performance variables: *Market Cap*, $EPS_{withE.I}$ and $EPS_{withoutE.I}$.; input performance variables: *CR*, *ROA*, *ROE*, *FLI*, *GR* and *D/E*; and panel (2003 to 2007).

This may appear to be a rather small accrual set for benchmarking purposes; however, it is important to realise that we have used a very focused accrual protocol differentiated on the Effect variable and so probably have reduced the blocked – within effect – variation compared to the non-blocked – overall – variation for the two effect sets. Usually, even for the panel fixed-effect model which is, in general, less powerful than the random effect binary-comparison test, this moves in the direction of enhancing power relative to effect size if, as expected, the null is *unlikely* to be the state of nature. We will report the statistical profiles for the intra-effect fit that demonstrate the adequacy of the sample size by:

1 reporting the number of independent variables (*CR*, *ROA*, *ROE*, *FLI*, *GR* and *D/E*) which tested in the panel analysis to be lower than 0.05 in p-value – tacitly power for the independent variable relative to actual effect magnitude

2 also the Hausman test (1978) for fixed effects.

4 The panel risk assessment protocol (PRAP)

As indicated above, the idea is very simple for using the PRAP at the planning stage. The audit in-charge will create profile matching information for *the firm under audit* (AC) relative to:

- *Scenario A:* if AC appears *to fit into* or *aligns with* the firms in the weakness/HR group more than it *fits into* or *aligns with* the LR group, then the in-charge will have a benchmarking indication that the *risk level* for the AC may be recorded as *above* average. This assessed risk level would then be compared with all the other usual sources of risk evidence information collected at the planning phase, the interim phase and even possibly at the substantive phase as the risk calibration of the client.
- *Scenario B:* if AC appears *to fit into* or *aligns with* the firms in the LR group more than it *fits into* or *aligns with* the weakness/HR group, then the in-charge will have a benchmarking indication that the *risk level* for the AC may be recorded as *below* average. This assessed risk level would then be compared with all the other usual sources of risk-evidence information collected at the planning phase, the interim phase and even possibly at the substantive phase as the risk calibration of the client.

The operational question is, of course: *What is the measure of 'fit' for the AC relative to the scenario A & B groups?* This measure is detailed in the following section.

4.1 The measure of conformity relative to scenario risk profiling

The panel model which is used in the inference risk profiling for our study is presented by Greene (2008, Ch9, pp.180–252) and is coded by SAS (2013)/$ETSA_2$/JMPv.10 (Panel Analysis PROC: Panel) *version9.4*. This is a standard panel model and is well detailed in Greene, and also extensively illustrated in the SAS/ETSA reference manual noted above. For this reason, and at the suggestion of our critical readers and cognisant of the word-count limitations, we have decided not to present the equation sets of this standard model form. However, they are available from the corresponding author upon request.

The protocol format used in testing the risk assignment was:

1. The audit in-charge uses the AC as the *profile firm* – i.e., the h^{th}, firm to be entered in cross-section of the panel. In this way, the AC is used as the benchmark of all the other (h-1) cross-sectional firm entries.
2. Then the auditor enters a set of randomly selected HR and LR benchmarking firms; order does not matter as the comparison of the profile firm with the other firms in the cross-section is unconditional-pairwise. The time carve-out that we recommend is five years as this longitudinal time series fits well with most of the AICPA/PCAOB/SEC audits where five years of prior data is the norm reported in the 10-Ks. We have experimented with a variety of selections from the HR and LR accrual sets. We find that three to four firms from each in the cross-section is sufficient to parse the AC. Additionally, we have experimented with random selections from the response variable set as well as a randomisation of the six input

variables and the parsing profile seems robust to changes in these modelling parameterisations. In summary, for testing the PRAP we used the following *PRAP selection menu* for *randomly* forming the panel for testing:

- the 'AC' was selected from the set of HR or LR firms
- three to four firms from the HR accruals
- three to four firms from the LR accruals
- a response variable from the set of three response variables
- two to six input variables from the set of six input variables.

3 Finally, one needs to form a logical expectation as to the p-value relationship between the cross-section sets relative to the AC – i.e., the firm under audit. Given the nature of the profiling logic it is expected that:

- If the AC is, in nature, not different from the HR or the LR set of firms then one would expect that the cross-section pairwise p-values would be relatively high suggesting that the observed profile comparison relative to the null is expected a high percentage of the time.

Simply: the null of the pairwise comparison assumes that the two firms are 'identified as being from the same risk-effect group'. The p-value is the percentage of time that one would observe the actual comparative profile if the null were to be the true state of nature. In this case for the false positive error (FPE) profiling, we have *set* the differentiating p-value as a *p-value equal to or greater than [= >] 0.25*. This is called a *high* p-value and suggests that there is very little difference in the profile comparison and so that there *is scant evidence that this comparative difference is a rare occurrence*. Therefore, the firms in this pairwise contract are assumed to be not different in nature. If on the other hand the p-value is *low*, *p-value less than [<] 0.25*, this suggests that if the Null of no difference in the pairwise comparison were to be true that it is a rare random sampling event to have found an observed 'substantial' difference for the cross-sectional firms selected. This argues then for rejecting the null in favour of a structural difference between the two firms.

4 In this p-value triaging logic we have selected as the crisp-cut-point a p-value of 0.25. This suggests that:

a If the p-value for a specific pair-wise firm cross-sectional comparison is greater than or equal to 25% in the FPE domain, then one fails to reject the null of no difference. On the other side of this binary partition:

b If the p-value for the firm cross-sectional comparison is less than 25% in the FPE domain, then one rejects the null of no difference in favour of a structural difference between the comparison firm and the profile firm.

4.2 Operational caveat for the effect partitioning

We collected feedback from our critical readers that possibly these expectations are conditional on the nature of the audit firm as the audit profile firm. This is an asymmetrical profiling issue in the following way: For the LR firm set there are many discretionary accruals and/or initial parameterisations of the expense and revenue AIS protocols that allow the firms under the guidance of the controller, CFO, or the COO, to select from the GAAP: USA menu in such a manner so as to form highly differentiated short term trajectory profiles relative to the relationship between the response variable set and the input variable set. This likely will result in a higher proportion of LR firms profiled against LR benchmarks exhibiting low or differentiating p-values than will be the case for HR firms profiled into HR benchmarks. Such election possibilities seem more likely for firms in the LR profile set as they are 'better positioned' relative to firms in the HR set regarding the strategic or tactical alternatives that could be selected. Simply put *MMM[3841]:* the 3M Corporation, a clear member of the LR set of benchmarks firms, has many more options in electing strategic or tactical initiatives than a firm in a distressed financial position which has been flagged by the SEC for COSO Weaknesses; distressed firms are relatively more homogeneous and so will have relatively higher p-values compared to the more heterogeneous LR firms. We shall test for this expected differential profiling in examining the state of nature set comparisons for our accrual set of firms.

5 The test results using the state of nature benchmarks

To create the information needed from the PRAP to evaluate the nature of the risk of the AC, we randomly selected from the set of 20 firms presented in the Appendix the firm to be the AC. Recall that the risk assessment is related to the observed p-value of the Cross-Sectional pair-wise comparisons. In this testing protocol, we *know* the state of nature information in that we have accrued the sets of HR and LR firms. For the risk classification, we will collect the FPE p-value, here after p-value, for each of the pairwise Cross-Sectional groupings. Based upon these profiles, we can form a classification taxonomy and evaluate the same. At this juncture an example would be useful.

An illustrative example: assume that we have as the AC: ATCV (see the Appendix). Next we randomly selected from the set of HR firms the following three firms from the HR accrual set: HR cross-section (ETAK, BASI and CKX); further, we selected the following three LR firms (HD, KMB and LOW) from the LR accrual set. Then we randomly selected $EPS_{\text{Basic Including Extra-Ordinary Items}}$, as the response variable; and finally we selected: CR, ROA, ROE, FLI and GR as the input variable set. In this case, we found the p-value profile reported in Table 1:

Table 1 The HR (ATCV) audit firm panel profiled against the LR firms (HD, KMB and LOW) and HR firms (ETAK, BASI and CKX) relative to observed cross-section p-values

Contrast	*Cross-section p-value*	*Projected state*
ATCV r* HD	<0.0001	Likely differ
ATCV r KMB	<0.0001	Likely differ
ATCV r LOW	<0.0001	Likely differ
ATCV r ETAK	0.4618	Likely do not differ
ATCV r BASI	0.3991	Likely do not differ
ATCV r CKX	0.0686	Likely differ

Note: *Where r means the pairwise panel cross-section *relative* comparison.

Additionally, for the input variables we find: CR (0.2631), ROA (0.7041), ROE (0.0092), FLI (0.0096) and GR (0.3543) where the p-value is in brackets; the Hausman test was $p < 0.0001$ a strong rejection of NO fixed effects. With this information, the auditor can anticipate or access the likely group membership of the AC. As the auditor knows the likely state of nature for the LR firms (HD, KMB and LOW) and also the likely state of nature for the HR firms (ETAK, BASI and CKX), the likelihood for the AC is strongly for the ATCV to be a HR firm as that profile has the sample likelihood odds of 83% in support of that profile; the alternative is then that the Firms is actually a LR firm where the only support for that is the tuple (ATCV r CKX) which happened 17% of the time. Alternative summary explanation: The low cross-sectional p-values suggest that the AC: ATCV is not similar in variable profile to HD, KMB, LOW and CKX. The high cross-sectional p-values suggest that the audit firm: ATCV is similar in variable profile to ETAK and BASI. In this profiling there is one 'misclassification' where ATCV does not align with CKX where it was expected to be similar to CKX in variable profile. Hoverer, in five of the remaining cases, the profiling is correct: low cross-sectional p-values are found for the LR firms (HD, KMB and LOW) and so ATCV, a HR firm, is judged as not similar to the LR set. Finally, ATCV, a HR firm does align with the other two HR firms (ETAK and BASI) as the high p-values suggest. This is correct five times of six or 83% of the time.

Consider now the full testing of the p-value triage risk assessing protocol of the PRAP. In this testing we have randomly formed a development dataset where we may, *if needed*, make a re-alignment of the *a-priori* cut point for the p-value triage and then re-test that modified cut-point with a Holdback dataset. After we consider these results, we suggest a risk assignment modelling form based upon the p-values in the cross-section of the PRAP.

6 The developmental dataset profile for the PRAP

The firms that we randomly selected for the development testing phase were:

- developmental phase LR firms: BA, HPQ, HD, KMB, LOW and MMM
- developmental phase HR firms: BLFS, CKX, LPTH, BASI, NTWK, ETAK and ATCV.

The classification profile using the p-value cut-point of 0.25 is presented in Table 2.

Table 2 Overall classification based upon the p-values of the cross-section for the developmental test

*Contrast**	*Correctly classified*	*Misclassified*
LR_{AC} r LR_{Set}	22.9% [11 of 48]	77.1% [37 of 48]
LR_{AC} r HR_{Set}	97.9% [47 of 48]	2.1% [1 of 48]
HR_{AC} r LR_{Set}	91.7% [44 of 48]	8.3% [4 of 48]
HR_{AC} r HR_{Set}	79.2% [38 of 48]	20.8% [10 of 48]

Notes: *Where the sub-script *AC* indicates audit firm, the h^{th} cross-sectional firm and the sub-script *Set* indicates the randomly selected comparison set.

Recall that *each* of the comparisons was made after random selections from the menu: *PRAP selection menu*. In this case, there is strong practical evidence that the p-value cut-point of 0.25 is an effective triage point except for the case where there is a LR firm as the AC profiled against other LR firms. For example, for the LR AC as projected into the LR accrual set only 11 times of 48 trials (22.9%) was the cross-sectional p-value high suggesting correctly that the LR firm did profile or align with the firms in the LR accrual set. This is consistent with the asymmetrical discussion above in that frequently the LR AC does not align very well with the firms in the LR accrual group. The simple summary of Table 2 is that the PRAP works in the majority of cases. It seems a slightly better and consistent screen when projecting into the HR set of firms. For this reason we have elected to retain the p-value cut point of 0.25. We have shaded these comparative instances. As a robustness check on these development results we formed as a Holdback test those firms not used in the developmental set.

7 The holdback dataset profile for the panel AP: risk assessing model

The actual firms that we randomly selected for the holdback testing were:

- holdback LR firms: PG, SLB and ORCL
- holdback HR firms: TCCO, TOF, UAMY and WVVI.

The classification profile using the p-value cut-point of 0.25 is presented in Table 3

Table 3 Overall classification based upon the p-values of the cross-section for the holdback test

Contrast	*Correctly classified*	*Misclassified*
LR_{AC} r LR_{Set}	31.2% [15 of 48]	68.8% [33 of 48]
LR_{AC} r HR_{Set}	93.0% [66 of 71]	7.0% [5 of 71]
HR_{AC} r LR_{Set}	84.4% [38 of 45]	15.6% [7 of 45]
HR_{AC} r HR_{Set}	89.3% [50 of 56]	10.7% [6 of 56]

In this case, there is also strong practical evidence that the cut-point of 0.25 is an effective triage point given that the state of nature profiles are *a priori* known. Also the basic profile of these developmental and holdback profiles are effectively similar. As these developmental and holdback results are directionally consistent and as their $z_{\text{-calcs}}$ for the highest difference of their blocked comparisons is lower than 1.645 strongly suggesting that the null cannot be confidently rejected, we have combined them to offer an average profile of the p-value triage. In this regard we shall take advantage of the expected asymmetry as between the HR and LR profiles to develop an expectation as to the nature of the risk level of the audit firm.

Incidentally, as a power profile for the *PRAP selection menu*, we collected information for the number of p-values for the explicit input variables (randomly between 2 and 6 selected) that were <0.05. For the development and the holdback trials there were 262 variables used over the 75 panel runs; only two times in the 75 runs was there only one p-value for the explicit input variables <0.05. Further, there was never an instance where there were no cross-sectional p-values < 0.05. Also using the Hausman screen in 100% of the test cases the p-value was sufficiently low to reject the random effects likelihood of no fixed effects. This also, of course, speaks to the sufficiency of the power of the PRAP selection menu. This argues that the panel selection menu had sufficient power relative to the various effect sizes in the accrual set.

8 The aggregate profile and the triage risk profile: the p-value profile

Tables 2 and 3 speak to the classification efficacy where we knew the state of nature of all the datasets. We observe that there is a marked classification difference for the LR firms. As expected LR ACs do *not* seem to profile with the LR firms. However, for the HR accrual firms there is marked triage separation for the projection of a LR AC into the HR accrual set compared to projection of a HR AC into the HR accrual set. In this regard, to enhance the risk-triage effectiveness:

- If we switch the risk-triage to the magnitude of the p-values for projection of the AC *only* into the HR accrual set then the simplest and most effective triage from the PRAP cross-sectional p-value set is:
  - project the AC into the HR accrual set of firms and form the p-value profile of high p-value as the likely indication of the risk level of the AC.

Using this modification in the risk-classification logic, the triage separation for the development and the holdback sets for projection of a LR AC and a HR AC into the set of HR firm accruals is presented in Table 4.

Table 4 Conditional distribution of p-values = > 0.25 for the HR set of firms

Contrast	*HR accrual developmental*	*HR accrual holdback*
Audit Firm LR	2.1% [1/48]	7.0% [5/71]
Audit Firm HR	79.2% [38/48]	89.3% [50/56]

We tested the separation of the HR and the LR percentages reported in Table 4. The weighted average of the two proportions is 5.0420% and 84.6154%. The p-value of the non-directional difference is < 0.0001 clearly suggesting that one may reject the Null for the relative distribution of the high p-values suggesting that they are likely from different populations.

The logic of this dramatic separation was suggested as part of the asymmetrical profiling for the two classes of firms: *HR*, distressed firms, are expected to exhibit less discretionary GAAP profiling for the usual impact variables that affect cash flow from operations and so they often share the same distressed operational profile. This is in contrast to the expectation for *LR*, non-distressed firms, where there are likely to be very different discretionary GAAP profiles. Simply: *LR*, well-managed, are market leaders with a variety of 'successful game plans' in place. *HR*, distressed firms, are in a survival mode and so their operational choices are very limited.

8.1 Conditional risk assignment rationale

Concentrating ONLY on the HR accrual set, we find that:

- *If* there are very *few* high p-values for the AC projected into the HR accrual set, *then* this is the likely case for a LR AC. Alternatively, *if* there are very *many* high p-values, *then* this is the likely case for a HR AC.

As we have argued that the four profiles from the development and the holdback sets are not significantly different and so we have combined them, we shall now form the logical $(1 - \alpha)$ confidence intervals, the service of which, is to assign the risk level to the AC. Blending the results of Table 4:

1 *PRAP: conditional characteristic of a LR AC as projected into a HR cross-sectional set of firms*. Using the weighted average of high p-values of Table 4 we find the sample average of: [1 of 48] with [5 of 71] or [6 of 119] gives 5.0420%. The $(1 - \alpha)$ confidence interval for this result is: *likelihood interval suggesting a LR AC:*

   - lower and upper bound: $0.050420 \pm [z_{(1-\alpha)} \times 0.020058]$.

   For the 95% confidence interval we find in percentage terms:

   - lower bound to upper bound: [1.1% to 9.0%].

2 *PRAP: conditional characteristic of a HR AC as projected into a HR cross-sectional set of firms.* Using the weighted average of high p-values of Table 4 we find the sample average of: [38 of 48] with [50 of 56] or [88 of 104] gives 84.6154%. The $(1 - \alpha)$ confidence interval for this result is: *likelihood interval suggesting a HR AC:*

   - lower and upper bound: $0.846154 \pm [z_{(1-\alpha)} \times 0.035379]$.

   For the 95% confidence interval we find in percentage terms:

   - lower bound to upper bound: [77.7% to 91.5%].

These two $z_{(1-.05)}$ confidence intervals can be formed into a line-score formula as illustrated for the 95% case:

- *Case A:* if the *AC*, *heretofore unknown as to risk class*, produces a proportion of high p-values, $\hat{\pi}$, for the projection into the HR accrual set such that $\hat{\pi}$ is in [1.1% to 9.0%] then the strong likelihood is that the AC is a *LR* firm.

- *Case B:* if the *AC*, *heretofore unknown as to risk class*, produces a proportion of high p-values, $\hat{\pi}$, for the projection into the HR accrual set such that $\hat{\pi}$ is in [77.7% to 91.5%] then the strong likelihood is that the AC is a *HR* firm.

In summary, for the line-scoring form the strong likelihood inference expectations for the LR and HR assignments lie *in* the two 95% confidence intervals respectively:

[0% _ [1.1% to 9.0%] __________________ [77.7% to 91.5%] ___ [100%]

Panel results outside of these intervals of course, a logical question is: *What if the values of $\hat{\pi}$ are not in these intervals?* There are three cases that could obtain: $\hat{\pi}$ in [0% to 1.1%), $\hat{\pi}$ in (9.0% to 77.7%) or $\hat{\pi}$ in (91.5% to 100%]. The two extreme or polar cases are simple. For $\hat{\pi}$ in [0% to 1.1%) by logical extension of the above results the polar-likelihood would be for a LR firm to have produced this result. For $\hat{\pi}$ in (91.5% to 100%] the logical polar-likelihood would be for a HR firm to have produced this result. For the third case: $\hat{\pi}$ in (9.0% to 77.7%), we have formed a $(1 - \alpha)$ indifference confidence interval around the mid-point of the range between the two adjacent CI boarders. Specifically, we found the weighted average of the midpoint between 9.0% and 77.7% which is: 41.02%. This average can be used to form the $(1 - \alpha)$ indifference confidence interval using the average of the sample experiences which is: 0.410164 ± $[z_{(1-\alpha)} \times 0.046581]$. For example, the 95% indifference confidence interval is: [31.9% to 50.2%].

8.1 PRAP line scoring taxonomy

This produced the following re-calibration of the above line scoring risk assessment intervals that we call the *PRAP line scoring taxonomy which is illustrated for the 95% confidence case*. We have created a decision support system (DSS) to aid in the creation of the various confidence interval where there are four choices 80%, 90%, 95% & 99% Confidence. The DSS is programmed in open access Excel™; VBA™ (2013) and available free as a download without restriction on its use (see the Appendix for the screen capture of the DSS).

The *LR* $\{F^{lr}\}$ *zone* is:

- [0% to 1.1%) *logically* suggests a LR AC
- [1.1% to 9.0%] *strongly* suggests a LR AC from the PRAP results
- (9.0% to 31.9%) *weakly* suggests a LR AC.

The *zone of indifference* between the LR and the HR zone [31.9% to 50.2%] $\hat{\pi}$ in this interval suggests that there is *no information from the PRAP* that can aid the auditor in intuiting a risk classification for the AC.

The *HR* $\{F^{hr}\}$ *zone* is:

- (50.2% to 77.7%) *weakly* suggests a HR AC
- [77.7% to 91.5%] *strongly* suggests a HR AC from the PRAP results
- (91.5% to 100%] *logically* suggests a HR AC.

Now that we have formed a line scoring taxonomy for the PRAP model let us address the nature of the jeopardy of the use of the PRAP model where there *is* a misclassification error.

8.3 Misclassification error in the PRAP parameterisation

As we are projecting an AC with an *a priori* unknown risk profile into a HR set of firms the most *extreme* classification error is that the auditor accrues a set of LR firms but *incorrectly* assumes that *all* of these firms are HR in nature. In this case, averaging the results for Tables 3 and 4 and mistaking LR firms for HR firms we can estimate the exhaustive binary classification error for this extreme result as follows:

1 If AC were actually a LR firm then the percentage of high p-values would be [11 of 48] and [15 of 48] or $\hat{\pi}$ = 27.08%; this would weakly suggest that the risk of the AC is low as 27.08% is in (9.0% to 31.9%). In this case the classification is in fact correct as the AC is LR and so insofar as the binary classification is concerned there is no error.

2 If AC were actually a HR firm then the percentage of high p-values would be [4 of 48] and [7 of 45] or $\hat{\pi}$ = 11.8%; this would weakly suggest that the risk of the AC is low as 11.8% is in (9.0% to 31.9%). In this case the classification is in error as the AC is in fact in the HR class and so insofar as the binary classification is concerned there is an error.

In either case, the PRAP metric that we have suggested above will weakly classify the AC as LR and it is, indeed, very unlikely that the AC would *ever* be classified as HR if a LR accrual set were incorrectly assumed to be HR. This would mean that the classification error would compromise the risk classification as the auditor would in some percentage of the cases underestimate the risk of the AC. This percentage of the cases is *a priori* not estimable as it depends on the prevalence of HR ACs for the Audit LLP. If the proportion of ACs that are HR is low then the classification jeopardy or incidence is also low. However, as a practical matter the scenario where the auditor makes a 100% misclassification by believing that all of the firms in the accrual set are HR when they are all in fact LR does seem inconsistent with an auditor qualified under the GAAS of the AICPA. In summary, such a misclassification: classifying LR firms as HR firms is *possible* but not *probable* for experienced auditors; and so *grosso-modo* the jeopardy of using the PRAP to assign a risk level to the AC seems acceptably low. This, of course, is an extension of the PRAP metric model and would be an interesting testing opportunity.

8.4 The final recommendation of the stages of the PRAP modelling system

Specifically, for assigning an initial risk level of the AC using the PRAP the auditor will:

1 Use the *AC* as the h^{th} *cross-sectional* firm in the panel.

2 Collect an accrual set of only *HR* firms.

3 Select a set of relevant *response* variables and *input* variables.

4 Using random selection from the *PRAP selection menu* detailed above to parameterise the PRAP model where the AC is profiled only against the set of HR firms in the accrual set. In our experimental work on the PRAP, we found that a set of ten HR firms and 40 to 60 cross-sectional p-values generated from the panel analysis is sufficient to operationalise the PRAP using the PRAP selection menu.

5 Calculate $\hat{\pi}$: the percentage of the number of p-values = > 0.25. Then use the *PRAP line scoring taxonomy* to assign the risk level to the AC.

9 Summary, conclusions and limitations of the PRAP modelling protocol and future directions

9.1 Summary

Analytic procedures and their related impact on assisting in establishing the risk level of the AC has a long tradition. Kinney was a pioneer in recognising the role of longitudinal analyses in this endeavour. His works (1978, 1987) are classics and set the standard for integrating statistical analytics for Risk-setting protocols. Essentially, the AICPA (2012) reinforced the integration of statistical methods as part of the tools that can often be put into productive use in calibrating the risk of the AC. Our work is thus an up-dating of the thread begun by Kinney and an elaboration of the work of the AICPA. It is of course the case that the auditor will collect other measures of risk from other sources generated by other techniques as required by the GAAS (*SFW 2*).

Following, we offer a few practical suggestions from our experience that we find invaluable in enriching the collection of other risk-related indications:

1. *Discussions between the audit-team and management:* a rich source of possible risk indicators can be gleaned from the linguistic analysis of the interchanges [both formal – usually written including e-correspondence, and informal – usually verbal] between the audit-team, the audit committee, and management (see no. 6 following as an elaboration).

2. *Practitioner opinions:* often senior practitioners offer their insights as to ways that AP can and should be integrated into the execution of the audit. Such reflections often imply sources that may be useful in calibrating the risk of the audit. For example, we recommend the advice offered by Bettauer (1975) who detailed 12 reasons for extending audit procedures. This list was published over 40 years ago but is still current today; it is standard reading in our audit and assurance courses!

3. *Understanding the details of the SEC COSO flags:* there is a copious amount of inferential or implied risk calibration information in the detailed description of the 302 indications for SEC-Weakness flagged firms. For example, assume the auditor selects the HR accrual group firms that have been identified by the SEC as having COSO-Weaknesses pertaining to their AIS; recall that a SEC Weakness designation suggests that material errors in the financial statements are possible and so likely required extensive additional testing as a condition for listing. The SEC provides very useful descriptive details in the 87 COSO coding indications for each weakness published. For example, *SEC 302 Disclosure Controls: Code 49 [DC – Restatement (recent past or pending) evident*] has the following descriptions summary: "Identifies circumstances where a restatement is noted in a disclosure control filing either directly or indirectly. It could be for example that the restatement is noted indirectly because reference is made of it in a 404 opinion which includes the identification of a restatement. Further, it could be that the company has filed a restatement but not indicated such in a disclosure control declaration. A restatement is considered a significant factor in the determination of both 302 and 404 adverse statements".

4. *Filings and related information sets:* the MD&A section of past 10-K filings and, depending on the phase of the audit, the current MD&A for the audit year provide a wealth of information that may assist in the risk calibration. According to Lee et al. (2014) since Sarbanes-Oxley: 2002 the MD&A section has about doubled in size going from around 4,500 words to around 10,000 words. Such an increase in reporting detail can aid the auditor in assessing the risk level of the AC.

5. *Rating organisations assessments of the auditee:* all firms traded on major exchanges are 'tacitly rated' by their ordered ranking in their SIC/NAICS group. These rankings are variable specific and so profiles may be simply developed over impact-variable sets; we find these profiles to be objective and useful benchmarks. We suggest that the auditor would examine the relative position of the AC longitudinally for at least three to five years for sets of impact-variables.

6 *Screening and risk calibration from research sources:* there is a wealth of risk information that can be gleaned from the manner and style that is used in discussing the results of reported information in the financial statements or in the MD&A section of the 10-K. For more than two decades content analysis (screening on various linguistic indications) has been productively used to flag firms where there can be evidence of accounting irregularities (and sometimes fraud). For example, see the work of Churyk et al. (2009) and Lee et al. (2014).

9.2 Conclusions

There are many techniques and information sources that may be used by the auditor to specify the level of risk at the planning stage of the audit. In this paper, we have offered the PRAP as one of these many techniques that could be used in the best practice execution of the audit where *risk* is an important driver of the audit. After considering all the various *risk* indications collected at the planning stage, the *initial level of audit-risk* is fixed by the audit-team using their experiential judgement. This *initial level of audit-risk* as formed around the PRAP-AP information sets the tone of the audit and eventually rationalises the extensive and intensive nature of the testing in the execution of the audit. In our presentation of the PRAP, we have taken up the same model presentation format as introduced by the AICPA in their AP presentation using the *On the Go Stores* illustration. We have therefore *not* tried to avoid using statistical modelling protocols in the AP mode but rather to create the information that facilitates their understanding and encourages their use in executing the audit. This is critical so as to encourage the 'welcoming' of new ideas of a statistical nature into the audit. In this spirit we hope that our explanations and detailed illustration of the PRAP model are sufficient for it to find its place in the panoply of the auditor.

9.3 Limitations and future directions

To be sure there are scope and design limitations of our work. It is the case, that we have used differentiated pools of comparison and benchmarking firms. Most likely this produced within comparison structural 'noise' due to the idiosyncratic generating processes particular to specific SIC/NAICS groupings – to wit, there are likely structural variational differences between The *3M* Corporation [MMM: surgical and medical instruments and apparatus, SIC:3841] and Tofutti [TOF: ice cream and frozen desserts, SIC: 2024] over the impact variables that we have used. No doubt blocking on the SIC/NAICS group would reduce this variation and so enhance the specificity of the PRAP protocol. Specifically, as extensions we encourage research that would test the possibility of:

1 Using a *simplified* relational tests where the benchmarks: strong and weak firms would be selected from a percentile wide-spectrum binary-block within their industry. In this protocol, one could select firms in their SIC/NAICS group that over the Panel were *consistently* above the top Y% percentile. These would be labelled as a strong candidate set of benchmarks; additionally firms below the (1 – Y%) percentile would be the weak candidate set. This within-industry symmetric-blocking has the advantage of expanding the matching pool while maintaining the matches within a particular industry group; recall that we were limited regarding the weak

candidate set to those firms flagged by the SEC for COSO weaknesses which essentially limited the accrual-pool. With an industry-blocked percentile pool of benchmarks the auditor could profile the AC against these two diverse sets to determine if the AC were more like the weak or strong matches using simple discriminate tests at each accrual year (or overall). This would be a simple percentage scorecard protocol. The classification differences between the scorecard and the PRAP protocols could be informative and may obviate the need for a panel protocol.

2 Another important extension of the PRAP would be to determine if informed expert judgement could be used to form a risk assignment protocol. In this design, one would need to control the accrual to a specific industry: SIC/NAICS group. Then panel profiles could be derived from the impact-variable set. These impact-variable sets could be judgementally formed into triage profiles and tested on holdback firms. This judgemental protocol would be used to evaluate the AC as having an impact-variable set that is more like either the weak/problematic profile or the strong/leading-edge profile.

These are two obvious extensions of the PRAP that should be tested.

Acknowledgements

We wish to thank Drs. H. Wright, *Boston University:* Department of Mathematics and Statistics, Boston, MA USA, Razvan Pascalau and Dhimitri Qirjo of the *SBE SUNY; Plattsburgh*, Plattsburgh, NY USA, the participants at the SBE Research Workshop at SUNY: Plattsburgh, Mr. Manuel Bern, *Deloitte Touche* LLP, Audit and Assurance Services, Frankfurt (Main), Germany for their most helpful suggestions. Additionally, appreciation to Prof. Dr. J. Imoniana, Editor of *The International Journal of Auditing Technology* for his excellent stewardship and suggestions.

References

AICPA (2012) *Audit Guide: Analytical Procedures*, American Institute of Certified Public Accountants, Inc., New York, NY, USA [online] https://competency.aicpa.org/media_resources/208125-analytical-procedures-aicpa-audit-guide (accessed 15 January 2016).

Bettauer, A. (1975) 'Extending audit procedures when and how', *Journal of Accountancy*, Vol. 11, No. 11, pp.69–72.

Churyk, N., Lee, C. and Clinton, D. (2009) 'Early detection of fraud: evidence from restatements', *Advances in Accounting Behavioral Research*, Vol. 12, No. 6, pp.25–40.

Fraser, L. and Ormiston, A. (2013) *Understanding Financial Statements*, 10th ed., Pearson/Prentice Hall, Upper Saddle Creek, New Jersey, USA, ISBN: 13: 978-0-13-265506-4.

Fukukawa, H., Mock, T. and Wright, A. (2011) 'Client risk factors and audit resource allocation decisions', *ABACUS*, Vol. 47, No. 1, pp.85–108 [online] http://dx.doi.org/10.1111/j.1467-6281.2011.00332.x (accessed 21 January 2016).

Greene, W. (2008) *Econometric Analysis*, 6th ed., Pearson/Prentice Hall, Upper Saddle Creek, New Jersey, USA, ISBN: 13: 978-0-13-600383-0.

Hausman, J. (1978) 'Specification tests in econometrics', *Econometrica*, Vol. 46, No. 6, pp.1251–1272 [online] http://dx.doi.org/10.2307/1913827 (accessed 11 December 2015).

ISA (2009) *International Standard on Auditing: 520 Analytic Procedures*, International Auditing and Assurance Standards Board IAASB® [online] http://www.iaasb.org/ (accessed 15 January 2016).

Imoniana, J., Antunes, M-T., Mattos, S. and Maciel, E. (2012) 'The analytical review procedures in audit: an exploratory study', *Advances in Scientific and Applied Accounting*, Vol. 5, No. 2, pp.282–303.

Kinney Jr., W.R. (1978) 'ARIMA and regression in analytical review: an empirical test', *The Accounting Review*, Vol. 53, No. 1, pp.48–60.

Kinney Jr., W.R. (1987) 'Attention-directing analytical review using accounting ratios: a case study', *Auditing: A Journal of Practice & Theory*, Vol. 6, No. 2, pp.59–73.

Lee, C-H., Lusk, E. and Halperin, M. (2014) 'Content analysis for detection of reporting irregularities: evidence from restatements during the SOX era', *Journal of Forensic and Investigative Accounting*, Vol. 6, No. 1, pp.99–122 [online] http://www.bus.lsu.edu/accounting/faculty/lcrumbley/jfia/Articles/v6n1.htm (accessed 27 January 2016).

O'Donnell, E. and Perkins, J. (2011) 'Assessing risk with analytical procedures: do systems-thinking tools help auditors focus on diagnostic patterns?', *Auditing: A Journal of Practice & Theory*, Vol. 30, No. 4, pp.273–283 [online] http://dx.doi.org/10.2308/ajpt-10148 (accessed 21 January 2016).

SAS Institute (2013) *SAS Econometrics and Time Series Analysis 2 for JMP*, SAS Institute, Cary, North Carolina, USA, ISBN: 978-1-61290-658-4 (e-book); ISBN: 978-1-61290-586-0 (text).

Wheeler, S. and Pany, K. (1990) 'Assessing the performance of analytical procedures: a best case scenario', *The Accounting Review*, Vol. 65, No. 3, pp.555–577.

Appendix

Firms used in forming the panel for the development and the holdback phases

Ticker [SIC]: D: development or H: holdback	*Popular company name*	*SEC coded COSO flags**
BA[3721]:D	BOEING CO	None during 2003 to 2012
HPQ[3570]:D	HEWLETT-PACKARD CO	None during 2003 to 2012
HD[5211]:D	HOME DEPOT INC	None during 2003 to 2012
KMB[2621]:D	KIMBERLY-CLARK CORP	None during 2003 to 2012
LOW[5211]:D	LOWE'S COMPANIES INC	None during 2003 to 2012
MMM[3841]:D	3M CO	None during 2003 to 2012
PG[2840]:H	PROCTER & GAMBLE CO	None during 2003 to 2012
SLB[1389]:H	SCHLUMBERGER LTD	None during 2003 to 2012
ORCL[7372]:H	ORACLE CORP	None during 2003 to 2012
TCCO[3663]:H	TECHNICAL COMMUNICATIONS	1 [17, 42, 44]
TOF[2024]:H	TOFUTTI BRANDS INC	2 [17, 42, 44]
UAMY[3330]:H	U S ANTIMONY CORP	4 [4, 17, 42, 44]
BLFS[3845]:D	BIOLIFE SOLUTIONS INC	1 [17]
CKX[1311]:D	CKX LANDS INC	1 [4, 17, 19]
WVVI[2080]:H	WILLAMETTE VALLEY VINEYARDS	4 [4, 12, 17, 22, 44]
LPTH[3674]:D	LIGHTPATH TECHNOLOGIES INC	2 [4, 17]
BASI[8731]:D	BIOANALYTICAL SYSTEMS INC	1 [17]
NTWK[7372]:D	NETSOL TECHNOLOGIES INC	1 [5, 17, 43]
ETAK[4813]:D	ELEPHANT TALK COMM INC	2 [17, 44]
ATCV[3714]:D	ATC VENTURES GROUP INC	1 [44]

Note: *See SEC 302 Disclosure Controls.

The above table presents in *column 1*, the ticker [SIC designation]: [D or H] indicating the development or the holdback phases of the study. The shaded cells indicate the n = 11 HR or weakness flagged-firms; the other cells are the n = 9 LR firms, *column 2*, the legal firm name as used in the popular press and in *column 3*, the number of SEC reported weaknesses and SEC coding designations. For example, LPTH [3674] is LIGHTPATH Technologies, Inc. and was used in the developmental phase of the study. In 2008, LPTH was reported to have two (2) weaknesses coded as 4 and 17.