Privacy-preserving Information Sharing within an Audit Firm

Alexander Kogan and Cheng Yin

Introduction

This paper explores the possibility of sharing firm-level information within an audit firm in a privacy-preserving manner and of exploiting the benefits of doing this, under the assumption that the same audit firm serves multiple clients competing in the same industry. We are going to show how to use aggregated contemporaneous data from different clients without violating clients' confidentiality to achieve better accuracy prediction and error detection performance in the audit process. Specifically, this paper addresses three questions.

- *First*, do auditors benefit from sharing contemporaneous peer audit data within the same audit firm in a privacy-preserving manner?

- *Second*, how does the sharing scheme guarantee clients' privacy ?

- *Third,* how can we evaluate the reliability and validity of sharing data and detect "cheating" (providing fake audit data) behavior during the aggregation and / or encryption process?

Related Work & Research Questions

- *Literatures*:

- The effectiveness and usefulness of using peer firms as a benchmark.
- Based on the usefulness of peer firms, previous papers also have investigated the way of choosing peers (economicallycomparable firms).
- A number of both financial accounting and auditing studies have extensively examined the importance of information transfer and industry expertise in providing high-quality audits.
- <u>Research Questions</u>:
- The first research question examines whether models with aggregated peer information lead to different mean absolute percentage error (MAPE) in comparison to models that do not incorporate peer data.
- The purpose of the second research question is to test the error detection performance regarding different information sharing schemes.

Sample Selection

- <u>Data</u>:

- 20 industries that contained the largest number of firms and experienced various sales growth rates from 1991–2015 were selected through 4 digits SIC codes.
- Quarterly data of total revenues, cost of revenues, accounts receivable, and accounts payable was downloaded from the Compustat fundamentals quarterly database for the period 1991 – 2015.

SIC	Num- ber of Firms	Ac- count Paya- ble	Cost of Goods Sold	Account Receiva- ble	Reve- nue	Growt h Rate
7372	320.00	28.82	35.71	109.56	166.50	14.05%
6798	236.00	68.56	65.47	204.38	105.89	11.22%
1311	216.00	205.40	307.19	202.43	438.93	23.63%
7370	180.00	209.38	245.57	470.39	468.90	19.30%
2834	152.00	160.39	156.02	403.59	597.10	25.22%
3674	140.00	85.14	120.34	142.48	292.63	12.98%
4911	134.00	280.97	516.87	350.11	734.13	4.86%
5812	124.00	48.97	209.97	37.91	287.86	9.52%
7373	120.00	42.22	79.94	112.21	133.65	13.57%
2836	111.00	64.21	47.92	107.46	158.05	26.51%
3845	100.00	14.80	21.99	49.34	62.58	17.13%
4813	99.00	359.07	438.57	662.59	873.85	14.88%
3663	82.00	187.46	292.45	269.43	461.72	11.05%
3841	68.00	34.87	38.31	65.74	92.55	17.46%
9995	67.00	94.34	45.49	154.13	60.18	5.92%
7990	65.00	28.08	100.50	41.79	171.47	14.63%
3714	63.00	244.00	403.37	338.75	497.17	9.75%
6331	62.00	1807.07	975.98	3666.59	1143.98	10.15%
6211	60.00	9534.34	379.00	12763.90	741.76	12.74%
3576	58.00	35.81	96.10	151.49	267.99	7.86%
3661	54.00	15.84	27.89	39.87	54.83	12.98%

Research Design

- <u>Peer Selection</u>:
- Based on size rank and growth rate rank
- <u>Sharing Schemes</u>:
- Low-level sharing standardized errors from peer companies.
- Medium-level sharing standardized predicted value from peer companies.
- High-level sharing standardized true value from peer companies.

Without Sharing	Company Specific	Model
Sharing Schemes	Peer Sharing Mo	del
Low Level	-(ε_(i≠j))	3
Medium Level	y _p	У
High Level	y_i, i≠j	У

- Model Specification:

IND_(XXX_t)=	=(∑1_1^iZ_i)/i	
SALE_t= α + β _1	$[[SALE]]_{(t-12)+\beta}]_2$	[[AR]] _t+ε_t (1)
$COGS_t = \alpha + \beta_1$	$[[COGS] _(t-12)+\beta] _2$	$[AP] _t+\varepsilon_t (2)$
SALE_t= $\alpha + \beta_1$	$\llbracket \ \llbracket SALE \rrbracket \ _(t-12)+\beta \rrbracket \ _2$	$[AR] _t+ [[[IND_ERROR] _t+\epsilon]] _t (3)$
$COGS_t = \alpha + \beta_1$	$\llbracket \ \llbracket COGS \rrbracket \ _(t-12)+\beta \rrbracket \ _2$	$[AP] _t+ [[IND_ERROR] _t+\epsilon] _t (4)$
SALE_t= $\alpha + \beta_1$	$[[SALE]]_{(t-12)+\beta}]_2$	$[AR] _t+ [[IND_PREDICT] _t+\varepsilon] _t (5)$
$COGS_t = \alpha + \beta_1$	$\llbracket \ [COGS] _(t-12)+\beta] _2$	[[AP]] _t+ [[[[IND_PREDICT]] _t+ε]] _t (6
SALE_t= $\alpha + \beta_1$	$[[SALE]]_{(t-12)+\beta}]_2$	$[AR] _t+ [[IND_ACTUAL] _t+ \epsilon] _t (7)$
$COGS_t = \alpha + \beta_1$	[[[COGS]]_(t-12)+β]]_2	[[AP]] _t+ [[[[IND_ACTUAL]] _t+ε]] _t (8)

Results

- *Estimation Accuracy*:

SIC	Number of Firms	Mean of Mape from Original Model	Mean of Mape from Error Shar- ing Model	Mean of Mape from Prediction Sharing Model	Mean of Mape from Actual Sharing Model
7372	316	1.57	0.73	0.73	0.73
1311	212	4.54	2.00	1.85	2.06
7370	180	0.77	0.45	0.41	0.34
2834	150	0.58	0.42	0.34	0.38
3674	140	0.39	0.25	0.23	0.24
4911	126	0.31	0.18	0.19	0.18
5812	121	0.11	0.08	0.08	0.08
7373	120	0.33	0.20	0.21	0.20
2836	111	1.92	0.85	0.85	0.82
3845	100	0.53	0.31	0.29	0.31
4813	98	0.31	0.21	0.22	0.21
3663	82	0.28	0.21	0.20	0.20
4931	73	0.25	0.14	0.15	0.14
3841	68	0.44	0.29	0.24	0.28
9995	67	0.90	0.60	0.60	0.54
7990	65	0.30	0.19	0.16	0.17
3714	63	0.20	0.13	0.13	0.13
6331	62	0.30	0.21	0.20	0.21
6211	59	0.26	0.17	0.16	0.17
3576	58	0.27	0.18	0.18	0.18



Discussion and Limitations

- Sample Selection Risk:

• Companies with no peers after the first three years are dropped from the sample but still need to be audited. The firms that do not have uninterrupted peers error are also removed from our sample.

- Simulated Data Risk:

• The results in this study should be carefully applied. The findings in this study are based on interpolated data points and not on real data. Thus there may be some abnormal data in our data set causing serious problems (outlier values of MAPE). While these results support our conjectures of general prediction improvements and the superior performance of error detections.



A tale of two standards: A Pan-American study on XBRL standards for U.S. and Brazilian local governments Deniz Appelbaum, Hussein Issa and Steve Kozlowski

The DATA Act in the United States: Pilot Program:

- Affects states and local governments, charities and non-profits that receive federal grants or contracts
- Currently there exists little standardization
 across jurisdictions and recipients of ac counting and data standards
- Reports will be generated in a machinereadable transparent format

Federal Level:

 All agencies will report in one mandated data standard and format

U.S. and Brazil: Similarities and Differences

	Brazil:	United States:	Degree of similarity
Number of States	26 states	50 states	Somewhat similar
Format	Z	Single Audit format controlled by Fed- eral agencies with different require- ments	Not similar
Reporting structure	Reports structured by law LRF	Reports loosely structured by GASB	Not similar
Data format	PDF Reports	PDF Reports	Similar
Matching reports	QDCC doesn't match required annual forms	Single Audit filings often do not match other reports	Similar
Data availability	QDCC data is detailed and publicly available	Detailed checkbook data is just now being made available by several local governments	~ = becoming more similar
	Challenges:	Challenges:	Challenges
	Need to standardize chart of accounts in order to consolidate information	Accounting conventions will need to be uniform for grant reporting	Somewhat Similar
	Obsolete method of collection	Obsolete method of collection	Similar
	Objectives:	Objectives:	
	Replace obsolete data collection sys- tem	Replace obsolete data collection sys- tem	Similar
	Improve data quality	Improve data quality	Similar
	Simplify reporting process for entities	Simplify reporting process for entities	Similar

 Reports will be posted at one central web site www.usaspending.gov.

DATA Act Grantee Pilot Program Timeline:

Timeline of the Grantee Pilot Program:

Date	DATA ACT/PILOT PROGRAM EVENT
5/09/2014	DATA ACT passed into law
5/09/2015	Pilot Program begins with selected grant recipients
5/09/2017	Pilot program finishes
5/09/2018	OMB reports results to Congress
8/09/2018	OMB decides to require (or not) grant recipients to report in the format required by the DATA ACT

S.N		XBRL Item	
	Recommended Data Element	Туре	Data Definition
1	Federal Expenditures	Monetary Item	Amount of federal expenditure.
2	Catalog of federal domestic assis- tance numbers (CFDA)	String Item	The title of the program under which the Federal award was funded in the CFDA.
3	Pass Through Grantors Number	String Item	Pass through grantors number.
4	American Recovery and Reinvest- ment Act Applicable	Yes No Item Type	Denotes the American Recovery and Reinvestment Act applicability.
5	Program Title [Axis]	Dimension Item	Axis to represent program title
6	Agency [Axis]	Dimension Item	Axis to represent agency name
7	Grant Nature [Axis]	Dimension Item	Axis to represent grant nature
8	Grantor [Axis]	Dimension Item	Axis to represent grantor title
9	Basis of Presentation [Text Block]	Text Block Item	The entire disclosure for basis of presentation.
10	Sub recipients [Text Block]	Text Block Item	The entire disclosure for sub recipients.
11	Federal Direct Student Loan Pro- gram [Text Block]	Text Block Item	The entire disclosure for Federal Direct Student Loan Pro- gram.
12	Loan Program [Text Block]	Text Block Item	The entire disclosure for Federal Perkins Loan Program, Nursing Student Loan Program, and Income Contingent Loan.

XBRL: Machine and Human Readable

• Sample XBRL Output (Computer Readable)



Sample Review Report (Human Readable)

Home Insert Page Layout P	ormulas Deta	Review Vi	ew Acrobat									
X Cut Cariben + 30 Cariben + 30 Cariben + 30 ■ X U + 10 + Cariben + 30 ■ X U + 10 + Cariben + 30	· A A =	= = ×·	t Merge & Center -	S - % +	51.0	Conditional Formatting -	Format Cell as Table - Styles -	Intert De	ete Format	E AutoSum * 2 Fill * Sort Clear * Fills	à Find à	
87 - 6 20430	k0285			1023241			spas			Carrie		
A		c	D	-		() () () () () () () () () ()		G	н	1	1	1.
utgers, The State University of New Jers	tx .											
SCHEDULE OF EXPENDITURES OF FEDERAL	12 Months Ended	Prefix	TaxonomyD		Tel	pe .	Period Type	Scaling	Units	Balance Type	Negation	1
AWARDS (Details) (USD 5)	Aun. 30, 2013										1.1.1	
atalog Of Federal Domestic Assistance	r deral	agency	CatalogOffederalDomestic	Assistan abr	rlistringite	mType	duration			-	FALSE	Catalog Of Per
ederal Expenditures	\$704,360,285	agency	FederalExpenditures	abr	limonetan	(TernType	duration	0	USD N	debit	FALSE	Federal Expen
irect Grant [Metuber]									1			
ederal Expenditures	205,180,123	opency	FederalExpenditures	abr	ril monetary	iten Type	duration	0	USD	deb/t	FALSE	Federal Expen
gency For International Development		1										
ederal Expenditures	177,205	a and	FederalExpenditures	*0*	rikmonetan	iners/Type	duration	0	010	deb-t	FALSE	Federal Expen
entral Intelligence Agency [Member] [10.00		
ederal Expenditures	639,588	and .	repera Expenditures	1	rii monetari	interes yes	duration	0	050	ees e	74638	Federal Exper
Indexal Parent Long	747 589		Federal Exceptioner	1	di modetan		duration		1050	daha	F.M. 18	Testeral Tener
ational Science Foundation [Member]				1~								
ederal Expenditures	52,323,443	APPENCY	Federal Expenditures	1.00	ril monetary	(TerriType)	duration	0	uso	debit	PALSE	Federal Exper
ational Science Foundation [Member] Pass										1		
ederal Expeny Hures	3,827,125	agency	FederalExpenditures	*	rii monetary	tem?ype	duration	0	USD	debit	FALSE	Pederal Exper
egartment of Commerce [Member] Direct				1						1		
ederal Experilitures	4,852,137	agency	federal@xpenditures	str	monetary	terr "ype	duration	0	USD	and a state	FALSE	Federal Exper
epartment of Commerce [Member] Pass			and the second second second		1					1		
ederal Expenditures	1,507,570	agency	federalExpenditures	abr	ritmenetary	riters/Type	duration	0	050	des.	FALSE	Pederal Expen
epartment of Defense [Member] Direct									1200			
ederal Expenditures	17,892,002	agency	Patera Expenditures	10	TR. MORTON	intere Type	duration	0	050	orege	PALSE	Federal Exper
		1 M 10 10 10 10 10 10 10 10 10 10 10 10 10				14			10			
SUMMARY ROLEL ROLES ROL	3 / ROLE4 / RO	100 100			-							
	A Cut Cutors - JD Promat Pareter Promat Par	A Col Composed Fermat Parenter Fermat	A Cut Control Control	Cot Came 10 A* Image: Construction of the construction	A Cut Cutom 10 + 10 + 10 + 10 + 10 + 10 + 10 + 10 +	Cut Cut me 10 X Image: Cut me Cut me Cut me Cut me Fermat Pareter Image: Cut me <	Column Column	Column 10 X Image: Second Secon	Column 10 Column 10 Column Column <thcolumn< th=""> Column <thc< td=""><td>Column 10 A Image in the set of th</td><td>Column Column Colum Column Column</td><td>Column Column Column</td></thc<></thcolumn<>	Column 10 A Image in the set of th	Column Colum Column Column	Column Column



Moving towards continuous audit and big data with audit analytics: Implications for research and practice Deniz Appelbaum, Alexander Kogan and Miklos Vasarhelyi

Current Audit Landscape:

"Advances in technology and the massive proliferation of available information have created a new landscape for financial reporting. With investors now having access to a seemingly unlimited breadth and depth of information, the need has never been greater for the audit process to evolve by providing deeper and more relevant insights about an organization's financial condition and performance—while maintaining and continually improving audit quality.

Does this mean that core elements of the audit such as the current "pass/ fail opinion" that external auditors are mandated to provide—and that has served investors well for years, need to expand? Absolutely!" (Liddy, 2015)

AUDIT BY EXCEPTION:









Questions in this current audit landscape:

- 1. Should new (modern) analytics methods be used in the audit process?
- 2. Which of these methods are the most promising?
- 3. Where in the audit are these applicable?
- 4. Should auditing standards be changed to allow / facilitate these methods?
- 5. What are the competencies needed by auditors in this environment?

A few questions that should be addressed by research:

- 1. The appropriateness of the method for a particular audit function?
- 2. How can predictive technologies be used to set comparison models against which match actuals. How to set allowable variance (Vasarhelyi & Bumgartner, 2015)?
- 3. What additional verification processes would be desirable with the extant analytic technology?
- 4. What parts of the audit standards and processes must be progressively abandoned due to their obsolescence?
- 5. How can validation function be developed that link corporate information with big data variables to validate the dimensionality and predict variances?



Perform Audit Procedures on Similar Contracts Using Text Mining

Zhaokai Yan, Kevin Moffitt and Miklos Vasarhelyi

Introduction

Contracts have been widely used in commercial activities these days. However, among these contracts, frauds or errors may exist that will lead to a significant impact to the business. This contract problem has not been discussed or considered in the audit literature due to that (1) the nature and volume of contract makes it impossible for auditors to manually review all the contracts and that (2) accounting standards ignore future liabilities due to contract obligation. In this project, we try to contribute to the resolution of this contract audit problem by proposing one possible method to perform audit procedures on similar contracts.

Framework

Some contracts ("similar contracts") are created by filling a fixed contract template ("template") with varying information ("variables"). By extracting and examining the "variables" part, auditors are allowed to perform audit procedures to identify anomaly or economic effects from these similar contracts.

- Use TF-IDF score and cosine similarity to determine which contracts can be classified as similar contracts
- Compare similar contracts to identify the identical words or sentences among contracts and create the "template"
- Compare other similar contracts against this "template" to detect the "variables"

Methodology

Sample: 10 reinsurance contracts

TF-IDF: stands for "Term frequency—inverse document frequency".

- Remove stopwords (e.g. "am", "and", "the", "it")
- Stem words (e.g. "run, ran, running" to "run")
- Term Frequency (TF) = (Number of times term t appears in a document) / (Total number of terms in the document)
- Inverse Document Frequency (IDF) = log(Total number of documents / Number of documents with term t in it)
- TF-IDF = TF \times IDF

Cosine Similarity: the similarity between any two contracts (A, B) can be calculated as: The attribute vectors A and B are the TF-IDF vector of contract A and B.

Similarity =
$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

• Perform audit procedures on "variables"

Limitation

- Many errors occurred when converting contracts from PDF to text files.
- Lack of sample data. Only 10 contracts have been analyzed in this study.

Future Work

- Identify variances from the contracts like handwritten annotations, deleted and added paragraph, etc.
- Some contracts contain handwritten paragraph. Algorithm for image similarity may be applied to analyze such content.
- Embedding this type of multicontractural examination into continuous auditing environment.

Preliminary Result

S

			Cosi	ine Similarit	y Score Bet	ween 10 Co	ntracts			
	Contract 1	Contract 2	Contract 3	Contract 4	Contract 5	Contract 6	Contract 7	Contract 8	Contract 9	Contract 10
Contract 1	100%	99.951%	99.951%	99.950%	99.855%	99.770%	99.760%	99.766%	99.767%	99.759%
Contract 2		100%	99.945%	99.950%	99.861%	99.773%	99.762%	99.772%	99.771%	99.761%
Contract 3			100%	99.950%	99.850%	99.767%	99.757%	99.769%	99.770%	99.759%
Contract 4				100%	99.853%	99.778%	99.763%	99.770%	99.774%	99.768%
Contract 5					100%	99.673%	99.709%	99.701%	99.699%	99.680%
Contract 6	i i					100%	99.915%	99.912%	99.914%	99.912%
Contract 7	,						100%	99.950%	99.948%	99.927%
Contract 8								100%	99.961%	99.933%
Contract 9									100%	99.937%
Contract 1	0									100%

	Variables of Contrasts																
							variables	or contrac	15					1	1		
		Period of	Insurance				Limit of	Liability				Roton	tion		Retroact	Rotroactivo dato	
		I CHOU OI	mourance		Professior	nal Liability		G	eneral Liabilit	ty		neter			netroac		
Contract Number	Policy Number	From	То	Policy Aggregate	Each claim	Each Location Aggregate	Each claim	Each Location Aggregate	Products Completed Operations Aggregate	Personal and Advertisin g Injury Limit	Damage To Premises Rented to You Limit	Profession al Liability	General Liability	Gross premium USD\$	Profession al Liability	General Liability	
1	A4DB12012013	12/1/2013	12/1/2014	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$0	\$0	NA	11/1/2000	11/2/2000	
2	M0DB12012013	12/1/2013	12/1/2014	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$0	\$0	NA	5/1/2004	5/1/2004	
3	A11DB12012013	12/1/2013	12/1/2014	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$0	\$0	NA	11/1/2000	11/1/2000	
4	A12DB12012013	12/1/2013	12/1/2014	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$0	\$0	NA	11/1/2000	11/1/2000	
5	A17DB12012013	12/1/2013	12/1/2014	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$0	\$0	NA	6/1/2001	6/1/2001	
6	NA	12/1/2013	12/1/2014	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$0	\$0	NA	11/112000	11/1/2000	
7	NA	12/1/2013	12/1/2014	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$0	\$0	NA	5/1/2004	5/1/2004	
8	NA	NA	NA	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$0	\$0	NA	11/1/2000	11/1/2000	
9	NA	NA	NA	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$0	\$0	NA	11/1/2000	11/1/2000	



An Automated Rule-Based Approach for the Measurement of **Internal Control Effectiveness**

Abdulrahman Alrefai

Introduction

With the passing of the Sarbanes-Oxley Act of 2002, the focus on internal controls has tremendously increased, and now firms are required to report on the effectiveness of their internal control systems. The importance of having an adequate internal control system cannot be overstated. Prior research has found that when internal controls are weak, there is an increased likelihood of earnings manipulation by management (Chan et al. 2008; Ashbaugh-Skaife et al. 2008). Hence, the evaluation of internal controls has always been an important aspect to both auditors and management.

The aim and contribution of this study is to provide a methodology whereby the effectiveness of internal controls can be measured. Specifically, the paper presents a conceptual model that illustrates how a rule-based system can be used to test internal controls, and then be used to provide an overall measure of how effective the internal control system for a business process. Basically, the system attempts to run tests on a dataset relative to a specific audit function, produce results, and based on those results, provide a formalized measure for the effectiveness of the internal control system.

Literature Review

- Research has shown that qualitative methods are considered to be insufficient, and the assessments generated by qualitative methods alone are insufficient for developing comprehensive internal control evaluation models (Yu & Neter 1973; Mock & Turner, 1981; and Bierstaker and Wright, 2004).
- Since computers have the advantage of speed, accuracy, and great memory capacity in addition to the ability of evaluating large and complex models, a systematic internal control model should be introduced to aid auditors or management in evaluating internal control systems instead of the traditional manual methods (Bailey et al. 1985).
- The consideration of the whole population of transactions in testing can enhance the effectiveness of an audit and increases the probability that material errors, omissions, fraud, and internal control violations may be detected (Chan and Vasarhelyi 2011).
- determining the reliability of a control consists of aggregating the possibilities that the control is applied (compliance) and that it is effective (design) (Srinidhi and Vasarhelyi 1989).

Methodology

Analysis

After studying the business process and the controls that need to be implemented, the next step was to get familiarized with the data on hand.

The data used in this study relates to employee procurement cards, obtained from a multi-national firm, and averaging about 50,000 monthly transactions with 55 attributes.

The conceptual model developed in this study is intended to measure the effectiveness of internal controls on a transactional level. Particularly, it utilizes a rule-based system to test the effectiveness of controls in a business process, and then provide a score on the adequacy of that business process.



Several key controls were identified that need to be instituted by any firm adopting procurement cards to either deter or detect misuse a rule-based system was used on the dataset to generate exceptions and serve as the basis for the measurement of effectiveness of controls

Results

Based on the testing and measuring the effectiveness of the transactional level controls for the procurement card business process, the results indicate that it would get a score of 0.726 for the overall effectiveness of the internal control system for that business process. This is indicative of a deficient internal control system.

Measured Control	EoC	Level of Effectiveness
Transactional Limit	0.997	Effective
Daily Limit	0	Missing
Monthly Limit	0.999	Effective
Item Description	0.542	Significantly Deficient
Merchant Information	0.998	Effective
Purchase Dates	0.543	Significantly Deficient
MCC Restrictions	1	Effective

Applying Process Mining to Auditing: A Comparison of Process Mining Techniques

Tiffany Chiu and Miklos A. Vasarhelyi

<u>Apply Process Mining Techniques to</u> <u>Auditing</u>

- Van der Aalst et al. (2010) indicated that event log and process mining techniques enable new forms of auditing. For example, the alpha process mining algorithm can automatically extract a Petri net that concisely models behavior in the event log; in this case, the auditors can have an unbiased view of what has actually happened in the company.
- Bukhsh and Weigand (2012) indicated that process mining techniques can be applied to detect bottlenecks, examine conformance of processes, predict execution problems, and monitor deviations (e.g., comparing the observed events with predefined models or business rules).
- By applying process mining techniques and tools, one can analyze control flows, authorization rules, business data models, organizational models and business rules (Van der Aalst et al. 2003; Rozinat and Van der Aalst 2006; Rozinat and Van der Aalst 2008; Bukhsh and Weigand 2012).
- There are two main advantages of using event logs in auditing: (1) it provides the auditor with more data, (2) it provides a human-independent way of

Apply Process Mining Techniques to Audit Fraud Detection

- Jans et al. (2013) indicated that process mining of event logs can add value to auditing: (1) examines the entire population of data; (2) Event logs have been automatically recorded by the system rather than entered by the auditees; (3) Allows auditors to conduct audit procedures that are not possible with current audit tools; (4) Enables auditors to implement the audit risk model more effectively by providing effective ways of conducting the required walkthroughs of processes and conducting analytical procedures.
- Yang and Hwang (2006) applied process mining techniques to detect the potential fraudulent and abusive cases in healthcare service. The authors proposed a detection model using frequent patterns from clinical instances; the result indicated that the proposed model is capable of identifying several fraudulent and abusive cases which have not been detected by traditional methods.
- Jans et al. (2014) applied process mining techniques to auditing analytical procedures. The result indicated that process mining techniques can detect anomalous transactions that traditional audit analytical pro-

<u>Apply Process Mining Techniques to</u> <u>Accounting Data</u>

- Over the past few years, internal and external auditors have emphasized process audits.
- Traditionally, process auditors have relied on manually modeled processes based on interviews and additional information provided by the client; however, these methods are error-prone and timeconsuming.
- Mueller-Wickop and Schultz (2013) proposed an algorithm that determines an activity sequence from accounting data. By applying this algorithm, mined process instance graphs can be decomposed in a way that fit into sequential event log formats. As a result, event log based process mining techniques can be used to construct process models.
- Werner and Nuttgens (2014) proposed a novel process mining technique that is able to construct process models using the data dependent order of events instead of temporal order of events which has been widely applied in contemporary general purpose mining algorithms.

recording data (Jans et al. 2010; Bukhsh and Weigand 2012).

cedures fail to discover.



Petri Net - Alpha algorithm	Fuzzy Miner	Inductive Visual Miner
 Alpha algorithm is the first process mining algorithm to address the 	 Fuzzy Miner is able to deal with real life dataset. 	 IVM focuses on process exploration, which is the process of repeatedly trying
discovery and concurrency of process model (van der Aalst and van	 When dealing with unstructured processes, traditional process mining 	settings until a satisfactory model is discovered (Leemans et al 2014).
Dongen 2013).	techniques often discover	 Many existing process discovery
 Captures the patterns from the event logs and presents the process model of the given dataset. 	 Fuzzy miner is not limited to re-discovering what we have already 	parts of the event log based on chosen parameters.
 Cannot deal with real life dataset: has problem with noise, infrequent/incomplete behaviors and complex routing constructs. 	known in the past, it can be used to unveil previously hidden knowledge in the event logs (Gunther and Van der Aalst 2007).	 IVM enables setting different parameters and filters in the process exploration, it also supports choosing different perspectives.



Predicting Risk of Credit Card Delinquency: An Application of C&RT Ting Sun, Jun Dai and Miklos A. Vasarhelyi

Introduction

Approximately 1 out of 20 Americans with credit files are at least 30 days late on a credit card or other non-mortgage bill payment, as revealed by the Urban Institute (Ratcliffe et al., 2014). Credit card delinquency affects the health of credit card industry (Ausubel, 1997; Gross & Souleles, 2002; Holmes& Ghahremani, 2015). Thus it is essential to model and predict the risk of card delinquency for card-issuing banks or other financial institution, threatened by the resulting financial distress (Lin, 2009).

Different from previous study, which mainly compared the performance of various data mining approaches in predicting default or other related risk of credit card, this study, by using a real-life credit card dataset from a bank, aims at applying C&RT to explore predictors for the occurrence credit card delinquency.

Related Research

- Haughton & Oulabi (1997) explored several technical aspects of classification and regression tree (C&RT) and chi-square Automatic Interaction Detector (CHAID) models in predicting response to a direct mail package as an example of a solicitation. They found that the response lifts are very close for both types of models.
- Lee et al. (2006) demonstrated the effectiveness of credit scoring using C&RT and multivariate adaptive regression splines (MARS).
- Yeh & Lien (2009) examined the six major classification techniques in data mining and compared the performance of classification and predictive accuracy among them for forecasting default risk of credit card clients. Their results showed that artificial neural network is the only one that can accurately estimate the real probability of default.
- Nie et al. (2011) built a churn prediction model for credit card holders, using logistic regression

Methodology and Data

Method: C&RT

C&RT is a classical decision tree algorithm introduced by Breiman et al.(1984). It is primarily used in situations when the objective is to classify an object into two or more populations and a large number of potential variables are considered. Both categorical and continuous variables can be used as explanatory variables in C&RT model. (Lee et al., 2006).

Description of Dataset

1. The data source

An anonymous major bank in South America provides the data for this study.

2. The data structure

The data we are using contain two subsets (1) detailed transaction information in July 2013 and (2) the personal information as recorded in September 2003 (this subset includes our dependent variable, Indicator).

and decision tree approaches.

3. Data size

The original dataset are at account level. After integrated them at the level of client , we have 11,087 obs. blocked due to delinquency and 723,438 normal obs. Since we are interested in how personal characteristics and spending behaviors relate to the possibility of credit card delinquency, we focus on clients who have no "late payment" records in July 2013. After dropping obs. with late payment records. We have 6,537 observations blocked due to delinquency and 704,860 normal observations.

4. Data under-sampling

50% of data were used to train the model. The remaining data are testing set. To reduce data imbalance in the training data, which had 3,230 positive (blocked) obs. and 352, 072 negative (normal) obs., we randomly select 3158 obs. From the negative records in original training data. The new training data have 3, 230 positive obs. and 3158 negative obs.

Variables

Dependent Variable

Indicator: It equals 1 if the account of the client was blocked due to credit card delinquency in September 2013, and 0 otherwise

Preliminary Results

Results from C&RT model (when we use Boosting and set Misclassification Cost = 30):



Results from testing data:



Discussion

The overall classification accuracy is 99.48%. In assessing the performance of a model, another important consideration is the Type I and Type II error rates. A Type I error is committed when a client who commit credit card delinquency (Positive client) is classified as a normal client (Negative client). A Type II error is committed when a negative client is classified as positive one .

True Positive =2,155 True Negative=352,089 False Positive =699

False Negative =1,152

Type I error rate is 0.3%, while Type II error rate is 0.2%.

Precision=0.76. It means all the clients identified as positive ones by the classifier, 76% of them have been identified correctly.

Recall (or sensitivity)=0.65. It means , for all true positive clients, 65% have been correctly identified by the classifier.

F-measure=0.7

Big Data as Audit Evidence: Utilizing Weather Indicators

Kyunghee Yoon and Alexander Kogan

APs and Disaggregated Data

Kogan *et al.* (2010) compare the widest range of statistical models and find that VAR models and linear regression models tend to perform better than others. Additionally, previous literature indicates that disaggregated model (micro-level) is likely to deliver better performance than monthly, aggregated level models on segment or product line balance (macrolevel) on APs. Knechel 1988; Dzeng 1994; Allen et al. 1999).

H1: Firm-wide sales expectation developed from disaggregated individual location with peer stores data produce more accurate and more precise expectation than firm-wide sales expectation derived from aggregated firm-level data.

APs and NFI

SAS No 56 (AICPA 1988) suggests Nonfinancial information (NFI) should be considered when performing APs (AICPA 2002, 2007). According to SAS 56 (AICPA 1988) during APs to develop expectations of accounts factors such as financial data from prior periods, client financial budgets, and industry information could be used. Especially, it recommends analyzing the relation between financial information and NFI.

Big data audit evidence provides more disaggregated data such as daily or weekly updated information than NFI suggested in previous literature. In addition, sources of big data would be the accessibility of information.

H2: The model with big data produces more accurate and more precise prediction than the model without financial information.

Method (1/2)

1. Data

The data employed in this research was obtained from one of the world-wide served audit firms. The targeted firm is a multiplication service firm with homogeneous operation in the world, but in this research only observations from the U.S. are used. A total 24 monthly observations are provided, and especially it is for about 2,000 operating unit locations from fiscal year 2011 to fiscal year 2012.

2. Big Data

Weather information such as daily precipitation and maximum temperature is utilized as nonfinancial information because in particularly retail industry sales amounts are likely to be affected by weather condition (Engle et al. 1986; Maunder 1973; Starr-McCluer 2000).

Method (2/2)

3. Control Variables

This study is extended by the studies of Kogan et al.(2010) and Allen et al.(1999). Basically, there are two kinds of models tested in this study- the multivariate regression models and the vector autoregressive models. The store level model is supposed to have about 2,000 predictors which are observations from the other stores on the models, but too many independent variables causes full rank issues. Therefore, only highly correlated predictors are selected by stepwise selection methods.

4. Evaluation of models

MAPE= Abs (actual value –predicted value)/ actual value

Each model generates one-step ahead forecast by rolling forecast.

Prediction Model

Level	Model Descrip-	Model Specifications			
	tion				
Panel A: Models Without NFI					
Weekly	Vector Auto-	$X_{wt,1} = \alpha + \beta_1 X_{wt-1,1} \dots$			
Daily	regression	$X_{dt,1} = \alpha + \beta_1 X_{dt-1,1} \dots$			
Panel B: N	Models With NFI ir	n a Firm-Wide Level			
Weekly	Multivariate	$X_{wt,1} = \alpha + \beta_1 A_{t,1}$			
Daily	Regression	$X_{dt,1} = \alpha + \beta_1 A_{t,1}$			
Weekly	Vector Auto-	$X_{wt,1} = \alpha + \beta_1 X_{wt-1} + \dots + \beta_3 A_t$			
Daily	regression	$X_{dt,1} = \alpha + \beta_1 X_{dt-1} + \dots + \beta_3 A_t$			
Panel C: N	Models in Store le	vel Data			
Weekly	Multivariate	$X_{wt,1} = \alpha + \beta_1 X_{wt,2} + \cdots$			
Daily	Regression	$X_{dt,1} = \alpha + \beta_1 X_{dt,2} + \cdots$			
Weekly	Vector Auto-	$X_{dt,1} = \alpha + \beta_1 X_{wt,2-1} + \cdots$			
Daily	regression	$X_{dt,1} = \alpha + \beta_1 X_{dt,2-1} + \cdots$			
Panel D: I	Panel D: Models With NFI in Store level Data				
Weekly	Multivariate	$X_{wt,1} = \alpha + \beta_1 X_{wt,2} + \dots + \beta_3 A_t$			
Daily	Regression	$X_{dt,1} = \alpha + \beta_1 X_{dt,2} + \dots + \beta_3 A_t$			
Weekly	Vector Auto-	$X_{dt,1} = \alpha + \beta_1 X_{wt-1,2} + \dots + \beta_3 A_t$			
Daily	regression	$X_{dt,1} = \alpha + \beta_1 X_{dt-1,2} + \dots + \beta_3 A_t$			

Preliminary Results

Preliminary results show that as previous studies show high frequent data provides more accurate and precise expectations. And also, generally including weather indicators improves predictive powers.

		Store Level
	Firm Level (Aggregate Model)	(Disaggregate Mod-
		el)
Model	MAPE	MAPE
AR (1-7)	0.1097	0.1008
		With peer stores and
	With peer stores	
		weather indicators
Model	MAPE	MAPE
Regression	0.0594	0.0195
AR(1-7)	0.0724	0.1603



Continuous Auditing of Cloud Computing: A Privacy-preserving Data Security Design

Alexander Kogan and Yunsen Wang

Introduction

- In the age of big data, cloud computing is emerging as a promising IT architecture that has been widely researched and deployed both in the academic and industrial areas. However, the implementation of cloud service causes the separation of ownership and management of corporate data. As a result, the adoption of cloud computing has given rise to cloud users' concerns about data security and privacy.
- Moreover, since information technology (IT) auditors are responsible for the security and integrity of corporate data, the adoption of cloud auditing models requires them to extend their range of responsibility from the local enterprise information system to the cloud.
- We discuss the Cloud-based Continuous Auditing design able to automatically detect unauthorized modifications of business transaction data using homomorphic authenticators. If a manager tempers with the transactional data to manipulate earnings, the challenge-and-response protocol will provide the auditor with the metadata showing the trace of modification. This design enables the auditor to detect fraud without having to download

Research Objectives

• Basically, the companies run the enterprise information systems on the cloud, and store the data in the cloud database. The auditors have the access to audit companies' data on the cloud.



- In the Third Party Audit (TPA) model proposed by Wang et al. (2010), from technology perspective, they discussed the threats from cloud service provider against integrity and confidentiality of the remote stored data.
- We focus on the fraud behavior by the company. It is possible that a manager may collude with an IT employee to manipulate the historical transaction records to prepare a fraudulent report. This design allows the auditor to detect fraud without

Literature Review

Since Vasarhelyi and Halper initially developed the first practical continuous audit (CA) system in 1991, the CA research field has flourished with numerous advances such as novel analytical algorithms and CA system architecture designs (Kogan et al, 1999, Vasarhelyi et al. 2004, Alles et al. 2008, Issa 2013). The emergence of cloud computing creates new opportunities and challenges for continuous auditing.

Cloud Auditing	Protocol Instances	Relative
Models		Literature
Two party remote checking model	Remote IntegrityChecking (RIC)Protocols	Blum et al. (1991) Deswarte et al (2004) Sebé et al. (2008)
	Proof of Retrievabil- ity (POR) protocols	Juels and Kaliski (2007), Shacham and Waters (2008)
	Provable Data Possession (PDP) protocols	Ateniese et al. (2007, 2008)
Third party remo checking model	te Third Party Auditing (TPA)	Wang et al. (2009)
Advanced Third Party Auditing Protocols	Continuous auditing	Wang et al. (2011) Erway et al. (2009)
	Batch auditing	Zhu et al. (2012) Wang et al. (2011)
	Security and Privacy concern	Wang et al. (2009) Wang et al. (2010) Yang and Jia (2011)

the entire financial dataset.

looking through the details in the transactional data.

Framework of Cloud based Continuous Auditing System (CCAS)

Key Generation

- File generation and outsourcing
 - From time to time, the auditee's enterprise information system, e.g. ERP, generates a file of transaction data *F*.
- Blocks division
 - Given the file F, the owner first splits F into n blocks and s sectors $\{m_{ij}\}(1 \le i \le n, 1 \le j \le s)$ for each block.

•
$$F = \begin{pmatrix} m_{11} & \dots & m_{1s} \\ \vdots & \ddots & \vdots \\ m_{n1} & \dots & m_{ns} \end{pmatrix}$$

- Key generation
 - The auditee chooses s random numbers $x_1, x_2, ..., x_s \leftarrow \mathbb{Z}_N$ as the private key of this file sk = (x). She also chooses j random element $u_1, u_2, ..., u_j \leftarrow \mathbb{G}$, and computes $v_i = g^{x_i} \in \mathbb{G}$ and $w_i = u^{x_i}$ as the public key pk = (v, w, g, u).

Signature Generation

• Signature Generation:

Includes tag computation and file outsourcing. The auditor downloads the dynamically computed tags that could be used for comparison.

- Tag computation
 - Let $H: \{0,1\}^* \to \mathbb{Z}_N^*$ be a hash function, where $x = \sum_{i=1}^s x_i$ for all $x_i (i = 1,2,...,n)$. Let F_{nt} be the file name and time stamp. For each block m_i , the auditor computes the tag $\sigma_i = (H(F_{nt}|i) \cdot \prod_{j=1}^s u_j^{m_{ij}})^x \mod N$.

• Proof Generation: the auditor

of the outsourced files.

rity of the files.

sends the challenge along with the

public key to check the integrity

• Proof Verification: CSP (Cloud

Service Provider) generates the

proof message to prove the integ-

- Data storage
 - The file F is stored on the cloud together with the tags $\{\sigma_i\}.$

Proof Generation and Proof Verification

- Challenge
 - The auditor randomly chooses a subset of data block $Q = \{s_1, s_2, ..., s_c\}$ from set [1, n] and use the public key pk for each selected data block m_i . Then, it sends a challenge $chal = \{(i, v_i)\}_{i \in Q}$ to the server.
- Proof
 - the server computes and sends back both the tag proof $\sigma = \prod_{i \in Q} \sigma_i^{v_i} \mod N$ and the data proof $\mu' = \sum_{i \in I} v_i \cdot m_i$.

$$e\left(\sigma,g\right) = e\left(\prod_{i=s_1}^{s_c} H\left(i\right)^{v_i} \cdot u^{\mu}, v\right)$$

File updating

• Whenever the transaction data file is updated, the tags could be accordingly updated and downloaded by the auditor.

• Key Generation:

Run by the auditee to set up the auditing delegation. The auditee first generates a pair of keys, public key and private key, then shares the public key with the auditor.

Contribution and Further Research

- Our paper is the first research that applies Third Party Auditing technology to continuous auditing. The main contribution is that the CCA design creates a relationship between the cloud computing and continuous auditing disciplines. The CCA design can be seen as a supplementary tool that enables auditors to assess the financial report files on the cloud in a more economical and dynamic way.
- There are three suggestions for further research: (1) The prototype of the CCA design could be programmed to simulate the fraud detection process; (2) Measure the overhead of the CCA design so that the program can be optimized; (3) The security and auditability of the CCA design need rigorous proof in theoretical method.
- This design enables the auditor to detect fraud without downloading the entire population of financial records and ensures data security and privacy.
 - RUTGERS Rutgers Business School

Newark and New Brunswick

Text Mining to Uncover the Severity of SEC Comment Letters Yue Liu and Kevin Moffitt

Introduction

The Security and Exchange Commission (SEC) comment letter is the correspondence between SEC staff and SEC filers about the filers' public information disclosure. The SEC periodically reviews the filings of public companies to evaluate the adequacy of disclosures and the filings' compliance with Generally Accepted Accounting Principles (GAAP) and issues comment letters to companies whose filings are perceived to be somehow deficient. Severity of comment letters can be a reflection of perceived deficiencies in the filings, and may provide information about the disclosure quality and cost of comment remediation.

This paper aims at using text mining to uncover the severity of SEC comment letters. Specifically, a contextual measure of severity based on tones (strong or weak) of comment letters is developed. Empirical analysis is conducted on a sample of initial comment letters related to 10-K filings. Preliminary result shows that the proposed severity measure (the decile rank of strong word ration of the comment letter) is positively associated with the probability of restate-

SEC Comment Letter

According to SOX Section 408, the SEC's Division of Corporation Finance shall review registrants' filings at least once every three years to evaluate the filings' compliance with SEC and GAAP disclosure requirements and to ensure the quality of information presented to stakeholders. As a result of the review, the SEC staff will issue a comment letter to companies whose filings are regarded as deficient and could be improved in some way.

The company receiving the comment letter is required to respond within 10 business days. Upon receiving the response, the staff will review the response letter and see whether the response is satisfactory. If not, additional comment letters will be issued to the filer until the SEC is satisfied with the response and issues a "no further comment" letter.

The SEC comment letter does not represent the official opinions of the SEC and is inquisitorial in nature; however, it could result in serious consequences such as definitive restatement of the reviewed filing (Cassell et al., 2013). In addition, comment letter remediation is a costly process in terms of the time and resources it takes to make the remediation. The

Literature Review

•Prior literature on SEC comment letter has focused on the determinants and consequences of receiving comment letters for IPO filings, 8-K filings, and 10-K filings, etc.

•Research that conducted analysis on comment letter severity:

	Findings regarding comment letter severity	Severity measure
Chen and Johnston (2010)	higher comment letter severity could result in more disclosure quality improvement	conversation time
Cassell et al. (2013)	comment letter severity is positive- ly related to the probability that the comment letter will result in a re- statement	number of com- ment topics
Gietzmann et al. (2015)	comment letter severity moderates the impact of comment letter re- view on CFO turnovers; there will be a permanent shift up of the CFO turnover once the firm receives a severe comment letter	conversation time

•All the prior studies that investigate comment letter severity choose proxies such as the number of topics or the conversation time, and none of them uses textual measures.

•However, the content of comment letters could also provide some information about the comment letter se-

ment of the reviewed 10-K filings.

Measure Development and Research Method

•Severity Measure

This paper focuses on the comment letters related to 10-K filings, and only uses the initial letter for empirical analysis.

Loughran and McDonald's Modal Strong and Model Weak word lists are used to identify strong/weak

words in the comment letters. Strong/weak word ratio is calculated as the number of strong/weak words divided by the total number of words in the clean texts; further, the decile rank of the comment letter's strong/weak word ratio is calculated so that comment letters in the decile having the highest strong/weak word ratio are assigned a value of 10 and so on down to 1 for comment letters with the lowest strong/weak word ratio.

•Research Model and Sample Selection

Using the following logistic regression to study the association between comment letter severity and restatement of reviewed 10-K filings:

 $restatement = \beta_0 + \beta_1 strongrank + \beta_2 weakrank + \beta_3 icw + \beta_4 sz + \beta_5 companyage + \beta_6 loss + \beta_7 bankrupt cyrank + \beta_8 growth + \beta_9 m \& a + \beta_{10} restructuring + \beta_{11} litigation + \beta_8 growth + \beta_9 m \& a + \beta_{10} restructuring + \beta_{11} litigation + \beta_8 growth + \beta_9 m \& a + \beta_{10} restructuring + \beta_{11} litigation + \beta_8 growth + \beta_9 m \& a + \beta_{10} restructuring + \beta_{11} litigation + \beta_8 growth + \beta_9 m \& a + \beta_{10} restructuring + \beta_{11} litigation + \beta_8 growth + \beta_9 m \& a + \beta_{10} restructuring + \beta_{11} litigation + \beta_8 growth + \beta_9 m \& a + \beta_{10} restructuring + \beta_{11} litigation + \beta_{10} restructuring + \beta_{11} litigation + \beta_{10} restructuring + \beta_{11} litigation + \beta_{10} restructuring + \beta_{10} restructuring + \beta_{11} litigation + \beta_{10} restructuring + \beta_{11} litigation + \beta_{10} restructuring + \beta_{11} litigation + \beta_{11} restructuring + \beta_{11} restr$

 $\beta_{12}big4 + \beta_{13} \sec ond - tier + \beta_{14}resign + \beta_{15}dismiss + yeardummy + industry dummy$

Data used in the empirical analysis are obtained from SEEKINF website, Compustat, and Audit Analytics. Final dataset includes 10,109 firm-year observations. more severe the comment letter is, the higher cost of remediation could be.

Preliminary Result

Table 1. Summary statistics for the sample

Variable	Obs	Mean	Std	Median	Min	Max
restatement	10109	0.108	0.311	0.000	0.000	1.000
strongratio	10109	0.003	0.002	0.003	0.000	0.013
weakratio	10109	0.009	0.003	0.008	0.000	0.024
strongrank	10109	5.503	2.875	6.000	1.000	10.000
weakrank	10109	5.528	2.869	6.000	1.000	10.000
icw	10109	0.095	0.293	0.000	0.000	1.000
SZ	10109	6.546	2.286	1.000	-4.096	13.348
companyage	10109	1.101	0.301	0.000	1.000	2.000
loss	10109	0.305	0.460	5.000	0.000	1.000
bankrupt-						
cyrank	10109	4.840	3.037	0.068	1.000	10.000
growth	10109	0.440	8.200	0.000	-12.937	438.000
m&a	10109	0.159	0.365	0.000	0.000	1.000
restructuring	10109	0.308	0.462	0.000	0.000	1.000
litigation	10109	0.301	0.459	0.000	0.000	1.000
big4	10109	0.725	0.446	1.000	0.000	1.000
secondtier	10109	0.069	0.254	0.000	0.000	1.000
resign	10109	0.015	0.120	0.000	0.000	1.000
dismiss	10109	0.049	0.217	0.000	0.000	1.000
		COG	efficient	t-value)	p-value
strongrank		0	.025*	2.16		0.03
weakrank		-	0.019	-1.62		0.11
icw		0	.270*	2.5		0.01
SZ		-	0.027	-1.2		0.23
companyage		(0.142	0.55		0.58
loss			0.117	1.39		0.17
bankruptcyrar	nk	-	0.024	-1.75		0.08
growth			0.000			0.96
m&a			0.071			0.48
restructuring			0.01			0.90
litigation			-0.07			0.59
big4		0	0.262*			0.03
secondtier			0.277			0.06
resign			0.015			0.96
dismiss		0.21		1.48		0.14

-2.206*

constant

n

year dummy industry dummy

Pseudo R2

-2.03

controlled

controlled 10052

0.0308

0.04

verity.

Limitation and Future Research

•Limitations

 \rightarrow First, some comment letters review several different filings simultaneously, but in this study the strong word ratio is calculated on the overall letter level rather than based on the specific 10-K related comments due to inability to identify exact comments on 10-K filings. However, even the overall comment letter severity level is found to be positively related to the probability of restatement of 10-K filings, suggesting that the effect could be stronger on the specific 10-K related comment level.

 \rightarrow Second, this study applies the Loughran and McDonald's Modal Strong and Modal Weak word lists directly in the severity measure development. However, comment letters may be somehow different from other financial filings; as a result, the word list may not be perfectly suitable to capture the severity of comment letters.

•Future Research

Future study could extend the current study by (1) Developing a severity measure for the specific 10

-K related comments;

(2) Modifying the word lists to better fit the comment letter situation.



Designing CA/CM to fit Not-for-Profit Organizations Deniz Appelbaum, Stephen Kozlowski, Miklos Vasarhelyi and Joel White

Background

There has been much public attention recently regarding fraud and abuse in the not-for-profit (NFP) sector, such as recent articles appearing in the Washington Post (Zapotsky, 2013). The NFP sector has been subject to increasing pressure to conduct internal audits in a more effective and efficient manner. Many large and generally publicly-held organizations have implemented Continuous Auditing and Monitoring (CA/CM) tools into their internal and managerial activities to some degree (Alles et al, 2008), to support their internal audit function. Small-to -Medium sized (SME) organizations, and especially NFPs, have not had an opportunity to participate in the benefits of CA/CM technology due to cost/benefit concerns, a general lack of product designed and priced for smaller organizations, and a lack of technical expertise in the organization.

Objective

The overall purpose of this Rutgers Accounting Research Center (RARC) case study is to undertake a CA/CM implementation by working with Small-to-Medium sized (SME) NFP organizations. The intent of this case study is to explore whether or not existing CA/CM tools and techniques can be implemented and generate appropriate analyses in an SME NFP setting. This paper discusses a case study where the project team and case NFP successfully implemented CA/CM protocols in the payroll/HR process. The project team envisions that this case study may serve as a benchmark for CA/ CM adoption by NFP organizations as well as by consulting organizations that support their NFP clients in undertaking such implementations.

Methodology

The Case Study approach, as outlined by Robert K. Yin (Yin, 2012, 2009, 1999, and 1984), assists in the understanding of complex issues or objects by extending current experience to previous research. Yin defines the case study method as an empirical investigation of a relevant issue in the context of a real life application. Yin (2012 and 1984) and other researchers (Stake, 1995; Eisenhardt, 1989; Simons, 1980) have suggested techniques for case studies, which may be combined as follows (Soy, 1997):

- Determine and define the research questions
- Select the case and determine data gathering and analysis techniques
- Prepare to collect the data
- Collect data in the field
- Evaluate and analyze the data

Project Execution Steps



Test Results-Graphic Representation



<u>Example Tests for</u> <u>Payroll/HR</u>

The initial list of potential tests is drawn from a review of audit assertions and their corresponding substantive analytical procedures as provided by several of the CA/ CM application providers. This review of existing tests provides a foundation for the review and decision of which processes to undertake.

- Prepare the report

The case study format seems ideally suited for the two research questions of this project: first, can existing CA/CM tools and techniques be implemented and generate appropriate analyses in an SME NFP setting? And second, what is the motivation for an SME NFP to implement CA/ CM technology? The first question addresses the issue of "how" and the second, "why". However, further exploration of CA/CM research and the domain of small to medium sized not-for-profits will be required before determining this feasibility, as is additional discussion of the project motivation.

Test	Purpose	Script
Duplicates by Employee and Amount	Identify possible duplicate pay- ments to employees in one pay cycle	Pre-defined duplicates test as configured in the CA/CM tool, based on employee ID (test 1) and check amount (test 2)
Check Date after Termination Date	Identify payments to terminated employees after termination	'Check Date' > 'Termination Date'
Overtime Pay at least ½ of salary amount	Identify possible overpayments to employees	'Overtime Amount' >= 'Regular Pay'/2
New hire pay before Hire Date	Identify payments prior to em- ployment commencing	'Check Date' < 'Hire Date'
Payments without any benefits deductions	Verify that benefits deductions are being recorded properly	Search for payment/check where all 'Benefit Deduction' fields are -0-
Termination Date match	Verify the termination date coin- cides in both H/R and Payroll sys- tem databases	'H/R Termination Date' <> 'Payroll Termination Date'

Exploration of Audit Opinion Shopping: Evidence from Non-Federal Organizations

Desi Arisandi, Feiqi Huang and Miklos Vasarhelyi

Introduction

•Given the current economic challenges faced by governments and nonprofit entities, the need to supervise the implementation of federal funds is very crucial.

•Audit opinion is very important because it can influence the success or failure of the organization in the future (Citron & Taffler , 1992).

•Non-federal entity (Local Government and Non profit Organization) that receive qualified audit opinion can influence the possibility to receive future federal funding or even raise public awareness about the accountability of the organizations.

•The change of audit firm that motivated by the tendency of auditee to receive better audit opinion or so-called opinion shopping (Gomez-Aguilar & Ruiz-Barbadillo, 2003)

•This paper examines whether Non-Federal Organizations successfully engage in opinion shopping activities to avoid a qualified audit report.

Research Methodology

Research Methodology - Hypothesis 1





Literature and Research Question

- •Auditor primary contribution is to issue written report that express an opinion regarding the fairness of the financial statements based on the applicable accounting standards (Hackenbrack & Hogan, 2002; Arens, Elder, & Beasley, 2012).
- •To provide value to the consumer of audit service, the auditor should technically capable and independent in providing their opinion (Citron & Taffler , 1992).
- •When the organization is planning to change their image or require different set of expertise, the subsequent audit firm can also be different (Davidson III, Jiraporn, & DaDalt, 2006).
- Research Questions:
- 1. Do firm switch auditors more often after receiving qualified opinion than after receiving clean opinion?
- 2. When firms switch auditor immediately after receiving a qualified opinion, are they more likely to receive clean opinion the following year?

Data and Results





Research Methodology – Hypothesis 2



Conclusion and Discussion

- Auditor change ratios in qualified samples are significantly higher than ratios from unqualified samples.
- For qualified sample entities with auditor change, they are less likely to received qualified audit opinion in the following years.
- The above evidences imply the existing and successful engagement of audit opinion shopping activities in Non-Federal organizations.
- Future research: Incorporating advanced statistical model such as Lennox's model (2000)

local governments, and non-profit organizations. Based on circular no. A-133, Non-federal entities that expend \$500,000 (\$300,000 for fiscal year before December 31, 2003) or more in one year in federal awards shall implement a single audit conducted for that year. 3. Tribes
 4. Local Territories
 5. Non Profit Organizations
 Schools
 Hospitals
 Housing Organizations

Housing Organizations
Social Services

 Final sample is 9,504 matched sample (after eliminate missing variables and not fit sample)

Accumulative Auditor Change Ratio				
Year 1 2 3				
Qualified Samples	12.54%	19.70%	26.01%	
Unqualified Samples	8.08%	14.81%	22.64%	
Difference	4.46%	4.89%	3.37%	
Test				
Pearson Chi2(1)	12.784	9.916	3.658	
P-Value	0.000	0.002	0.056	

Qualified Audit Opinion Ratio				
Year 1 2 3				
Auditor changed	21.48%	16.24%	13.92%	
Non-Auditor changed	44.56%	32.49%	26.96%	
Difference	23.08%	16.25%	13.04%	
Test				
Pearson Chi2(1)	28.5733	23.9729	21.603	
P-Value	0.000	0.000	0.000	

Design of Apps for Armchair Audit of Government

Procurement Contracts

Qiao Li, Jun Dai and Miklos A. Vasarhelyi

Risk Type 1: Bid Rigging

Introduction

Examples of Proposed Apps

Background:

Armchair Audit

Armchair Audit is a form of "crowd sourced" accountability where citizens use open government data on spending, contracts and crime to hold public bodies to account. Governments, such as British and Brazilian Federal governments, have started open data initiatives to make their operations more transparent to their citizens. With open data, armchair auditors are able to identify potential risks in current operation processes of a government using data analytics techniques.

Audit Apps

Audit applications are formalized audit procedures that are performed through computer scripts. Ease of use, user-friendly interface, and low cost make audit apps a popular choice when performing data analytics.

Objectives of the Study:

1. Propose a list of audit apps that help armchair auditors to:

Analyze open government procurement data

			OF SERVICES HOLF ACTIVE FEU			
Purpose of App	Data	Risk Indicator	Purpose of App	Data	Risk Indicator	
Monopoly check	Industry data, bid- der Info	Only very few bid- ders	Duplicate billings check (for same prod-	Billings	Duplicate billings for the same prod- ucts or services	
Contract prices comparison (gov. VS other clients)	Prices to different clients	Contractor submit higher price bids to gov. for exactly same product /service	Address check (company's & delivery)	Addresses	Delivery location is not the office, plant, or job site	
Bidders with- draw detection (in a short time	B i d d i n g process in- formation	Qualified bidders in- explicably withdraw	Geographic in- formation check	Billings	Employees bill at multiple distant job sites on same day	
period) Risk Type 2:B of Interest	2:Bribery, Kickbacks, and Conflicts		Abnormal work- ing hours check	Billings	Employees bill for more hours than typically worked in a day	
Purpose of App	Data	Risk Indicator				
Relationship	Background information Employment of con- tractor or sub-, or	Other Types of Apps:				
(gov. personnel of both par- in government pe	their family member in government per-	Purpose of App	Data	Risk Indicator		

Risk Type 3: Changing for Products not Used or Services not Rendered

· Identify high-risky contracts, and detect exceptions and anomalies

2. Provide illustrations of proposed apps to identify risks in Brazilian federal government contracts

Orders chang-	Tracks of
es check	changed
(times, costs)	ders

3. Anomaly Detection

VS contractor)

Many changed orders/ Orders changed or- changed by a high percent of original costs

sonnel

check	ue
(unusual "0"	tra
and small val-	
ues)	

Contract values Initial val-

ues of con-	in the values,
tracts	such as 0,
	0.01,0.05

Unusual number

Illustration of Audit Apps

ties

1. Descriptive Analysis



Software : Qlik Sense

-- Dashboard for Visualization

App script	Sample resul	ts	
Sub Main	IDENTIFICADOR	DO_CONTRATO	
Call Excellmport() 'C:\Users\Administrator\Downloads\integrity check.xisx Call DiredExtraction() 'integrity check3-Sheet2.IMD	1 11460650000011984		
End Sub	2 15404753000011984		
File - Import Assistant Excel Function Excellmport Set task = Client.CetImportFask("ImportExcel") doName = C.'UsersVAdministrator(Downloads/integrity checksts/" task:FileToImport = "Sheet2" task:CutputFilePreter = "Integrity check3" task:FileToImport = "Sheet2" task:CutputFilePreter = "TRUE" task:EmpthNumericFieldAsZero = "FALSE" task:PerformTask doName = task:Odup/FilePath("Sheet2") Set task = Hothing Client.OpenDatabase(doName) EndFunction	3 15404753000011986		
	4 17011650000011988		
	5 51208450000011992		
	6 51211150000011992		
	7 17011950000011990		
	Integrated results: For contracts that lost contractor records, 90%		
"Data: Direct Extraction Fundion DirectExtraction Set db - Client OpenDatabase("integrity check3-Sheet2.MID") Set task = db Extraction	belong to waived bidding		
task.IncludeAllFields dbName = "Integrity check - Contractors.IMD"	 In 470,683 contracts, 35,516 contracts lose contractor informat 6,167 contracts lose bidding mode 		
task.AddExtraction dbName, ", "@IsBlank(CONTRATADA)" task.CreateVirtualDatabase = False			
task.PerformTask 1, db.Count Set task = Nothing			
Set db = Nothing	1 000 contracts loct up	d dataa	

bsolute Deviation: 0.00126 sion: Acceptable conformity	V First Two	VALOR_INICIAL o Digits - Positive Values			
	33 37 41 45	49 53			
15 17 21 25 25		Digit S	juence	, , , , , , , , , , , , , , , , , , , ,	
	Opper Bound	- Lower	uspicious Suspicious		
A	L .	1 1			
App script	or/Downloads\initial value xlsx		Sample r		
App script	orDownloads\initial value.xlsx IMD		Sample r IDENTIFICADOR_DO_CON 7007854000011994 7005854000011996	esults ITRATO VALOR_INICIA 0. 0.	
App script	t or/Downloads'initial value xisx MD	E>	Sample r IDENTIFICADOR_DO_CON 7007854000011994 7005854000011996 5326654000011996	esults ITRATO VALOR_INICIA 0. 0. 0.	
App script	t orDownloadslinitial value xisx MD	⊏>	Sample r IDENTIFICADOR_DO_CON 7007854000011994 7005854000011996 5326654000011996 53266540000011996	esults ITRATO VALOR_INICIA 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.	
App script Alin Call Excellmport) Call Excellmport) Call Excellmport Sub -import Assistant: Excel Ion Excellmport Set task = Client GetImportTask("ImportExcel") dName = "CiUsersVidministratorDownloadsVi	t orDownloads'initial value xisx IMD Initial value xisx"	5	Sample r IDENTIFICADOR_DO_CON 7007854000011996 5326654000011996 5322654000011996 5322654000011996 5325454000011996	esults ITRATO VALOR_INICIA 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.	
App script Vain Call Excellmport() Call DirectExtraction() 'Initial value2-Sheet1.1 Sub -Import Assistant. Excel Set task - Client GetImportTask('ImportExcel') dbName = 'C:\UserSivdministratorDownloads'i task.ShelTolmport = Sheet1'	t or/Downloadslinitial value xisx IMD initial value xisx*	⊏>	Sample r IDENTIFICADOR_DO_CON 7007854000011994 7005854000011996 53226554000011996 53226554000011996 5325454000011996 5301752000011996	esults ITRATO VALOR_INICIA 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.	
App script Main Call Excellmport) Call DirectExtraction() 'C:UJsersV4dministrator Call DirectExtraction() 'Initial value2:Sheet1.1 Sub -Import Assistant: Excel fion Excellmport Set task : Client GetImportTask('ImportExcel') dbName = 'C:UJsersV4dministratorDownloads/ task: SheetToImport = Sheet1' task: OutputFilePrefix = 'Initial value2' task: ExcelToImport = 'Sheet1' task: ChipturFilePrefix = 'Initial value2'	t orDownloads'initial value xlsx IMD initial value xlsx*	=>	Sample r IDENTIFICADOR_DO_CON 7007854000011994 7005854000011996 5326654000011996 5325454000011996 5325454000011996 5301752000011996 5301752000011996	esults ITRATO VALOR_INICIA 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	
App script Main Call ExcelImport() Call DirectExtraction() Sub *-Import Assistant: Excel ction ExcelImport Settask = Client GetImportTask["ImportExcel") dbName = "C:USersVdministratorDownloadsV task.FirefToImport = "Sheet1" task.ForefToImport = "Sheet1" task.ForefToImport = "TRUE" task.EmpthNumericFieldAsZero = "FALSE"	t orDownloads'initial value xisx IMD initial value xisx*	€>	Sample r IDENTIFICADOR_DO_CON 7007854000011994 7005854000011996 5326654000011996 5322654000011996 5322654000011996 5301752000011996 01004540000012000 5303252000011996	esults ITRATO VALOR_INICIA O O O O O O O O O O O O O	
App script Main Call Excellmpont) Call DirectExtraction() 'C'Users/Administrato Call DirectExtraction() 'initial value2-Sheet1.1 Sub a-Import Assistant: Excel ction Excellmport Settask = Client GetImportTask("ImportExcel") dbName = 'C:Users/AdministratorDownloadsvi task/DieToImport = 'Sheet1' task.OutputFilePrefix = 'initial value2' task.FirstRowlsFieldName = 'TRUE' task.EngtNumericFieldAsZero = 'FALSE' task.PerformTask dbName = task.OutputFilePath("Sheet1') Settask = Nothing Client.OpenDatabase(dbName)	t orDownloads'initial value xlsx IMD Initial value xlsx*	=>	Sample r	esults ITRATO VALOR_INICIA 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.	
App script Call Excelimport) Call Excelimport) Call DirectExtraction() 1 Sub e - Import Assistant Excel ction Excelimport Settask - Ctlent GetImportTask("ImportExcer") dbName = "C:USersVAdministratorDownloads" task SheetToImport = "Sheet1" task OutputFilePrefix = "Initial value2" task FirstRowtsFoldName = TW2E' task PerformTask dbName taskSoutputFilePath("Sheet1") Settask = Nothing ClientOpenDatabase(dbName) IFunction	t or/Downloadslinitial value xisx IMD initial value xisx*	₽	Sample r IDENTIFICADOR_DO_CON 7007854000011994 7005854000011996 5326654000011996 5325454000011996 5325454000011996 5301752000011996 5303252000011996 5502652000012000	esults ITRATO VALOR_INICIA 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.	
App script Main Call ExcelImport() Call DiredEbtraction() 'Initial value2-Sheet1.1 Sub se-Import Assistant: Excel ction ExcelImport Settask = Citent GetImportTask("ImportExcel") dbName = "C:USers!Vaministrator/Downloads// task FileToImport = dbName task.SheetToImport = dbName task.SheetToImport = dbName task.SheetToImport = "Sheet1" task CutputFilePrate = "Initial value2" task.FileToImport = "Sheet1" task.FileToImport = "Sheet1" task.FileToImport = "Sheet1" task.FileToImport = abName task.SheetToImport = "Sheet1" task.FileToImport = abName task.SheetToImport = Sheet1" task.FileToImport = abName Tubert = task.CutputFilePrath("Sheet1") Settask = Nothing Client.OpenDatabase(dbName) Function	t orDownloadslinitial value xisx IMD initial value xisx*	¢	Sample r IDENTIFICADOR_DO_CON 7007854000011994 7005854000011996 5326654000011996 5322654000011996 53225454000011996 5301752000011996 530252000012000 5303252000012000	esults ITRATO VALOR_INICIA 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.	
App script Main Call ExcelImport() Call DirectExtraction() 'Initial value2-Sheet1.1 Sub a - Import Assistant: Excel ction ExcelImport Settask = Client GetImportTask['ImportExcel') dbName = 'C:USersVdministratorDownloadsV task FileStask = Client GetImportTask['ImportExcel') dbName = 'C:USersVdministratorDownloadsV task FileStask = Client GetImport = dbName task.SheetToImport = dbName task.SheetToImport = 'Sheet1' task CutpufFilePratk("sheet1') Settask = Nothing ClientOpenDatabase(dbName) IFunction ta: Direct Extraction ction DirectExtraction Settask = db Extraction Settask = db Extraction	t orDownloads'initial value xisx IMD initial value xisx*		Sample r IDENTIFICADOR_DO_CON 7007854000011994 7005854000011996 5326554000011996 5325454000011996 5325454000011996 5301752000011996 5501752000011996 5502652000012000	esults ITRATO VALOR_INICIA 00 00 00 00 00 00 00 00 00 0	
App script Main Call Excellmport() Call DiredExtraction() 'Subar *-Import Assistant Excel ction Excellmport Settask = ClientGetImportTask("ImportExcel") dbName = 'C:USersNdministratorDownloads// task Filerformport = Sheet1" task CutputFilePTerk = "initial value2" task Filerformport = Sheet1" task CutputFilePTerk = "initial value2" task FilerformTask dbName = task.CutputFilePath("Sheet1") Settask = Nothing ClientOpenDatabase(dbName) IFunction ta: DiredExtraction ction DiredExtraction Settask = db Extraction task.IncludeAliFields dbName = task.CutputFilePath("Sheet1") Settask = db Extraction task.IncludeAliFields	t orDownloads'initial value xisx IMD initial value xisx*		Sample r IDENTIFICADOR_DO_CON 7007854000011994 7005854000011996 5326554000011996 53254554000011996 53254554000011996 5301752000011996 5303252000011996 5502652000012000	esults ITRATO VALOR_INICIA 00 00 00 00 00 00 00 00 00 0	
App script Alian Call ExcelImport) Call DirectExtraction() 'C:USers/Administrator Call DirectExtraction() 'Initial value2-Sheet1.1 Sub a - Import Assistant Excel idion ExcelImport Settask = ClientGetImportTask['ImportExcel') dbName = C:USers/AdministratorDownloadSV task.FilerIoImport = CNAme task.SheetToImport = "Sheet1" task.OutputFilePrefix = "initial value2" task.FilerIoImport = CNAme task.SheetToImport = "Sheet1" task.CutputFilePrefix = "initial value2" task.FilerIoImport = CNAme task.SheetToImport = "Sheet1" task.Empt/NumericFieldAs2ero = "FALSE" task.PerformTask dbName = task.CutputFilePath('Sheet1') Settask = db Extraction ClientOpenDatabased(Initial value2-Sh Settask = db Extraction task.IncludeAliFields dbName = "small value IMD" task.AddExtraction dbName, "," VALOR_INICIAL	t orDownloads'initial value xisx IMD initial value xisx* Heet1.IMD")		Sample r IDENTIFICADOR_DO_CON 7007854000011994 7005854000011996 5326554000011996 53254554000011996 5301752000011996 5301752000011996 5502652000012000 5502652000012000	esults ITRATO VALOR_INICIA 00 00 00 00 00 00 00 00 00 0	
App script Main Call Excelimport() Call DirectExtraction() 'solution Excelimport Set task = Client GetImportTask['ImportExcel') dbName = 'C:USersVdministratorDownloads' task/BriedToimport = 'Sheet1' task/DutpuFilePrefix = 'initial value2' task/BriedToimport = 'Sheet1'' task/DutpuFilePrefix = 'initial value2' task/BriedToimport = 'Sheet1'' Set task = Nothing Client/OpenDatabase(dName) Function task/IncudeAliFields dbName = 'small value MD'' task/AdExtraction dbName, ", 'VALOR_INICIAL task/CreateVirtualDatabase = False task/PerformTask 1, do.Count	t orDownloadsiinitial value xisx IMD initial value xisx" eett IMD") L < 0.1"	¢	Sample r IDENTIFICADOR_DO_CON 7007854000011994 7005854000011996 532654000011996 532654000011996 5322454000011996 5301752000011996 5502652000012000 5502652000012000 Integrated res 501 contracts value after rem pertaining to departments;	esults ITRATO VALOR_INICIA O O O O O O O O O O O O O	
App script b Main Call Excellmport() Call DirectExtraction() 'C:IUsers/Administrato Call DirectExtraction() 'Initial value2-Sheet1.1 d Sub e - Import Assistant: Excel tiction Excellmport Set task: FilerCollmport = 'Sheet1' task: CutputFilePrefix = 'Initial value2' task: FilerCollmport = 'Sheet1' Set task = Nothing Client OpenDatabase('Initial value2-Sh Set task = do Extraction Set db = Client.OpenDatabase('Initial value2-Sh Set task = do Extraction task: IndueAliField8 dbName = 'small value.IMD' task: AddExtraction dbName, ", ''WALOR_INICIAL task: CredeVirtualDatabase = False task: PerformTask 1, db.Count Set task = Nothing Set db = Nothing	t orDownloadslinitial value xisx IMD initial value xisx" eeet1.MD")	E>	Sample r IDENTIFICADOR_DO_CON 7007854000011996 532654000011996 5322454000011996 5322454000011996 5322454000011996 5301752000011996 5301752000011996 5502652000012000 Integrated res 501 contracts value after rem pertaining to departments; 527 contracts h	esults ITRATO VALOR_INICIA O O O O O O O O O O O O O	
App script DMain Call ExcelImport) Call ExcelImport) Call DiredExtraction() initial value2-Sheet1.1 ad Sub ille - Import Assistant: Excel Inction ExcelImport Settast - Client CetImportTask("ImportExcel") dthame = C:IUsersVdministratorDownloadsW task.FileToImport = Sheet1" task.CutputFilePrefix = Initial value2" task.FileToImport = Sheet1" task.CutputFilePrefix = Initial value2" task.FileToImport = Sheet1" task.CutputFilePrefix = Initial value2" task.FileToImport = Sheet1" task.PerformTask dthame = task.CutputFilePath("Sheet1") Settask = Nothing Client OpenDatabase(dthame) ndF Unction Set db = Client OpenDatabase("initial value2-Sh Settask = db Extraction Set db = Client OpenDatabase("initial value2-Sh Settask = AdbExtraction Set db = Client OpenDatabase("initial value2-Sh Settask = AdbExtraction task.AddExtraction dthame, ", "VALOR_INICIAL task.CreateVirtualDatabase = False task.PerformTask 1, db.Count Set ask = Nothing Client OpenDatabase (dthame) ht Eurotion	t orDownloadslinitial value xisx IMD initial value xisx* eeet1.IMD*)	E>	Sample r IDENTIFICADOR_DO_CON 7007854000011996 5326554000011996 53254554000011996 53254554000011996 5325454000011996 5301752000011996 5301752000011996 5502652000012000 Integrated res 501 contracts value after rem pertaining to departments; 527 contracts h <1; the values	esults ITRATO VALOR_INICIA O O O O O O O O O O O O O	

Software: SAS



-- Matching with Other Data

Sample results			
Contractor	Frequency		
29.213.386/0001-00	164		
64.198.393/0001-72	141		
59.363.937/0001-74	113		
00.394.494/0016-12	106		
00.656.865/0001-00	103		

ワ Integrated results:

- 40,942 contracts are signed with companies •
 - that do not have any information in the contractor file
- 18,778 unique suspicious contractors

Sources (Suppliers)



Understanding Usage and Value of Audit Analytics: An Organizational Approach

He Li, Jun Dai, Tatiana Gershberg, and Miklos A. Vasarhelyi

Introduction

Audit analytics is the process of using various analytic techniques such as statistical and quantitative analysis and explanatory and predictive models upon financial, operational, and other forms of electronic data to identify risks, provide assurance, and support decision-making (KPMG 2012; Schneider et al. 2015; Cao et al. 2015). Audit analytics not only improves the audit process through increasing operational efficiencies and reducing costs (KPMG 2012), but also identifies potential frauds and anomalies in time to provide a higher level of assurance (EY 2014). Although internal auditors are increasingly aware of the importance and value of data analytics, some surveys conducted by practitioners show that audit analytics is not being fully utilized by the majority of companies. Many auditors are not able to effectively incorporate audit analytics in their work and only use it on an ad-hoc basis. Therefore the objective of this paper is to examine factors at organizational level that have an impact on audit analytics adoption and usage, and

Theoretical Model

We choose to use the Technology-Organization-Environment (TOE) framework because it focuses on the post adoption stage and it studies the technology adoption and usage in an organizational context. Based on prior literature, we hypothesize that technological competence, IT complexity, firm size, management support, standards, and professional help, will have a posi-

tive impact on audit adoption. IT Complexity this paper posits that audit ana-Technological lytics adoption, pro-Competence fessional, and technocompetence drive the usage. Fi-Management Analytics Usage nally, both the adop-Support Adoption tion and usage are adding value to inter-Value nal audits. In our ad-Size dition analysis, we al-Audit Analytics expect publicly Usage held firms and pri-Professional firms Help have different pat-Standards H6

its value to internal audit.

terns.

vately held

SO

analytics

Further,

logical

Methods & Participants

An online survey was administrated to test our hypotheses. Participants of the survey are clients of a major data analysis software vendor. The questionnaire was developed by referring to prior literature and consulting experts. The survey was distributed to 4820 firms and we collected 427 responses, representing a response rate of 9%.



Perceived Benefit across Different Maturity Level



SEM Results

Our results indicate that the audit analytics adoption by internal auditors is driven by their perceived level of importance and technological capability. Encouragement by management and regulators are the most important factors in shaping how internal auditors use audit analytics. Factors that relate to the firm's characteristics, such as IT complexity and firm size, do not have significant influence. The audit analytics usage is influenced by professional help, technological competence, and audit analytics adoption. It supports the argument that advanced audit analytics tools require expertise in statistics and technology, which can be acquired by frequently using audit analytics throughout the audit process, or by enhancing technological competence, or by seeking help from vendors. Finally, both audit analytics adoption and audit analytics usage provides value to internal audit function. In our additional test, publicly held firms and privately held firms are separated. Standards are shown to be affecting publicly held firms in audit analytics adoption, but not privately held firms. Management support, on the other hand, is a key factor for privately held firms, but not for publicly held firms. Finally, privately held firms seem to only benefit from audit analytics adoption. Exploring various analytics tools does not provide significant value.

Results in this paper should be valuable to both practitioners and regulators. Software vendors can use the factors identified in our model to promote their products efficiently. In addition, since standards have been proved to be the most powerful indicator, regulators can develop rules or guidance to encourage audit analytics usage.



A Vision of an ENHanced ANalytic Constituent Environment: ENHANCE

Stephen Kozlowski, Hussein Issa and Deniz Appelbaum

Background

Constituent demands for improved transparency in governmental reporting have increased since the recent (2008-2009) financial crisis in the U.S. that impacted the financial well-being of a number of U.S.-based governmental entities at both state and local levels. Several governmental entities in the U.S. have responded by the effort to provide an open data environment with very limited analytic capabilities. Recent Federal initiatives with respect to open data, in particular the 2014 DATA Act legislation, will drive further open data initiatives that will require advanced analytics to fulfill con-

Objective

The purpose of this research is to develop an ENHanced ANalytic Constituent Environment (ENHANCE) facilitated by open government data that fulfills the reporting requirements of the various governmental stakeholders, such as citizens, analysts, bond investors, creditors, and oversight officials. Auditors should also be considered within the group of governmental stakeholders.

.<u>Methodology</u>

The first step in the procedure is the availability of open governmental data, presented in a standardized and usable format. The second step is that of designing a series of analytic 'apps' that can provide meaningful information to the entity's stakeholders. The third step is the development of ENHANCE where apps can function, acting upon the standardized government data to support constituent reporting requirements.

An Example of a Governmental Open Data Portal



stituents demands.

Typical Checkbook (Spending) Data Attributes (State of Ohio)

Attribute:	Description:
Checkbook Unique Identifier (9 -digit numeric)	Unique identifier for the journal entry line item (system-generated?)
Transaction Date	Date of the transaction
Vendor Name	Name of the recipient of the payment
Voucher Journal ID	An agency-specific eight- digit number that records an authorized transaction with a vendor. There may be multiple transactions (vouchers) grouped into a single vendor payment (check). Unique identifier for each journal entry (Numeric and alphanumer- ic)
Check Number	Unique identifier for each payment (Numeric and al- phanumeric)
Amount	Amount
Entity Code (alphanumeric)	Code of the entity for which the expenditure was made
Business Unit Code (alphanumeric)	Code for the Business Unit (or detailed department) within the entity for which the expenditure was made
Fiscal Year	Fiscal year in which the expenditure was made
Entity Name	Name of the entity for which the expenditure was made

Analytics to Support an Enhanced Analytic Constituent Environment

Analytic:	Function:
Cluster analysis	Identify spending outliers by agency, department, or other entity
Fraud detection	Search for duplicate payments and other inconsistencies
Pattern recognition	Identify increased spending with a particular vendor
Time series analysis	Identify spending trends over time
Benchmarking	Develop benchmarks that can help define what a normal spending behavior is expected to be
Exceptional Exceptions	Identify and prioritize spending exceptions and anomalies
Cross-entity analysis	Compare spending for similar items across similar entities
Development of KPIs	Identify entities



The Effect of Transparency on Accountability and Integrity: A Study on a Resource-Rich Developing Country Ahmad AlQassar

Abstract

According to UNDP's corruption model, transparency, accountability, and integrity are considered essential in deterring corruption. Transparency might seem an obvious necessity to most, but some views in the literature argue that it is of minimal benefit or might even cause unwanted results if not used in the right context. This study is an attempt to measure the three factors and examine the relationship between them in a unique situation. Our results show that there is a strong and significant relationship between the three factors. This proves that transparency can encourage accountability and integrity if implemented correctly, thus reduce corruption. Such results have significant implications on governmental agencies and its stakeholders. Agencies might be more encouraged to promote such initiatives, while the public and Civil Society Organizations (CSOs) might be more stimulated to raise the transparency demands and introduce analytical tools that could deal with such information.

Corruption

The topic of corruption has received renewed interest and international attention in the past two decades, resulting in a large amount of policy initiatives that are aimed at reducing corruption (Peton and Belasen 2012). The most widely used definition of corruption is the abuse of public power for private gain. However, Transparency International defines it as the abuse of entrusted power for private gain, which really covers a wider area for corruption to take place in. So whether the activity is public, private, or non-profit, corruption will occur when someone has monopoly power over a good or service, has the discretion to decide who will receive it and how much he'll get, and is not accountable (Klitgaard 1998). However, the UNDP modified Klitgaard's model by adding integrity and transparency. Thus, identifying these three characteristics is crucial to prevent and fight corruption.

Sandoval-Ballesteros (2015) made an interesting point when he stressed, "Corruption is not just a question of low-level public servants filling their pockets at the expense of common citizens, but is an institutional and political problem that requires structural solutions". Therefore, the true concern is a corrupt system, which will disrupt the primary

Transparency

Transparency is simply defined as the stakeholders' timely and reliable access to economic, social, and political information (Bellver and Kaufmann 2005). Michener and Bersch (2013) state that transparency represents the two literal meanings of the word; visibility and inferability. Visibility refers to the degree the information is attainable, while Inferability refers to the ability to use the information to arrive at valuable conclusions. Initially, the advocates of transparency focused solely on the visibility part of the equation, until they realized that information is useless if you cannot infer useful conclusions. This made us arrive at our conclusion that the type of information is insignificant. The ultimate goal is to find relevant information that can aid us in understanding and analyzing the case at hand. Transparency might seem an obvious tool to prevent corruption, since lack of it makes corruption less risky and more attractive. Proponents of such a theory are abundant. However, there is another side to that argument as well. Theories of such opposing views range from stating that transparency is a necessary but not sufficient tool for fighting corruption, to the extreme views that see transparency as an element that might, in some cases, increase corruption.

functions of government.

Accountability

that transparency is not a goal in its own right, but a tool that can be used and taken advantage of in the fight against corruption. Such information enables citizens to assess the performance of their governments and detect corruption in order to hold them accountable (McGee and Gaventa 2011, Michener and Bersch 2013, Louren 2013, Wong and Welch 2004). It has been suggested that these accountability processes ultimately "curtail fraud and corruption, increase fiscal responsibility among government actors, improve the public's understanding of why performance goals may not be met, and help to establish trust" (Benito and Bastida 2009, Curtis and Meier 2006, Ferraz and Finan 2008, Halachmi and Holzer 2010). Public officials usually do not get highpowered incentive contracts. Instead, they receive a fixed salary. Thus, retaining their job is the only way to discipline or reward them. Due to the frustration with elections and bureaucratic procedures as the main instrument to hold officials accountable for their decisions, there has been developments in the form of transparency and accountability initiatives (Lambert- Mogiliansky 2015). These new initiatives not only mitigate the weaknesses of the bureaucratic process but also allow the public to act by involving society as a whole in the accountability process by creating millions of auditors (Kaufmann 2002).

Integrity

To have integrity is to be honest and have strong moral principles. It is also regarded as one's selfregulation when it comes to loyalty, dedication, effort, and initiative to the organization. These traits usually help achieve the organizations' goal (Barney 1986, Barney and Hansen 1994, Eisenberger et al. 1987). Integrity is a quality that is needed in all facets of life, and even more so, when in a position of a trustee. Fellow citizens trust officials to provide certain services without engaging in corrupt activities. In order to achieve that, the public must have a sufficient amount of trust and commitment to the government (Witherell 2002). And the integrity of the officials is essential to achieve that trust and commitment (Tsou and Wang 2008). Although integrity is rarely acquired by training or education (Caligiuri and Di Santo 2001), it is highly connected to the promotion of governance, thus determining the public's trust in the government (Tsou and Wang 2008).

Hypotheses

H1: Higher levels of transparency will be positively associated with greater accountability.

H2: Higher levels of transparency will be positively associated with higher levels of integrity.

H3: Higher levels of accountability will be positively associated with higher levels of integrity.



Results

- The results show that the model has good predictability, where all the paths in our model are statistically significant at the 0.001 level.
- The model explains 50% and 64% of the variance in accountability and integrity respectively.
- The results suggest that access to information does enable accountability, which ultimately curtails fraud and corruption.
- We can conclude that transparency's ability to reduce corruption by making the acts risky and the agents accountable can ultimately maintain norms of integrity and trust (Kolstad and Wiig 2008).

Corporate Usage of Social Media

Pei Li and Alexander Kogan

Background

- · Public companies seek ways to communicate and engage with shareholders and the investing market. (Fang and Peress, 2009).
- Securities and Exchange Commission (SEC) announcement on April 2013.
- Given the traditional channels of corporate disclosure, what is the corporate motives for using social media?
- Corporate Disclosure Examples
 - XL_Insurance @XL_Insurance · 9 Feb 2011 XL Group announced results, op net income Q4=\$242.4 mn, FY= \$809.7 mn. P&C op comb ratio 91.4% in Q4 http://bit.ly/g8GZeY #insurance 17 ★ 000 AEP @AEPnews · 23 Apr 2013 **SAEP Well-Positioned for Continued Earnings Growth. Shareholders**

Learn @ Company's Annual Meetingaep.com/?i=1807

27 ★ ...

AEP @AEPnews · 23 Apr 2013

Motivation

on social media.

• Reduce information asymmetry

(Blankespoor, Miller, and White, 2013)

- Firms may benefit from strategically posting corporate information on social media.
- It is still unclear whether corporate disclosure via social media can receive more interest or attention from investors.
 - \Rightarrow investors have already received information via traditional channels.
 - \Rightarrow market reaction to earnings release is more pronounced with unique features of social media.

Research Questions

- Possible Motives for corporate disclosure When firms disseminate earnings via social media?
 - · How earnings release via social media impacts the capital market response to earnings?

Social Media

Features

- Simultaneously reach investors
- Quick spread
- Social capital (Ellison et al.2007; Chiu et al. 2006)
- Sentiment (Bollen et al. 2011)
- Interaction

Two Proxies to represent the attention and interests of investors

• Retweet

AEP Increases Quarterly Dividend To 49 Cents A Shareaep.com/?i=1806 * 23 1

• Favorite

Sample Selection

This study constructs a data on the corporate usage of Twitter to disseminate quarterly earnings announcements by S&P 1500 firms from 2009 through 2013.

Sample Com- position	Reduced Firms	Number	
List of S&P 150	00 companies	1500	
Less	Firm without any twitter account	-527	
	Firm' twitter account without any tweet	-37	
	Firm with inactive twitter account (twitter account with less than 5 tweets)	-14	
	Firm with protected accounts (followers need the authentication from account's user, otherwise followers cannot see tweets)	-5	
Firms with val- id twitter ac- count		<u>917</u>	
Note: 917 firms with official Twitter accounts as Dec 10, 2013			

- Keywords matching to solve classification issue of tweets.
- · Categories: earnings-related tweets and not earnings-related tweets.

• When firms disseminate earnings via social media?

Models

 $Quarter_{EM} = \beta_0 + \beta_1 Meetforecast + \beta_2 LogFollower + \beta_3 LogMV$ $+ \beta_4 ReturnOnAssets + \beta_5 LogFollowingAnalyst + \beta_6 Book to Market$ $+ + \beta_7$ Debt to Assets + YearIndicator + IndustryFixed effect + ε

How earnings release via social media impacts the capital market response to earnings?

 $CAR = \beta_0 + \beta_1 Surprise + \beta_2 LogRetweet + \beta_3 LogFollower + \beta_4 LogMV$ $+\beta_5 ReturnOnAssets + \beta_6 FollowingAnalyst + \beta_7 Book to Market$ $+ \beta_8 Debt$ to Assets $+ IndustryFix Effect + YearIndicator + \varepsilon$

Descriptive Statistics

Variable	#Ohs	Mean	Std Dev	Min	Max
	π Ου 5	witan	Stu. Dtv.	171111	
Cumulative abnormal return (CAR)	7281	0.006	0.06	-0.089	0.109
Standard Earnings Sur- prise(SUE)	6856	-0.001	0.121	-8.762	1.252
Absolute Abnormal Accrual	6729	0.521	2.068	0	32.282
Log of Market Value	7289	7.794	1.651	3.068	12.589
Return On Asset (ROA)	7285	0.015	0.038	-1.062	0.268
Analyst Following	6863	10.122	7.936	1	47
Book to Market	7177	0.526	0.355	0.01	1.889
Debt to Assets	6977	0.17	0.174	0	0.802
Tweet	7289	0.372	1.323	0	32
Retweet	7289	0.841	11.985	0	672
Favorite	7289	0.26	3.87	0	159
Follower	7289	113800.8	626047.2	. 14	8732972

Conclusion and Contribution

- This study systematically examines that firms in S&P 1500 stock index utilize social media to disseminate corporate information.
- The study applies the natural language processing method to solve the classification issue of tweets.
- The study captures two proxies of investor' attention in the social media.
- Results indicate that firms do not strategically post earnings information on social media.
- Results indicate that social media increases the impact of earnings on stock market.