Procurement Card Fraud Detection Using Hidden Markov Models and Information Fusion:

A Fusion Study

Abdullah Al-Awadhi and Deniz Appelbaum

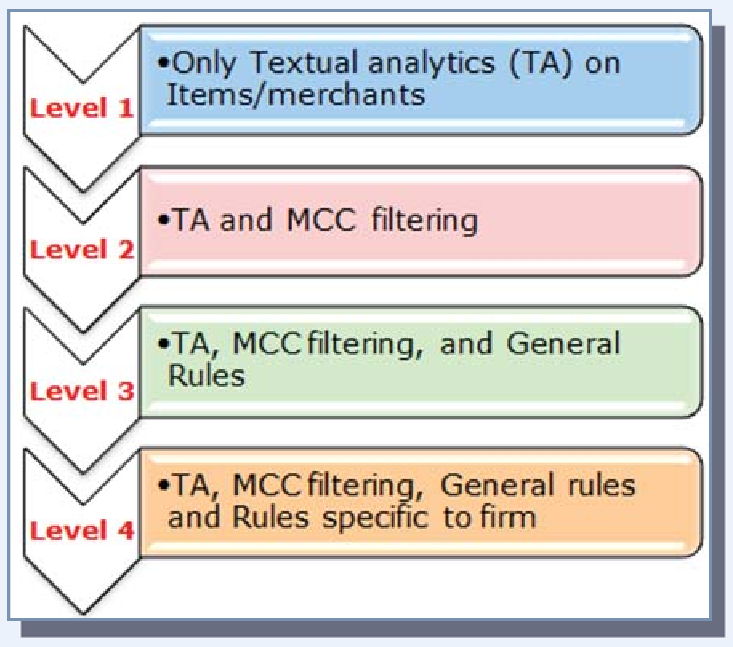
Introduction

- Large Multinational Consumer Goods Manufacturer
- Manual Batch Fraud Detection System
- Required Customized Approach, Commercial CAATs insufficient
- Many vendors do not report items that were purchased
- CARLab created a supervised system as first phase - ILisa

Measure for Jan 2013 to April 2014	Total Data Set	Missing Purchase Item Information Data Set
# of Transactions	741,710	194,528 (26% of total)
# of Employee IDs	4532 (cards are 5600)	4339 (95.74%)
Total $ Fin Original Currency	$157,115,184	$65,926,544 (42% of total)
Total # of vendors	101,900	41,258 (40.49%)

Background of ILisa

- ILisa uses duplicate testing, key words, and hundreds of association rules
- Can drill down to desired level of testing
- Elicits the expertise of the auditors


Level 1 •Only Textual analytics (TA) on Items/merchants

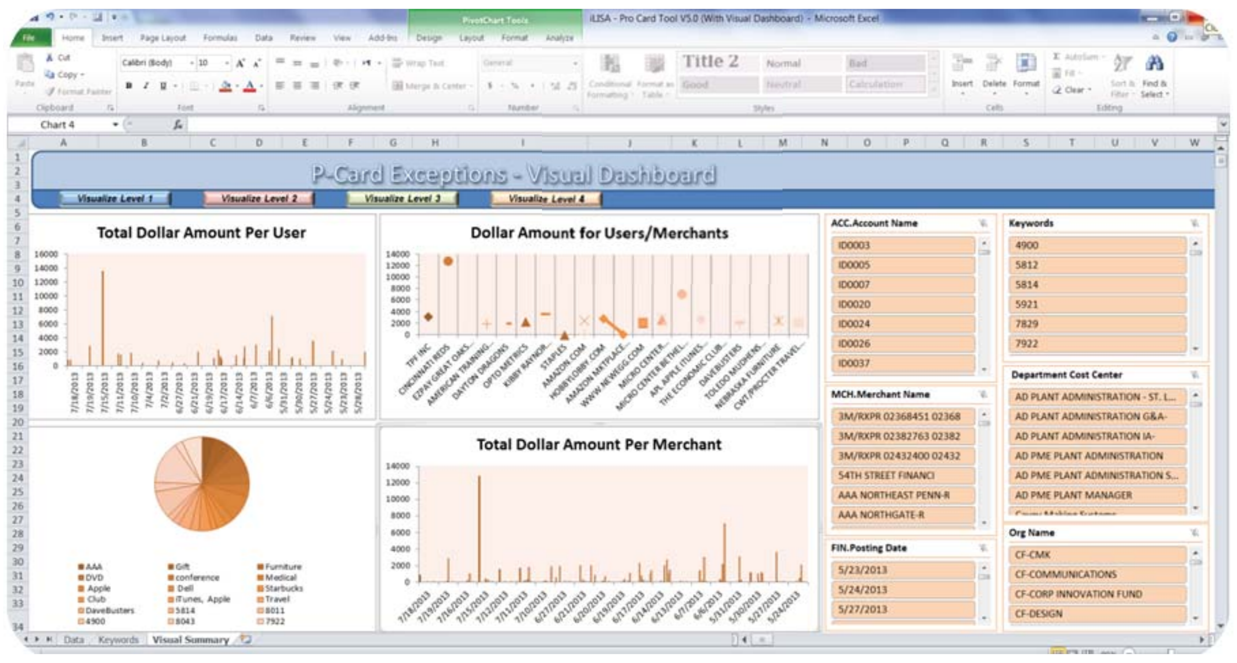
Level 2 •TA and MCC filtering

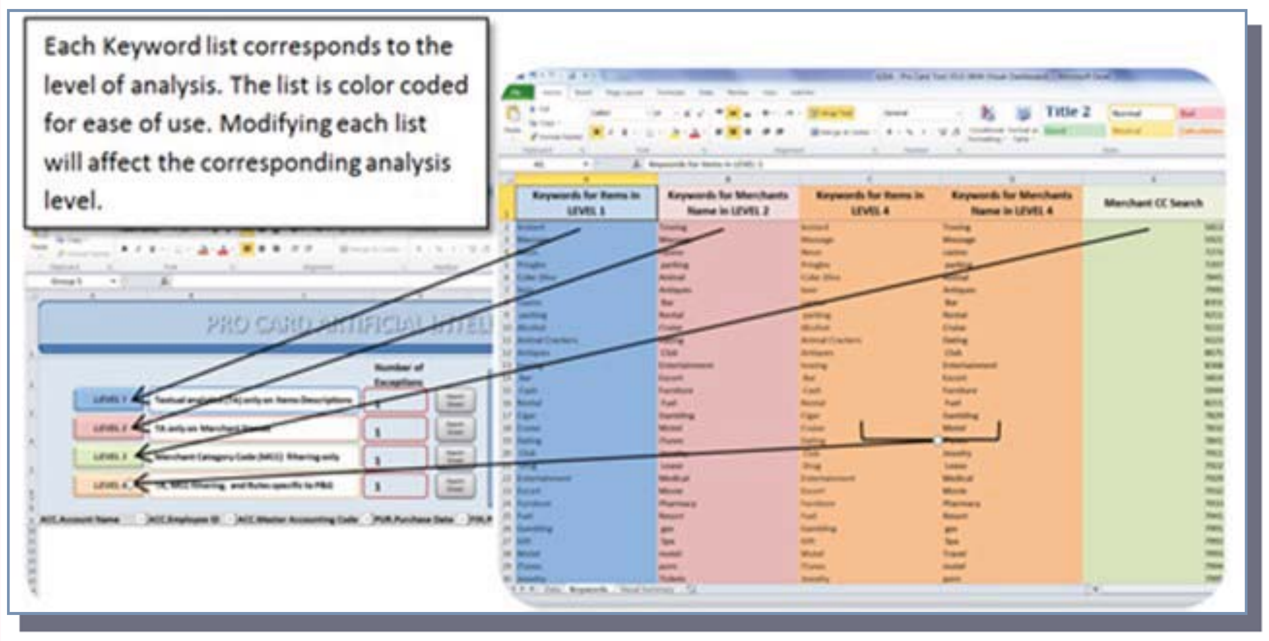
Level 3 •TA, MCC filtering, and General Rules

Level 4 •TA, MCC filtering, General rules and Rules specific to firm

ILisa Dashboards:

Levels of Analysis:

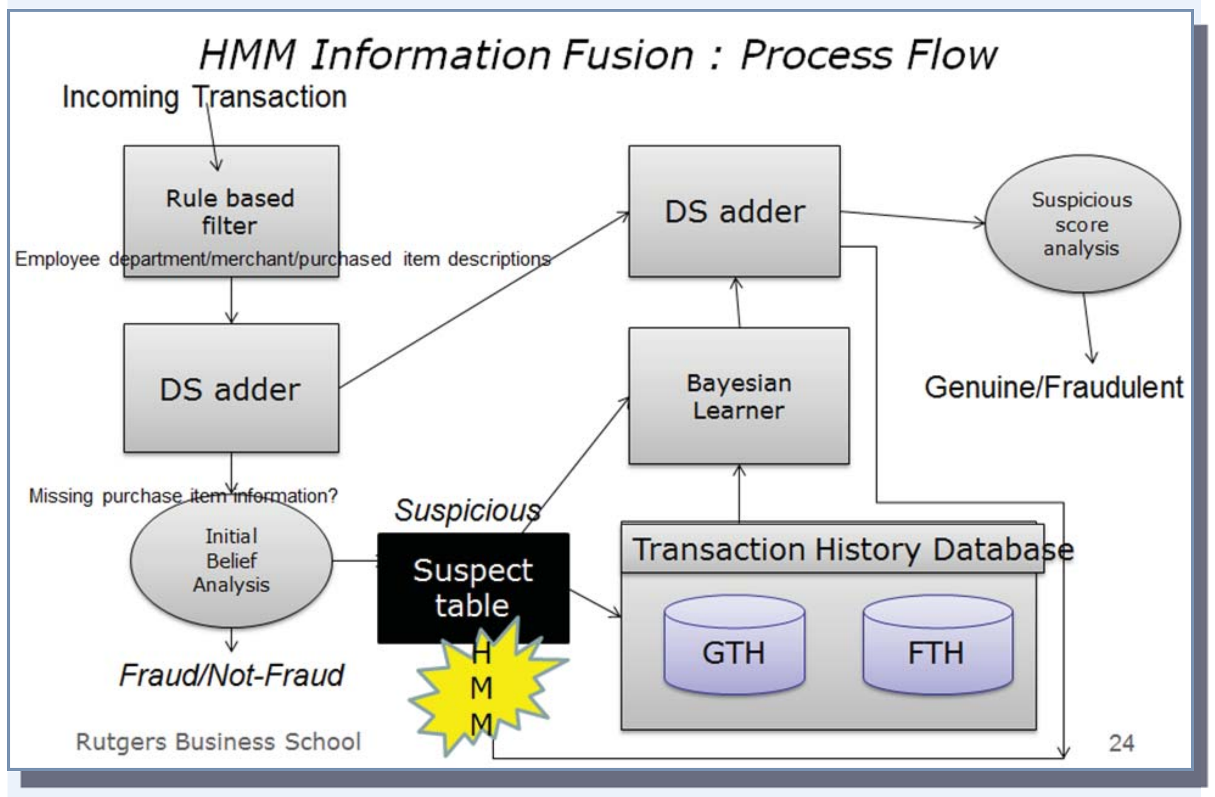



Each Keyword list corresponds to the level of analysis. The list is color coded for ease of use. Modifying each list will affect the corresponding analysis level.

PRO CARD ARTIFICIAL INTEL
Number of Exceptions
LEVEL 1
LEVEL 2
LEVEL 3
LEVEL 4
Keywords for Items in LEVEL 1
Keywords for Merchants Name in LEVEL 2
Keywords for Items in LEVEL 4
Keywords for Merchants Name in LEVEL 4
Merchant CC Search

A Fusion Study

- ILisa can't analyze the missing purchase item transactions
- New types of fraud need an unsupervised approach
- Using Hidden Markov Models embedded in a Belief Network with Dempster-Shafer:

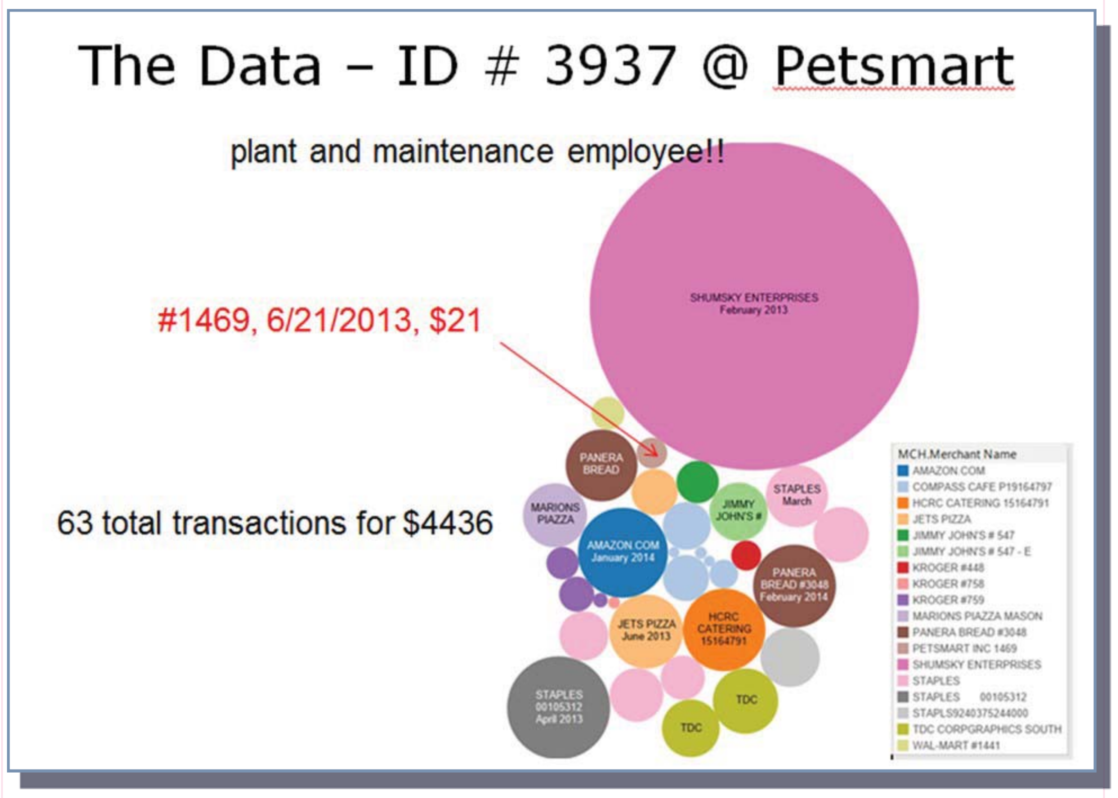


ILisa Supervised Results

The Data: 172 confirmed fraudulent transactions

CARD Numbers	TOTAL ALL Transactions	# of Confirmed Fraud	# with missing values
1351	232	29 (12.5%)	8
5014	68	1(1.5%)	1
3479	15	2 (1.3%)	2
789	71	3(4.23)	0
235	273	3(1%)	0
523	180	13 (4.64%)	0
1748	29	2 (6.9%)	0
1503	107	52 (48.6%)	0
5004	87	26 (29.89%)	7
3921	62	19 (30.65%)	1
450	65	1 (1.5%)	1
2734	14	1 (7.14%)	1
1522	276	1 (0.36%)	1
4404	52	19 (12.5%)	4
14 total cards	1531	172	26
		11.23% of transactions	1.6%all/15.1%fraud

$1689 of fraudulent transactions!


The Data – ID # 3937 @ Petsmart

plant and maintenance employee!!

#1469, 6/21/2013, $21

63 total transactions for $4436

Future Research

- Application of ILisa rules to the Information Fusion to detect false positives and negatives
- Missing Purchase Item Transactions to be analyzed with the Hidden Markov Model add-in
- Operational Efficiency of model needs to be tested
- Accuracy rate should be at 85%, based on previous research
- Missing Purchase Items is not atypical for the industry

RUTGERS
Rutgers Business School
Newark and New Brunswick

Securing Big Data Provenance for Auditors: The Big Data Provenance Black Box

By Deniz Appelbaum

Data Provenance

- **Origin** of Data
- **Lineage** of Data
- **Log files**
- Data **life cycle**
- Component of Data **RELIABILITY**

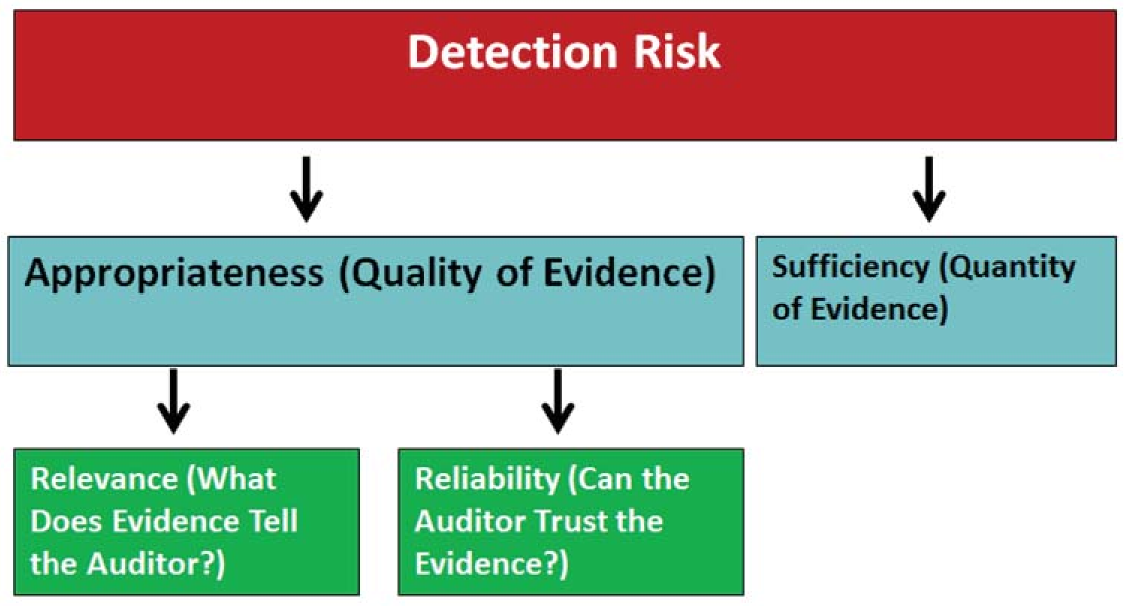
Big Data!!

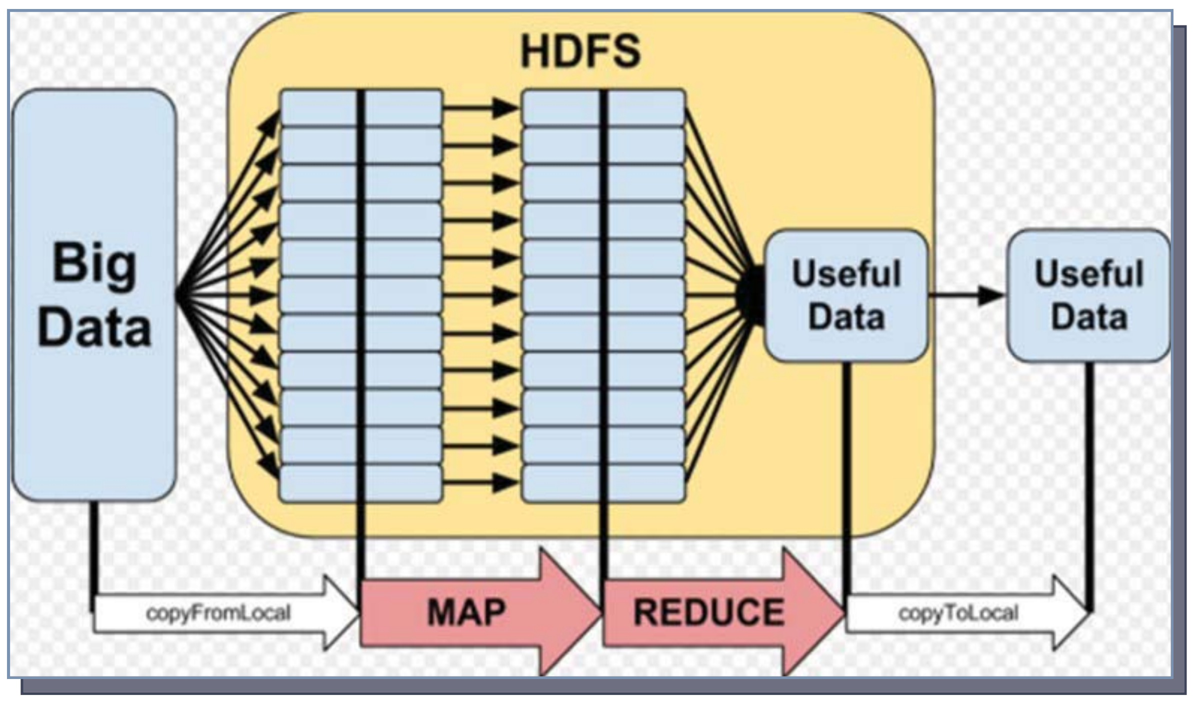
- **Velocity:** streaming very quickly and incessantly
- **Variety:** textual, social media, financial, sensor, pictures, audio
- **Volume:** massive, tetrabytes of data
- External sources to the firm

= Hadoop!

- Hadoop is used heavily by Big Data Processors, open source MapReduce
- **Map:** Reads, transforms and filters data from input files into intermediate records
- **Reduce:** splits the records with hashing and matches them to new files/buckets

Why is Data Provenance Critical?


Detection Risk
Appropriateness (Quality of Evidence)
Sufficiency (Quantity of Evidence)
Relevance (What Does Evidence Tell the Auditor?)
Reliability (Can the Auditor Trust the Evidence?)



Data Provenance in Hadoop

- **HadoopProv** (Akoush et al, 2013)
- 10% temporal lag in Hadoop
- Provenance is not secure
- **Black Box:** provenance is write once, read only (Alles et al, 2004)
- Digital signatures via hashing in the Black Box provide ultimate form of security
- These digital signatures reveal if the provenance records have been altered - the ultimate **Big Data Provenance Black Box!**

The Big Data Provenance Black Box:

- Based on HadoopProv
- Applies Black Box for secure data provenance for auditors!

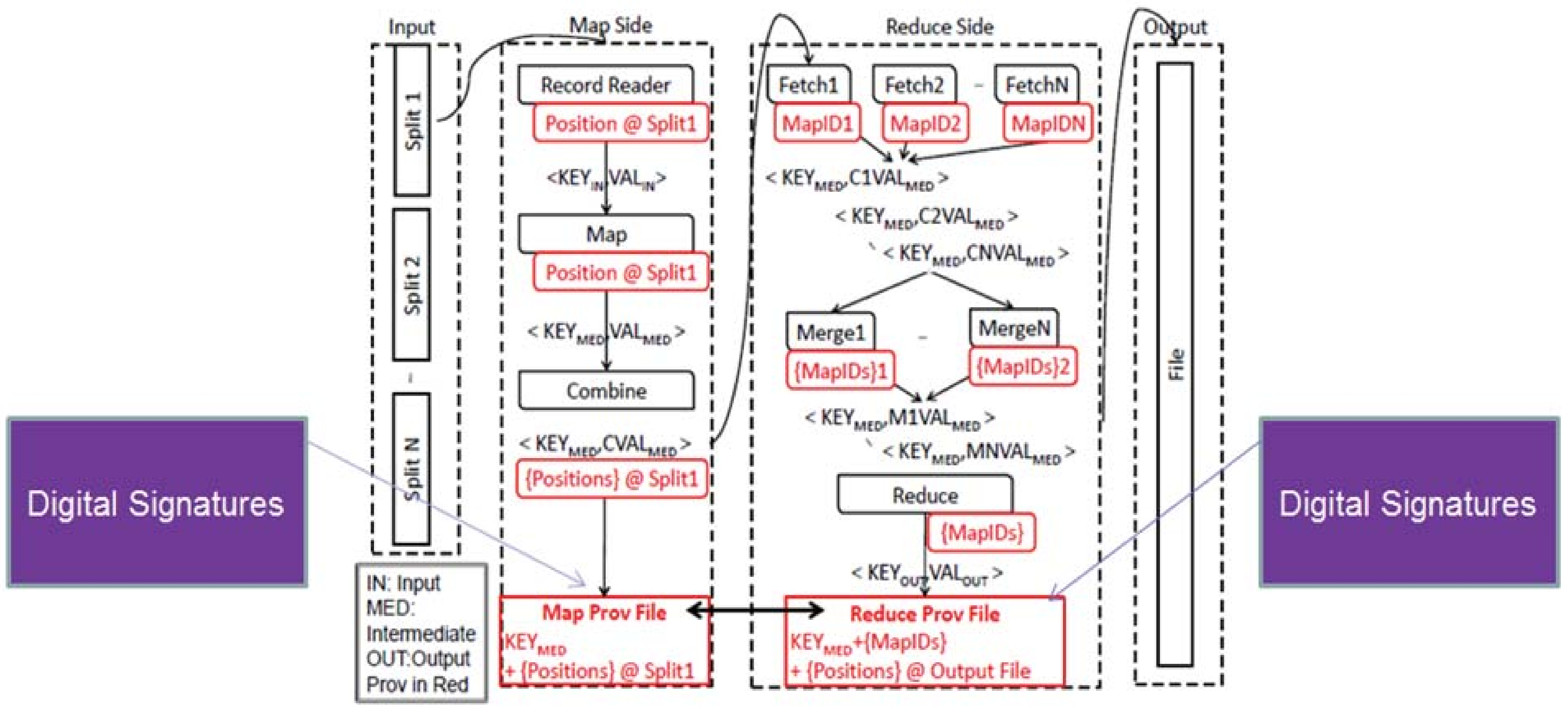


Figure 1: Provenance Capture in HadoopProv

RUTGERS
Rutgers Business School
Newark and New Brunswick

The Implementation of Exploratory Data Analysis (EDA) on State Data

Desi Arisandi and Miklos Vasarhelyi

Introduction

GASB Concepts Statement No. 1:

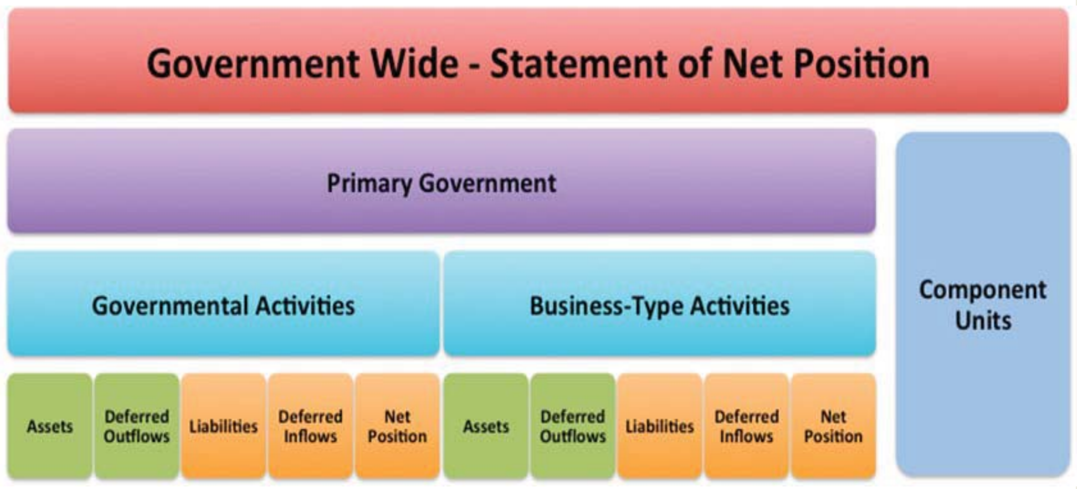
Accountability is the cornerstone of all financial reporting in government. Accountability requires governments to answer to the citizenry—to justify the raising of public resources and the purposes for which they are used.

The volume of government financial data and **the change** of structure of financial report due to standards implementation are raising the possibility of information overload. The overwhelming and dynamic information can increase the difficulty to understand the underline information of financial statements.

Exploratory data analysis (EDA) is one of contemporary methods that can assist to mine the information within substantial amount of data. The analysis can reveal every changes and dramatically flow of financial resources that are disclosed in the financial statements.

Government Financial Reports

The Comprehensive Annual Financial Report (CAFR) is a thorough and detailed presentation of the state's financial condition. It reports on the state's activities and balances for each fiscal year (GASB, 2014).



Government Wide - Statement of Net Position

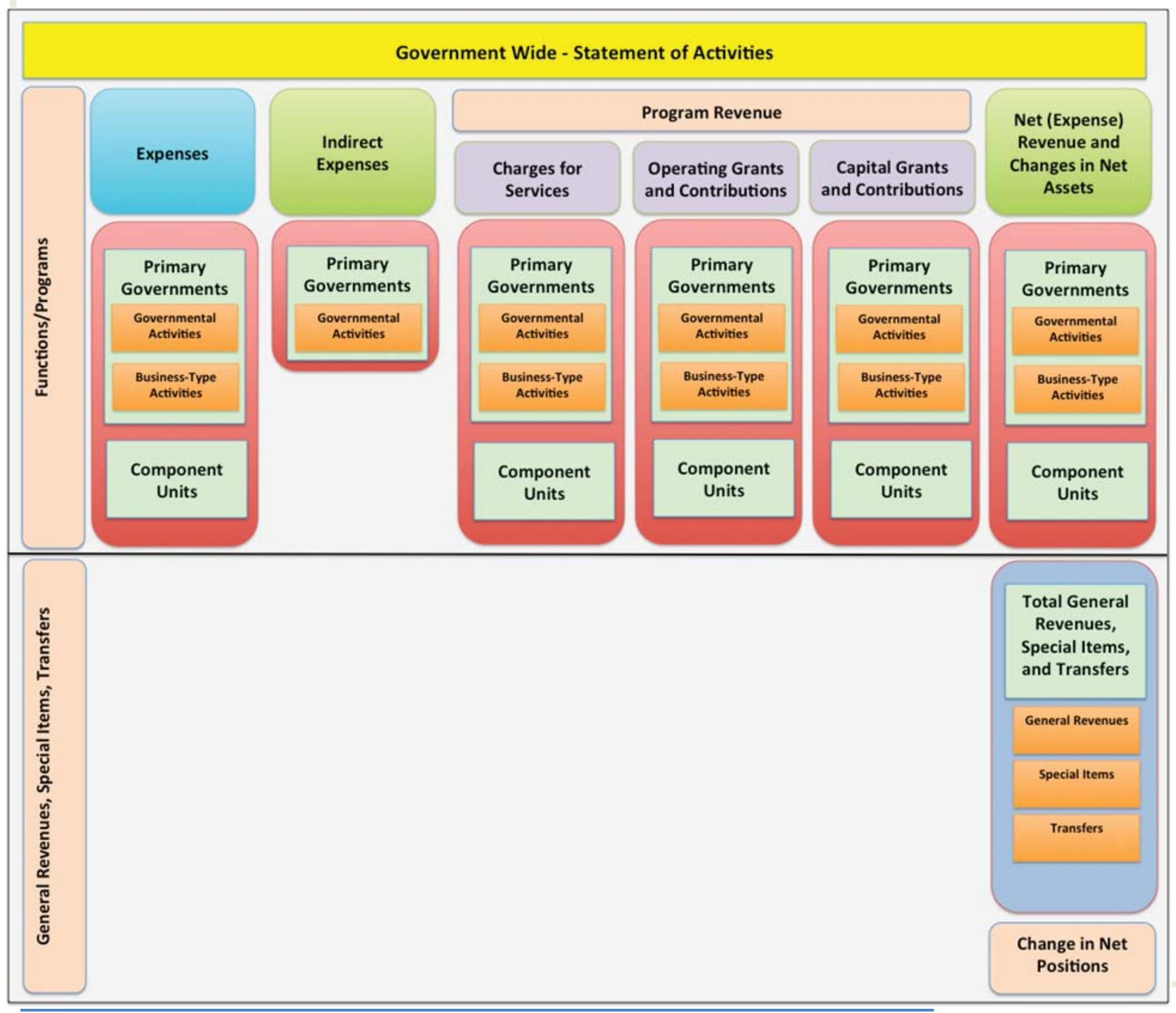
Primary Government

Governmental Activities

Business-Type Activities

Component Units

Assets | Deferred Outflows | Liabilities | Deferred Inflows | Net Position | Assets | Deferred Outflows | Liabilities | Deferred Inflows | Net Position



Government Wide - Statement of Activities

Functions/Programs

Expenses

Indirect Expenses

Program Revenue

Charges for Services

Operating Grants and Contributions

Capital Grants and Contributions

Net (Expense) Revenue and Changes in Net Assets

Primary Governments

Governmental Activities

Business-Type Activities

Component Units

General Revenues, Special Items, Transfers

Total General Revenues, Special Items, and Transfers

General Revenues

Special Items

Transfers

Change in Net Positions

CAFR is presented in three sections:

- Introductory section
- Financial section:
  - ⇒ Required Supplementary Information (RSI)
  - ⇒ Basic financial statement
  - ⇒ Notes to financial statement
  - ⇒ Audit Report
- Statistical section

In general the government activities can be classified into governmental, business, and fiduciary

Methodology

Data

- Financial data: CAFR of 51 of states level entities in United States and cover 10 years periods (2004-2013)
- Non-Financial data:
  - ⇒ Crime statistic (Source: FBI)
  - ⇒ Leadership change (Source: Entity's website)
  - ⇒ Transparency Award (Source: GFOA)
  - ⇒ Public employee corruption Conviction (Source: DOJ)

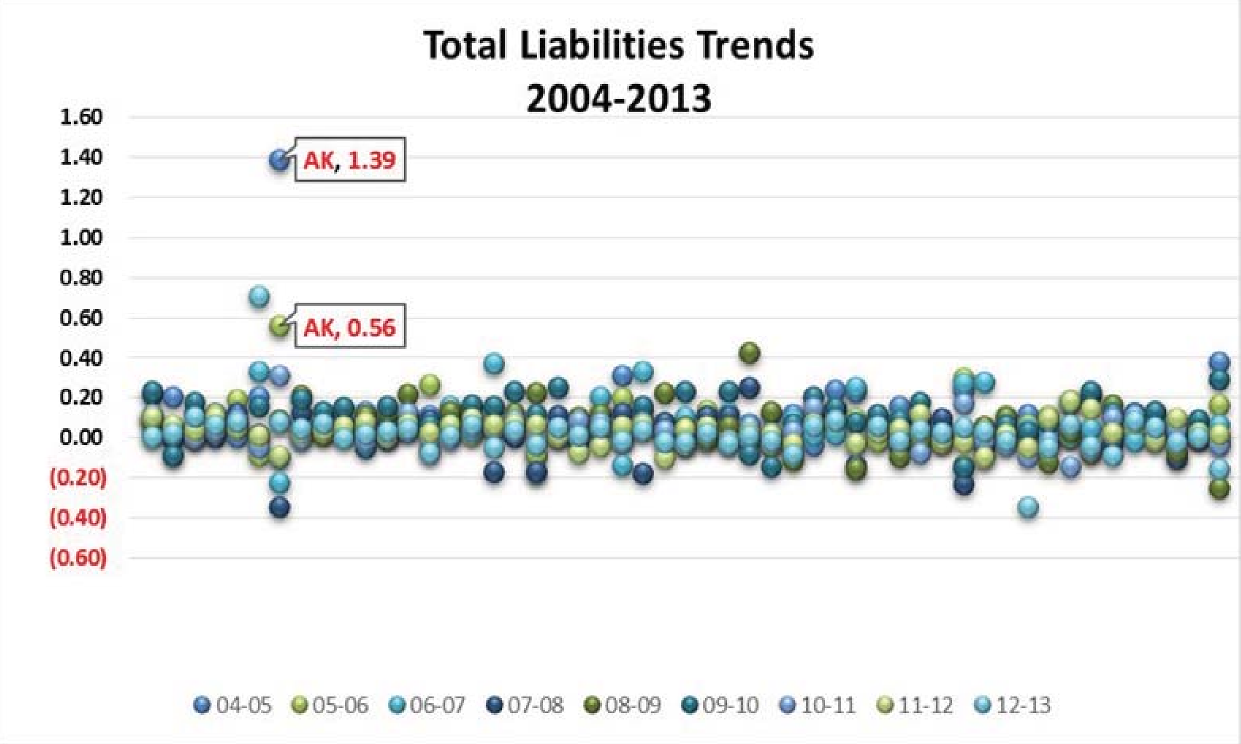
Analysis

- Ratio Analysis
- Cluster Analysis
  - ⇒ Initial cluster based on the classification of National Statistic Bureau (Geographical based)
  - ⇒ K-Mean cluster method

Expected Result

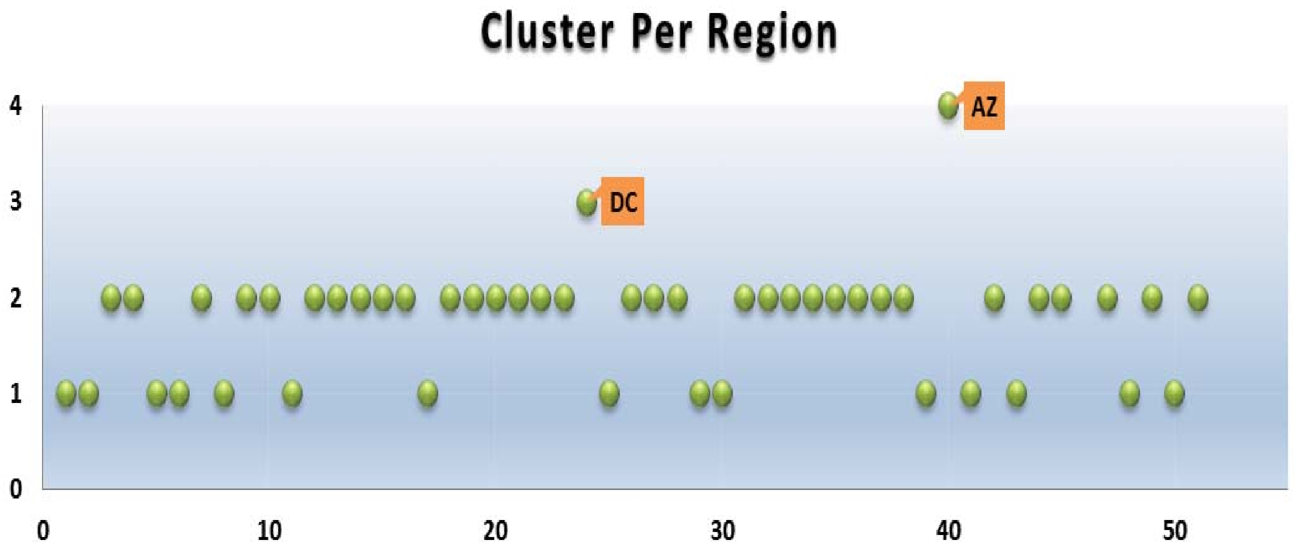
- Overall trends
- Cluster result based on the financial and non-financial information
- Anomaly Analysis

Results and Analysis



From Alaska MD&S (2004-2006):

- Change of leadership
- The state total debt increased due to the development of sport fishing infrastructure and international airport



Cluster Analysis:

- **Washington DC**
  - ⇒ Washington DC is not a state and based on U.S. Constitution this district is based on the Congress jurisdiction hence the deviation of law and budgetary structure with other states.
- **Arizona**
  - ⇒ Low percentage of Government Service Revenues/Total Program Revenues: AZ average = 0.05, whole state average = 0.15
  - ⇒ The state of Arizona is sinking in debt (State Data Lab, 2014). The State financial report shows $4.2 billion shortfall represents compensation and other costs

Conclusion

Trends and potential anomalies can be detected by implementing EDA. Furthermore, the graphical-based result can also support users understanding of the information.

Rutgers Business School
Newark and New Brunswick

Legal Risk Prediction Model for Credit Card

Feiqi Huang
Qi Liu
Miklos Vasarhelyi

Introduction

Legal risk is special and important for banking and finance. Companies are usually stuck by lawsuit which may cause extremely large expense. Meanwhile, customer's lawsuits against bank is a serious problem. Reports show larger global banks' legal tab is more than $100 billion. In addition, unlike most other operational risks, legal risk cannot be traded away in any market. However, legal risk is not like other operational risks which have been fully analyzed by quantitative analysis.

Prior literature claims that legal risk is an indicator of the weakness of internal control and reflection of bad operational performance in the future. SAS No.109 requires auditors have a sufficient understanding of the entity, environment and internal. Besides traditional audit which is backwards or retroactive, a new audit focus is forward looking or predictive. In business area, predictive models is a common way to exploit patterns found in historical data to identify risks and opportunities.

To the best of our knowledge, there is no existing literature that focuses on legal risk prediction.

Data Description

The data sets is related credit card business from a Major South American financial group. **Cardholder information data** describes each account holders' personal information which contains 289 variables and 67,049,047 observations. **Lawsuit data** records each lawsuit case's information and contains 256 variables and 1,495,673 instances. **Complain data** shows clients' complains records, which has 26 variables and 1,116,386 records. **Default data** contains 50 variables and 53,224,215 observations and presents credit card holders' default information. The last dataset is about **Credit card restriction**. It has 27 fields and 197,950,335 records. The combined data set contains 42,235,966 distinct clients and 598,431 of them (1.4%) have sued the bank.

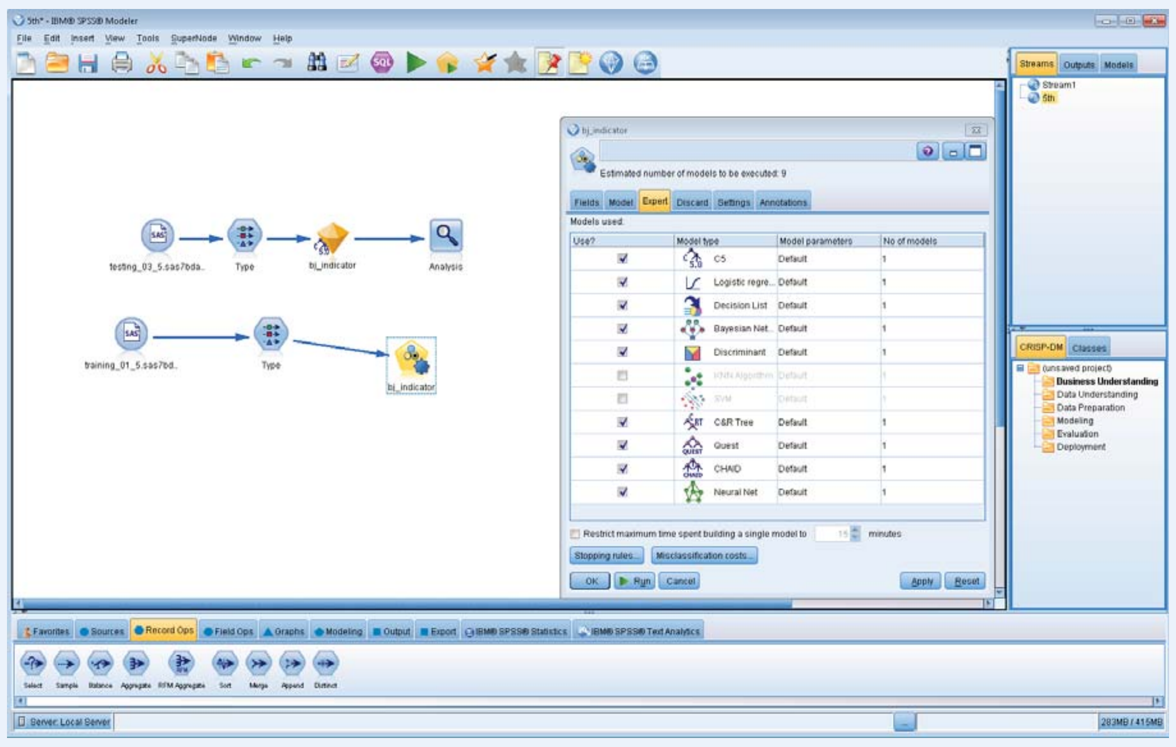
Future Work

- **Minimizing cost by cost matrix**
- **Dimension reduction**
- **Prediction potential conspired lawsuits**
- **Analyzing causes of lawsuits**

Methods & Measurements

In the process of building prediction models, authors use SAS to preprocess data and employ SPSS Modeler to build prediction models.

In the learning process, nine supervised algorithms are used to build prediction model: C5.0, CHAID, decision List, C&R tree, QUEST (Quick, Unbiased, Efficient Statistical Tree), Bayesian Network, Discriminant, Neural network and Logistic regression.



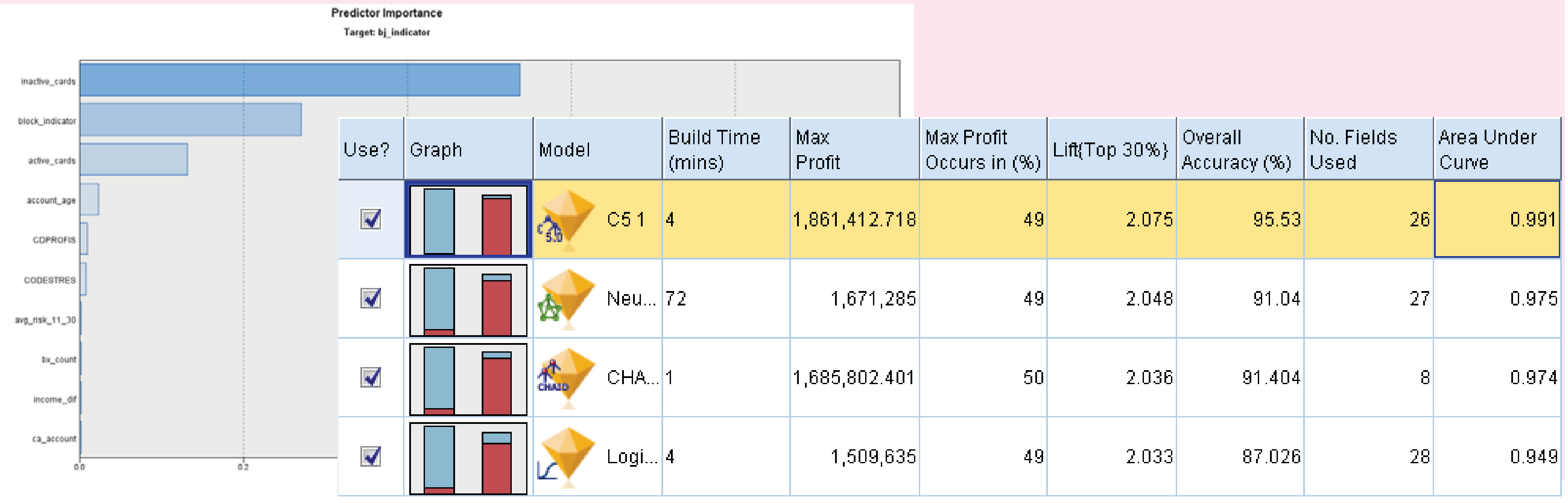
Preprocess
Learning
Measure

Data reflects less than 2% of clients have ever sued the bank. This feature (imbalance data) leads the predictive accuracy, the common measure of performance of prediction model, might not be appropriate. Receiver Operating Characteristic (ROC) curve, Recall and Precision are measurements of models performance.

Prediction Model

Trained by balanced training data, the best four algorithms are C5.0, Neural network, CHAID and logistic regression, which achieve 99.1%, 97.5%, 97.4% and 94.9% area under ROC curve respectively. When we applied the best model on testing data set, the C5.0 model achieves 95.63% Recall rate and 18.91% Precision rate. According to this model, 26 variables are used in the decision tree. The depth of the tree is 24 and contains hundreds of rules. The most five important variables in C5.0 model: number of inactive cards, indicator about whether the client's cards are blocked, number of active cards, age and indicator about whether the credit card is restricted.

Usually, trying to get higher recall will hurt precision. How to find the trade-off between recall and precision is related to many factors like business environment and management's goals. Managers and internal auditors can adjust cost matrix parameters to minimize the cost, based on the cost of failing to recognize a "lawsuit client" and misunderstanding a good customer.



Use?	Graph	Model	Build Time (mins)	Max Profit	Max Profit Occurs in (%)	Lift(Top 30%)	Overall Accuracy (%)	No. Fields Used	Area Under Curve
☑		C5 1	4	1,861,412.718	49	2.075	95.53	26	0.991
☑		Neu...	72	1,671,285	49	2.048	91.04	27	0.975
☑		CHA...	1	1,685,802.401	50	2.036	91.404	8	0.974
☑		Logi...	4	1,509,635	49	2.033	87.026	28	0.949

RUTGERS
Rutgers Business School
Newark and New Brunswick

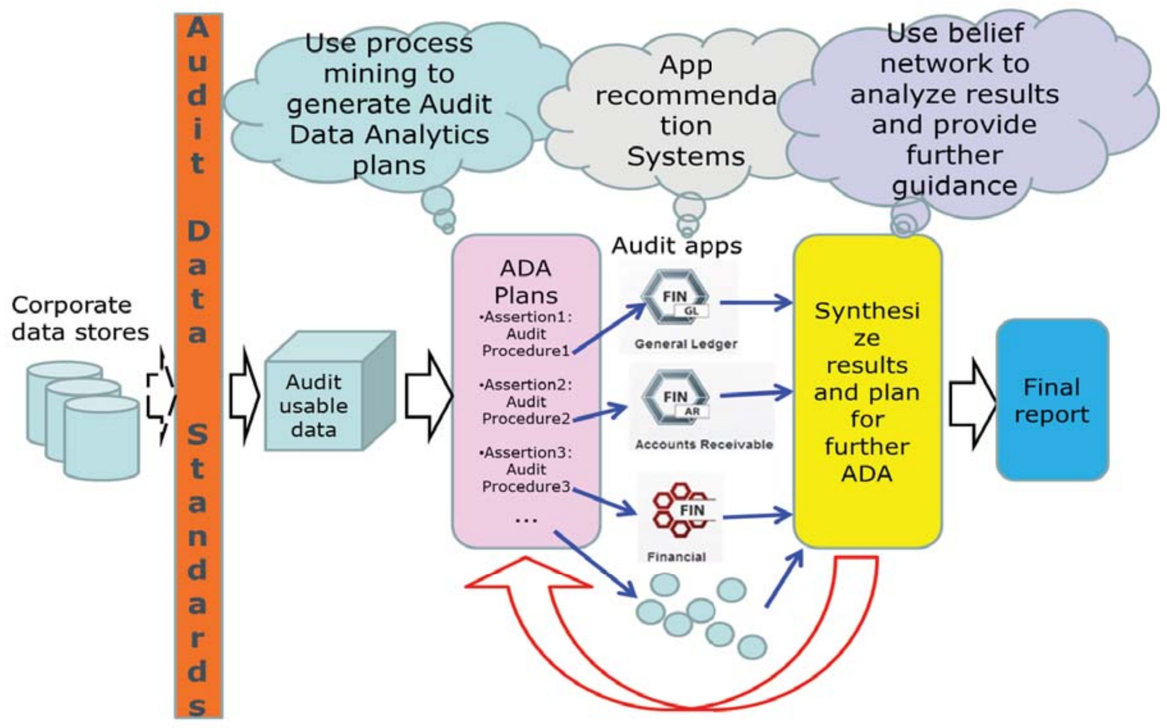
ANALYSIS OF ANALYSIS: Planning, Audit App Selection & Result Aids

Jun Dai*, JP Krahel# and Miklos A. Vasarhelyi*

* Rutgers Business School

# Loyola Univeristy Maryland

The Vision

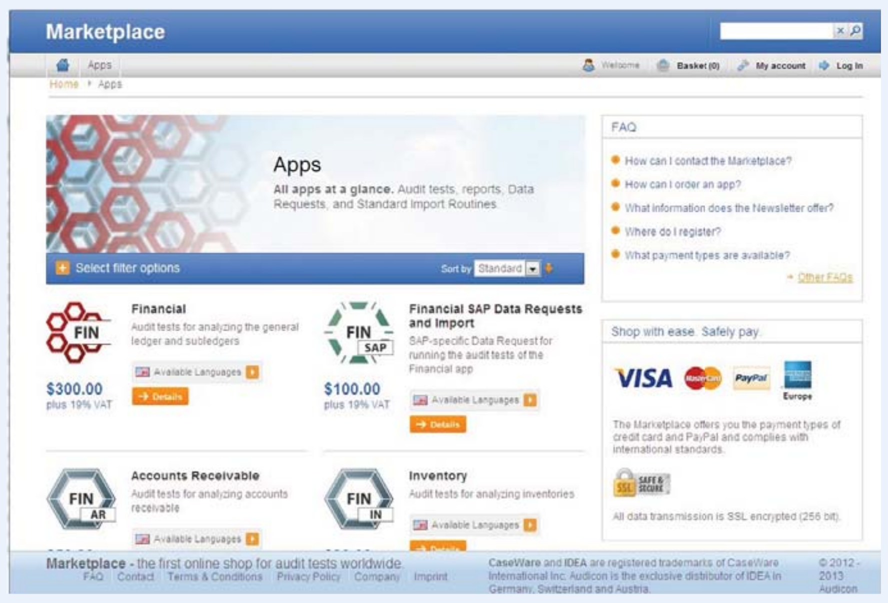

Audit Data Standards
Use process mining to generate Audit Data Analytics plans
App recommendation Systems
Use belief network to analyze results and provide further guidance
Corporate data stores
Audit usable data
ADA Plans
•Assertion1: Audit Procedure1
•Assertion2: Audit Procedure2
•Assertion3: Audit Procedure3
…
Audit apps
FIN GL
General Ledger
FIN AR
Accounts Receivable
FIN
Financial
Synthesize results and plan for further ADA
Final report

- Audit client data are standardized following the Audit Data Standard to facilitate analysis automation
- An audit plan for planning audit analytics is generated through processing mining
- The audit analytics plan is implemented by linking each audit procedure in the plan with the most appropriate audit app
- Results from all audit apps are synthesized, and used for improve the initial plan, until enough evident are collected
- The final report is generated

Audit App Selection

- Audit apps are formalized audit procedures performed through computer scripts.

Example – Caseware Marketplace



Marketplace

Apps | Welcome | Basket (0) | My account | Log In

Home › Apps

Apps

All apps at a glance. Audit tests, reports, Data Requests, and Standard Import Routines

Select filter options | Sort by Standard

Financial
Audit tests for analyzing the general ledger and subledgers
Available Languages
Details
$300.00
plus 19% VAT

Financial SAP Data Requests and Import
SAP-specific Data Request for running the audit tests of the Financial app
Available Languages
Details
$100.00
plus 19% VAT

Accounts Receivable
Audit tests for analyzing accounts receivable
Available Languages

Inventory
Audit tests for analyzing inventories
Available Languages

FAQ
- How can I contact the Marketplace?
- How can I order an app?
- What information does the Newsletter offer?
- Where do I register?
- What payment types are available?

Other FAQs

Shop with ease. Safely pay.

VISA PayPal Europe

The Marketplace offers you the payment types of credit card and PayPal and complies with international standards.

SAFE & SECURE

All data transmission is SSL encrypted (256 bit).

Marketplace - the first online shop for audit tests worldwide.
FAQ Contact Terms & Conditions Privacy Policy Company Imprint

CaseWare and IDEA are registered trademarks of CaseWare International Inc. Audicon is the exclusive distributor of IDEA in Germany, Switzerland and Austria.

© 2012 - 2013 Audicon

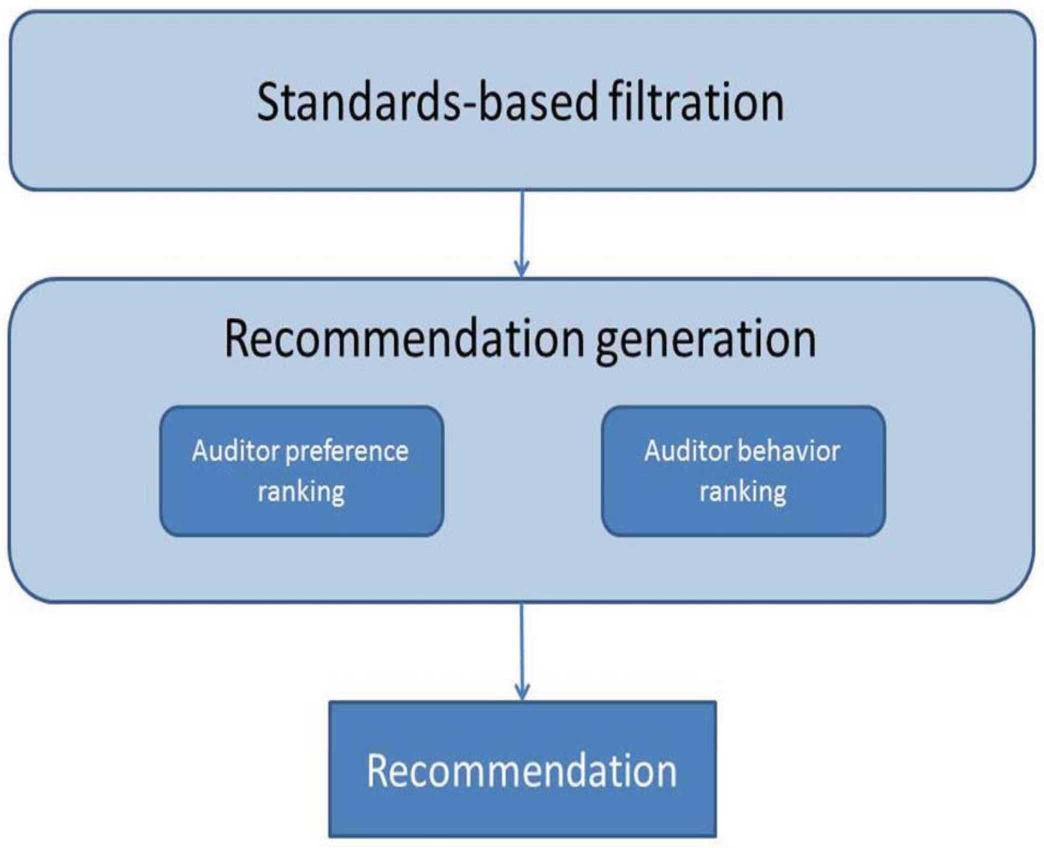
A Potential Problem

- The increase in number and variety of audit apps complicates the app selection process.
- Auditors, especially those with less experience, will likely desire guidance or assistance when selecting apps for specific audit clients.

Methodology

•The immense number and variety of tests necessitate a system that can assist auditors in discovering the most appropriate audit apps.

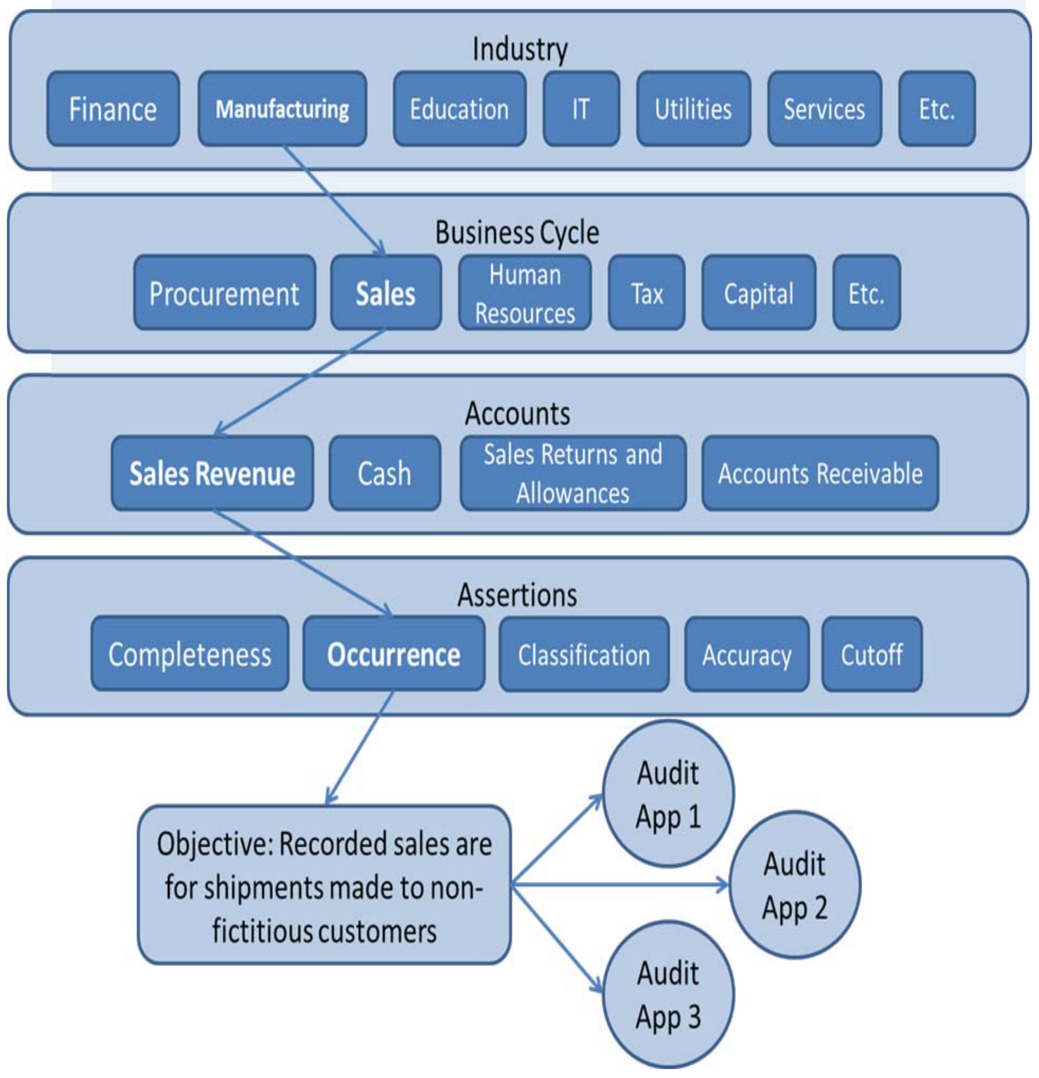
Design of an audit app recommender system


Standards-based filtration
Recommendation generation
Auditor preference ranking
Auditor behavior ranking
Recommendation

Standard-based Filtering

- The system filters audit apps by industry, business cycle, account, assertions, and audit objectives

Recommendations based on audit standards

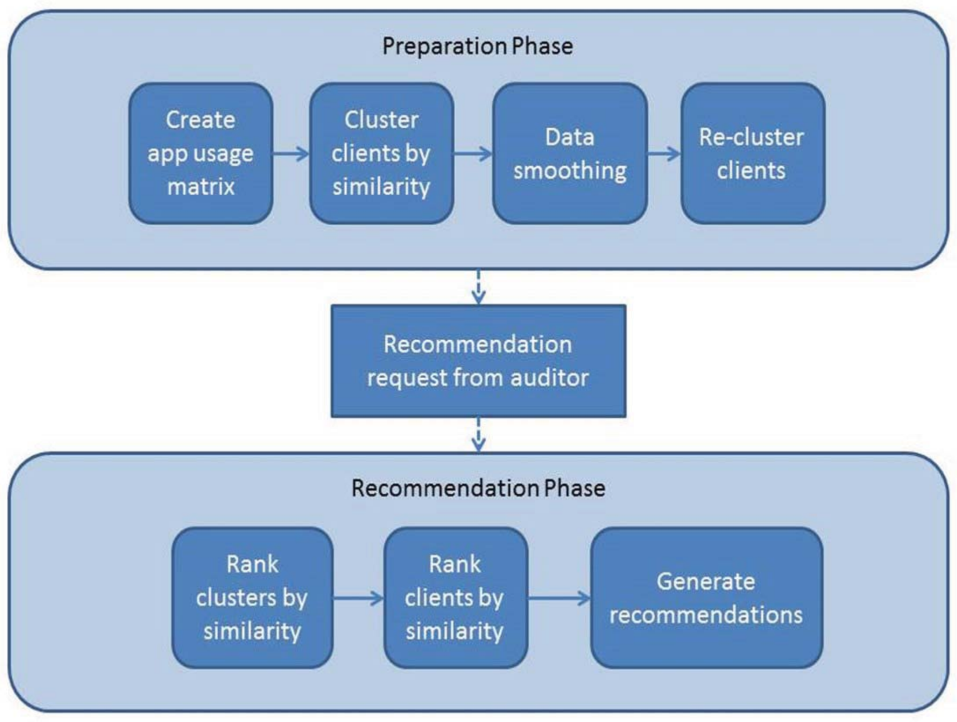


Auditors & Clients' Effects

Recommendations based on auditors' preferences

	Audit App 1	Audit App 2	Audit App 3	…
Auditor 1	5	5	1	…
Auditor 2	1	2	2	…
Auditor 3	5	5	1	…
…	….	…	…	…

Recommendations based on audit clients


Preparation Phase
Create app usage matrix
Cluster clients by similarity
Data smoothing
Re-cluster clients
Recommendation request from auditor
Recommendation Phase
Rank clusters by similarity
Rank clients by similarity
Generate recommendations

Final Recommendation

•Generate a score for each audit app by combining the predicted rating based on the auditor preference and the predicted rating based on audit clients

- *Score = δ *predicted rating from the auditor+ (1-δ) *predicted suitability of the audit app*
- Audit apps with high scores will be recommended to the auditor in a particular audit engagement

Conclusion

- In this paper, we propose an audit app recommender system to provide digital suggestion for auditor.
- By analyzing audit environment and auditors' historical behaviors, the recommender system can provide "personalized suggestions" for a particular auditor in a particular audit engagement.

RUTGERS
Rutgers Business School
Newark and New Brunswick

A Novel Method for Outlier Detection

Paul Byrnes

Abstract

Organizational fraud is a growing problem for which solutions are needed. In fact, both companies and auditors are becoming more active in addressing this problem. In alignment with this, outlier detection can assist with the fraud discovery process.

In this research, a unique, automated multivariate outlier detection method is developed and implemented. The approach relies upon four recognized measures that are used in both an individual and aggregated manner to identify anomalous objects. Individually, each measure separately determines the extent to which an object differs from a representative data point (i.e. median). In the aggregate, all measures are combined to produce an overall outlier score for each record. Preliminary results suggest that the outlier scoring method is useful for assisting with outlier detection in numerically represented data.

Introduction

Outliers have historically been described in a variety of ways. For example, Hawkins (1980) referred to an outlier as "an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism". Barnett and Lewis (1994) described an outlier as "an observation which appears to be inconsistent with the remainder of that set of data".

Irrespective of specific definition, an outlier, exception, or anomaly can be perceived as an object that is substantially different from other objects in the set to which it belongs. Outlier detection is a method for capturing those objects that are notably different from others (Zimek et al., 2014).

Background

In this study, outlier detection entails preliminary considerations. First, a suitable measure of central tendency is needed. While the mean might seem an obvious choice, it is only appropriate when the data approximates a normal distribution. However, data often deviates from this structure. Fortunately, the median is applicable in any case.

Second, the metric set to be used in anomaly detection is an important consideration because it heavily influences outcomes (Chandola et al., 2009). Zimek et al. (2014) caution that two measures of a particular type will tend to be more highly correlated than two metrics of differing types. They propose the use of ensembles in outlier detection whereby more than one measure is deployed. Given this, multiple indicators are selected for this study.

An array of potential distance measures are available for consideration, including Manhattan, Minkowski, Euclidean, and Mahalanobis. Furthermore, similarity measures exist such as the Jaccard Coefficient, Cosine Similarity, and the Tanimoto Coefficient (Tan et al., 2005). After evaluation of strengths and weaknesses, four measures are chosen: 1) Mahalanobis distance, 2) Euclidean distance, 3) Cosine similarity, and 4) Tanimoto coefficient. The last two are converted to dissimilarity measures. In this way, larger measurements always indicate higher outlier likelihood.

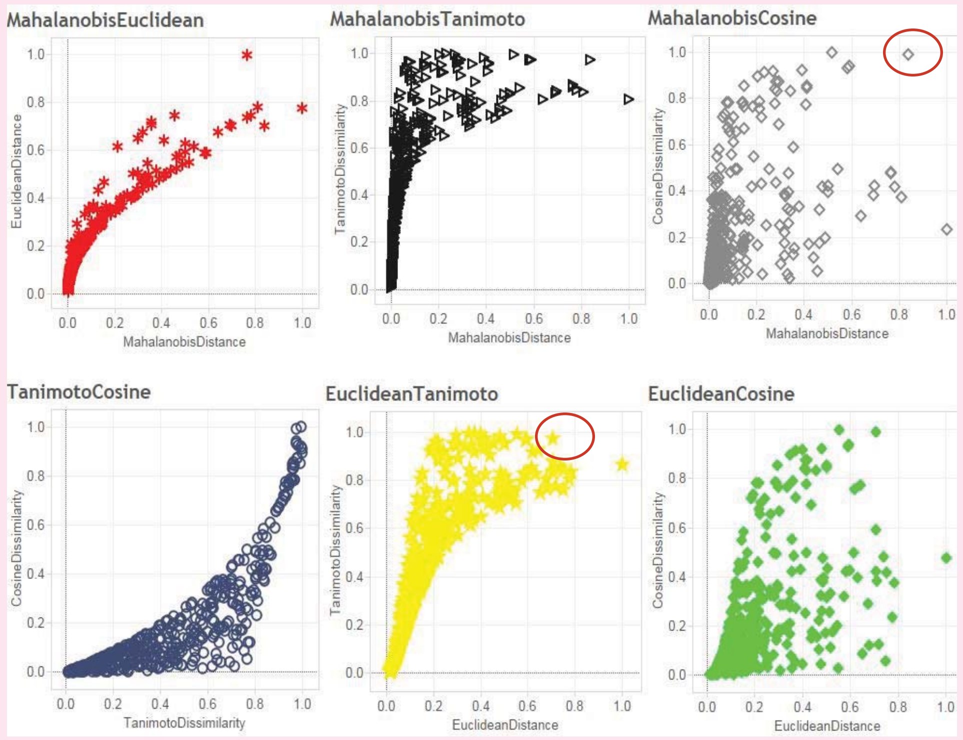
Method

In addition to initially examining each metric individually, a mechanism is used to aggregate outcomes for all four measures, thus producing an outlier score for each object. In achieving this, the results for each measure are normalized on a (0, 1] scale, which means that the maximum value for a particular metric is 1. The outlier score for a each object is computed as the sum of its normalized values for all measures. Consequently, the outlier score for a given record must lie between 0 and 4 (i.e. (0,4])

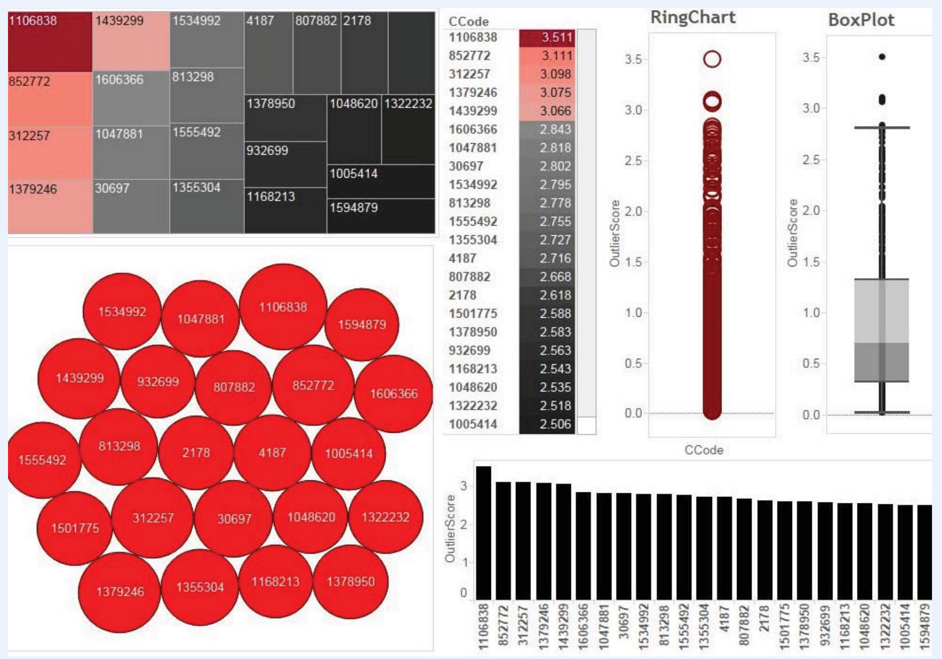
The object having the highest outlier score is deemed as most suspicious. Conversely, the record with the lowest outlier score is viewed as most similar to the median value, and, therefore, least problematic. To facilitate efficiency, the outlier detection process is substantially automated and produces rudimentary visualizations as well as an output file that can be readily explored in more sophisticated visualization software packages. In the following section, Tableau is used for image generation.

Analysis/Results

To gain initial insight, a series of plots representing all pair wise combinations of measures is created.



In each image, the median is at the origin, and objects farther from this are more anomalous. For example, the circled object in the upper right graph is identified as second most different from the median in terms of both Mahalanobis Distance and Cosine dissimilarity. This same object is again circled in the middle lower plot. While it remains far from the origin, its status as an outlier is less obvious in terms of Euclidean distance and Tanimoto dissimilarity. Next, outlier score visualizations are created to offer more specific insights.



CCode	
1106838	3.511
852772	3.111
312257	3.098
1379246	3.075
1439299	3.066
1606366	2.843
1047881	2.818
30697	2.802
1534992	2.795
813298	2.778
1555492	2.755
1355304	2.727
4187	2.716
807882	2.668
2178	2.618
1501775	2.588
1378950	2.583
932699	2.563
1168213	2.543
1048620	2.535
1322232	2.518
1005414	2.506

In the above dashboard, records with the most significant outlier scores are emphasized. For instance, 1106838 has the highest outlier score (i.e. 3.511), indicating it is most different from the median. In fact, all records with outlier scores above 3 are particularly suspicious (see box plot view). An initial data review process indicates that, as outlier scores increase, records become increasingly different from the median result.

Conclusion

Anomaly detection is becoming more important. In this study, a novel outlier detection method is developed and implemented. While this study is still evolving, initial results show that it can successfully identify and prioritize outlier candidates in numerically represented data.

RUTGERS
Rutgers Business School
Newark and New Brunswick

Internal Audit Scheduling Project

Factors affecting internal audit time duration and audit planning optimization

Qiao Li, Junming Liu, Miklos A. Vasarhelyi

Introduction

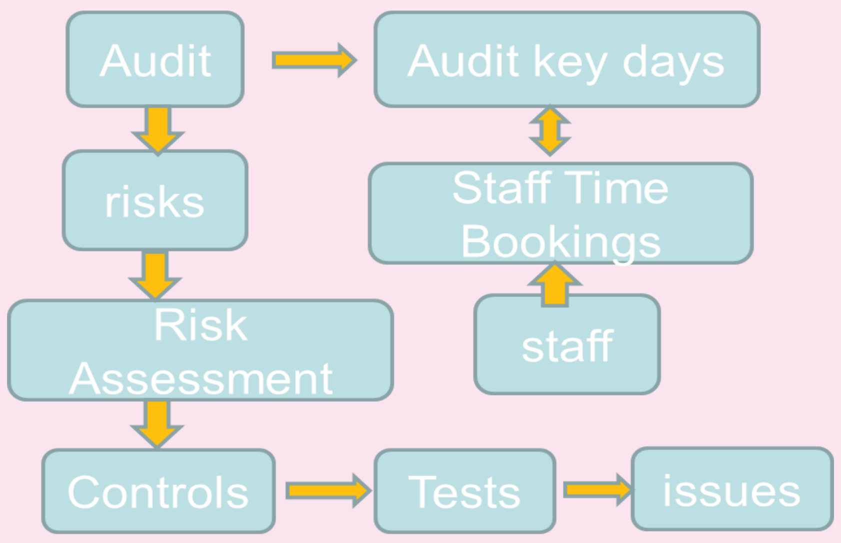
Interim Objectives:

- Which factors affect the elapsed time differences in completing audit engagements
- Budget Management
- A new risk-based audit planning/scheduling model

Preliminary Hypotheses

- Hypothesis 1: does audit elapsed time vary significantly with audit entity category?
- Hypothesis 2: do audits with higher risk levels need more audit time?
- Hypothesis 3: does the quarter affect audit time?
- Hypothesis 4: does the number and types of risks affect audit time?
- Hypothesis 5: do the issues reported after control affect audit time?

Problem Structure


Audit
Audit key days
risks
Staff Time Bookings
Risk Assessment
staff
Controls
Tests
issues

Previous study

Effective internal audit

- Consider other *factors at firm level*, such as *internal control system, organizational setting, staff expertise, auditee attitude, management support…* e.g.: Yismaw, 2007; Fadzil, 2006; Zain, 2006; E&Y, 2013; Goodwina, 2001
- Test hypotheses: *conceptual*, use *survey* or *did not do deeper analysis on real firm data*, e.g.: Mohamud,2013; Marco, 2003; Strouse,2010

Audit staff scheduling

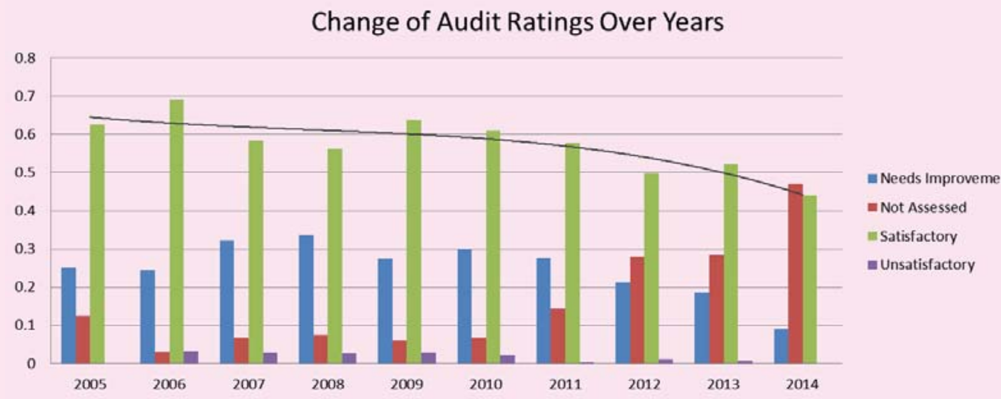
- Models to assign audit-staff to audit engagements in the most effective way (Balachandran,1981; Chan&Dodin, 1986)
- Objective linear programming (MOLP) model appropriate to the audit planning decision (Gardner, 1989)
- Integer linear program (ILP) for audit scheduling with overlapping activities and sequence-dependent setup costs (Dodin,1997)
- Linear Programming Analysis on audit staff assignment (Summers,1972)
- New TABU search procedure for audit-scheduling (Brucker, 1999)

Factors correlated with audit elapsed time

Before optimizing audit scheduling process, the reason why audit engagements take long time to issue is considered. In order to figure out which factors have significant effects on audit duration, some of the dimensions are observed and explored first: quarter the audit starts, audit rating, the level 1 entity or business line, the number of Critical, High, Medium, or Low risks, the number of reported issues, size of the audit (total number of budgeted hours), number of staff, and titles of staff working on the audit, etc.

Since the actual audit duration is unknown in the data set, it is defined as = last booking date of all staff involved in the engagement - first booking date of all staff involved.

The following graph shows the percentage of audits that were rated as satisfactory is declining across years

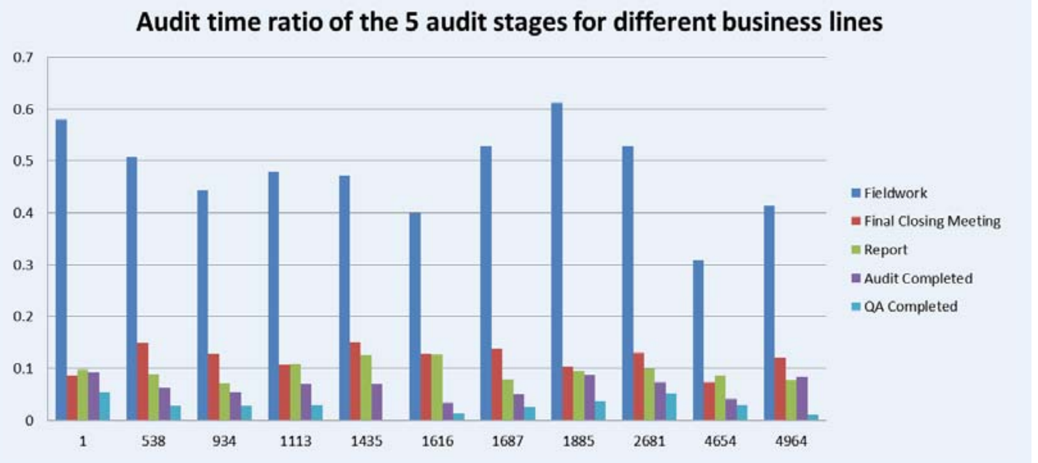


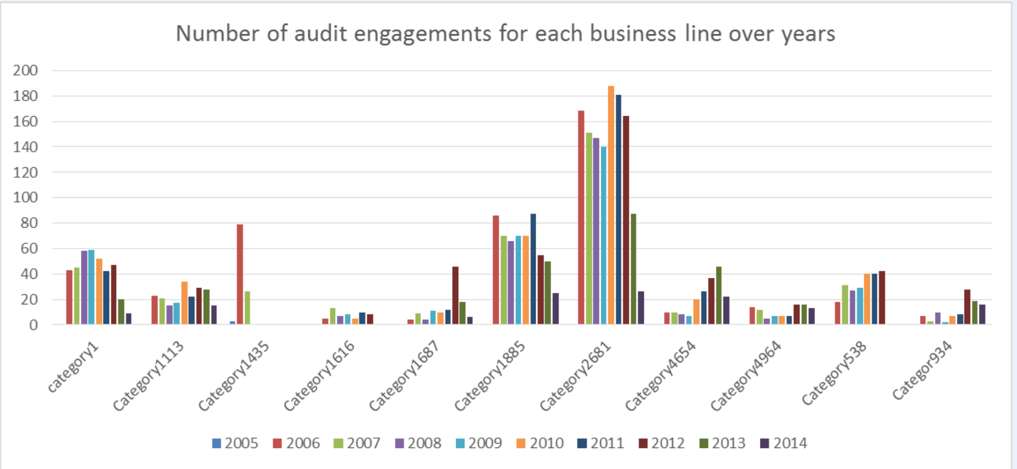
Correlation

Number of issues reported is more correlated with audit duration, and the correlation is higher if considering audit hours than days.

Pearson Correlation Coefficients, N = 3009 Prob > |r| under H0: Rho=0				
	plan-comp days	first_to_last	audit plan hours	actual hours
aud_globalRiskScore	0.05835	0.09304	0.06914	0.07520
	0.0014	<.0001	0.0001	<.0001
num of issues	0.19205	0.27156	0.46979	0.49659
	<.0001	<.0001	<.0001	<.0001

Despite Planning stage of audit scheduling, the allocation of each stage is relatively similar for all 11 business lines: fieldworks use more time than other stages.

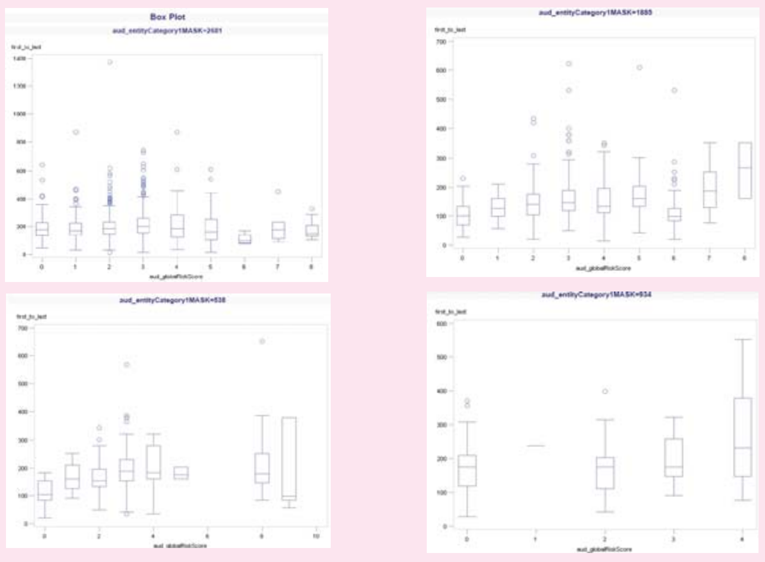




Factors Analysis

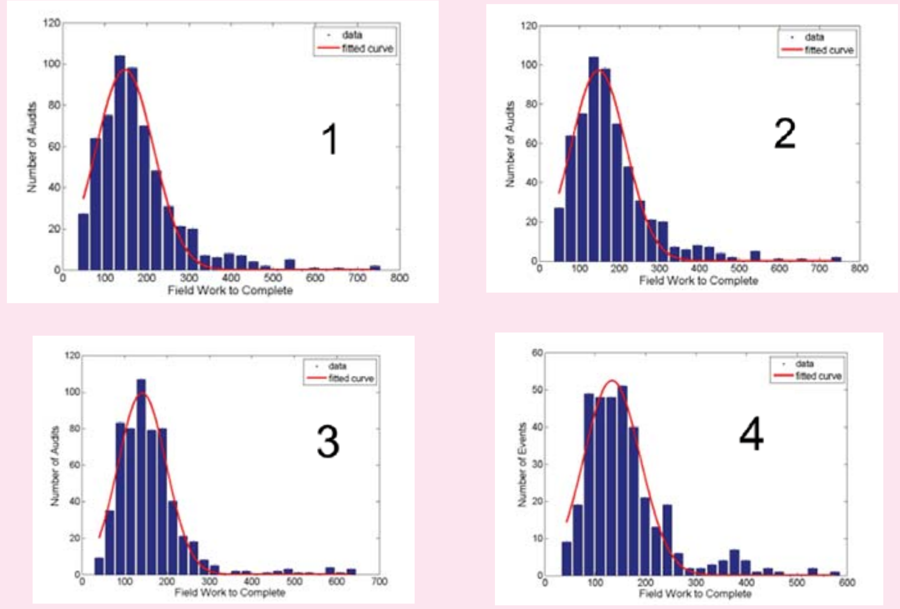
Global risk score& audit time duration (days)

-- the assumption that higher risk audits need more time applies to some business lines, but not to all.



Quarter of which the audit starts & audit time duration (days)

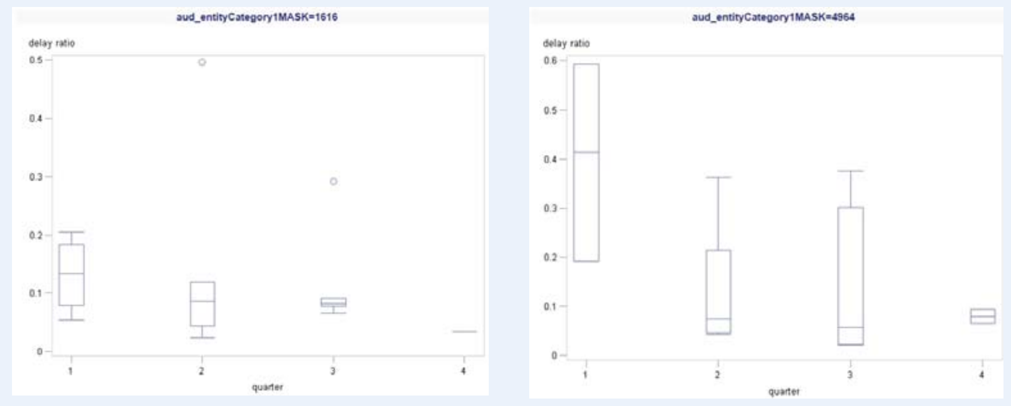
— shape of the four distributions are different, but engagements that start at quarter 4 actually were not completed in shorter period even the time is much closer to the end of a year.



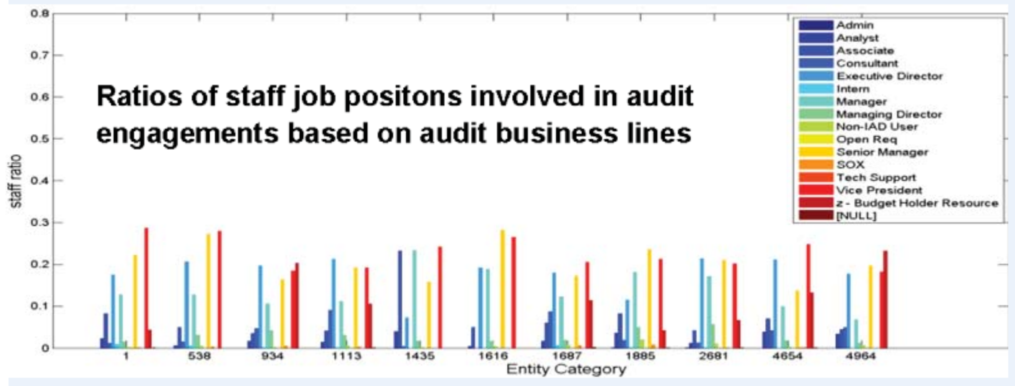
Factors Analysis

Quarter to start &delay ratio

— delay ratio = delay of days start working after planning completed/total days; for some business lines, delay ratio is higher if the audit engagement starts in early quarter (Jan. or Apr.)



Allocation of staff job position & business line



Continuing Work

- Finding out reasons of extreme and unusual
- Weighting risk assessment scores for audits
- Using classification and regression methods to predict whether an engagement will go beyond time limit

Pink Book Chapter 1: Re-conceptualizing the Continuous Audit

Miklos A. Vasarhelyi & Nancy Bumgartner

The Original “Red Book” and Its Expanded Conceptualization

	Vasarhelyi & Halper (1991), Red Book (1999)	Expanded conceptualization (1999-2014)	Notes
	CPAS / Prometheus effort	Several corporate experimental experiences	
Measuring	Metrics		
Creating a model	Standards	Of comparison	
		Of Variance	
Relating	Analytics	Representational equations	
		Continuity equations	
		Visualization	
		Dynamics self-configuring representations	
		Clustering and transaction level continuity equations	For automatic fraud detection and transaction correction
	Alarms (4 levels)		
	Measurement vs Monitoring	Measurement (indirect data acquisition)	
		Direct Data access	
		Introducing external comparative benchmarks	
		Probabilistic data relationships	Linking corporate ERP data to big data in the fringes
Dimension	Data	Continuous data audit (CDA)	Vasarhelyi & Halper 1991
	Control	Continuous Control Monitoring (CCM)	Vasarhelyi, Halper & Esawa, 1995; Alles et al, 2006
	Risk	Risks (CRMA)	Vasarhelyi, Alles, & Williams, 2010; chapter yy pinkbook
	Compliance	Compliance (CM)	Pink book chapter 1

CA Redefined

A continuous audit is a methodology that enables independent auditors to provide assurance on a subject matter, for which an entity’s management is responsible, using a continuous opinion schema issued virtually simultaneously with, or a short period of time after, the occurrence of events underlying the subject matter. The continuous audit may entail predictive modules and may supplement organizational controls. The continuous audit environment will progressively automated with auditors taking progressively higher and more progressive judgment functions. The audit will be by analytic, by exception, adaptive, and cover financial and non-financial functions.

The New CA

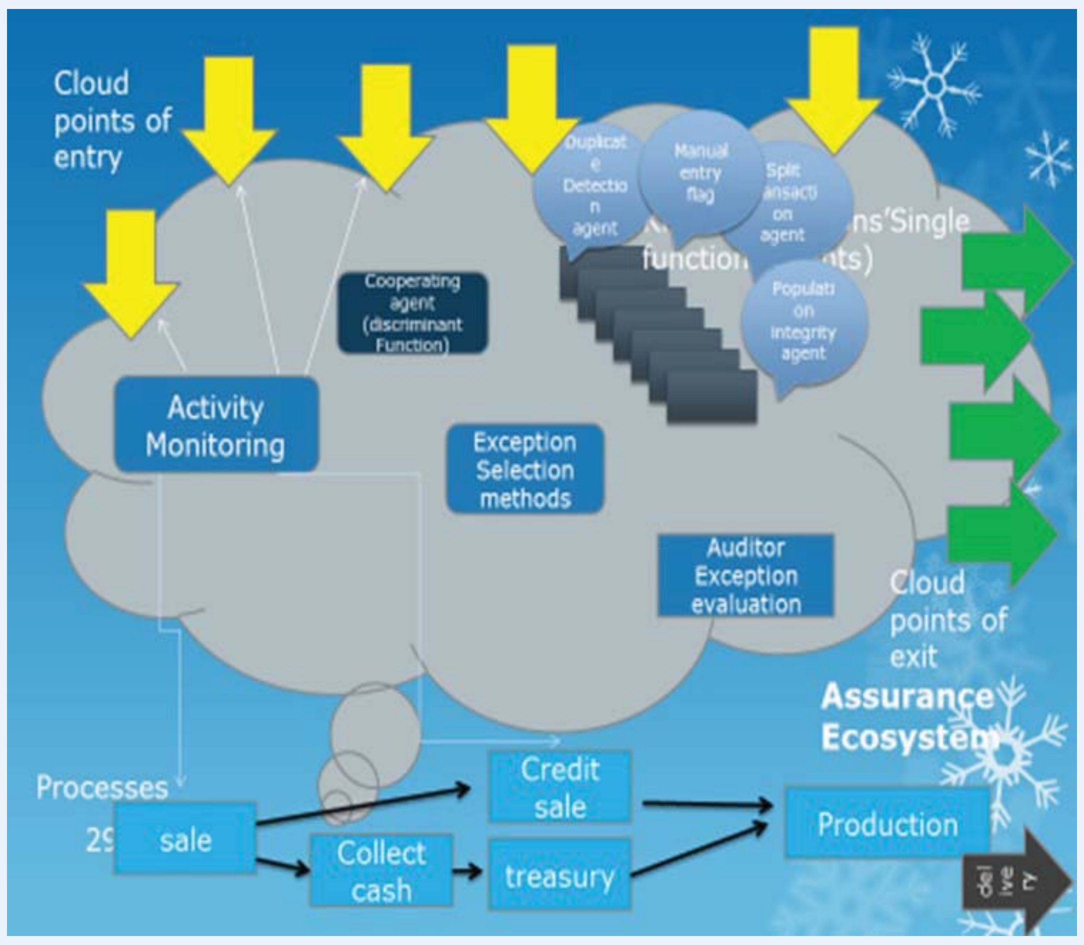
The major changes to CA that are emerging and should be permeating the audit environment and hopefully standards are:

1. Progressive adoption of a standard data interface to allow for the usage of assertion and analytic based “apps.”
2. The need to incorporate Exploratory Data Analysis into extant audit methodology. Liu (2014) proposes such a step where she expects intelligent modules to interface with a wide variety of data sources.
3. Progressive impounding of audit apps into the operating environment.
4. The evolution of an audit ecosystem with progressive level of automation over financial and non-financial systems.
5. An environment rich of software agents (krons and daemons) activated by conditions or timing and acting both over data received (inputs) from upstream system, data entry, and automatic capture and examining data to be fed to downstream systems in a predictive audit mode.

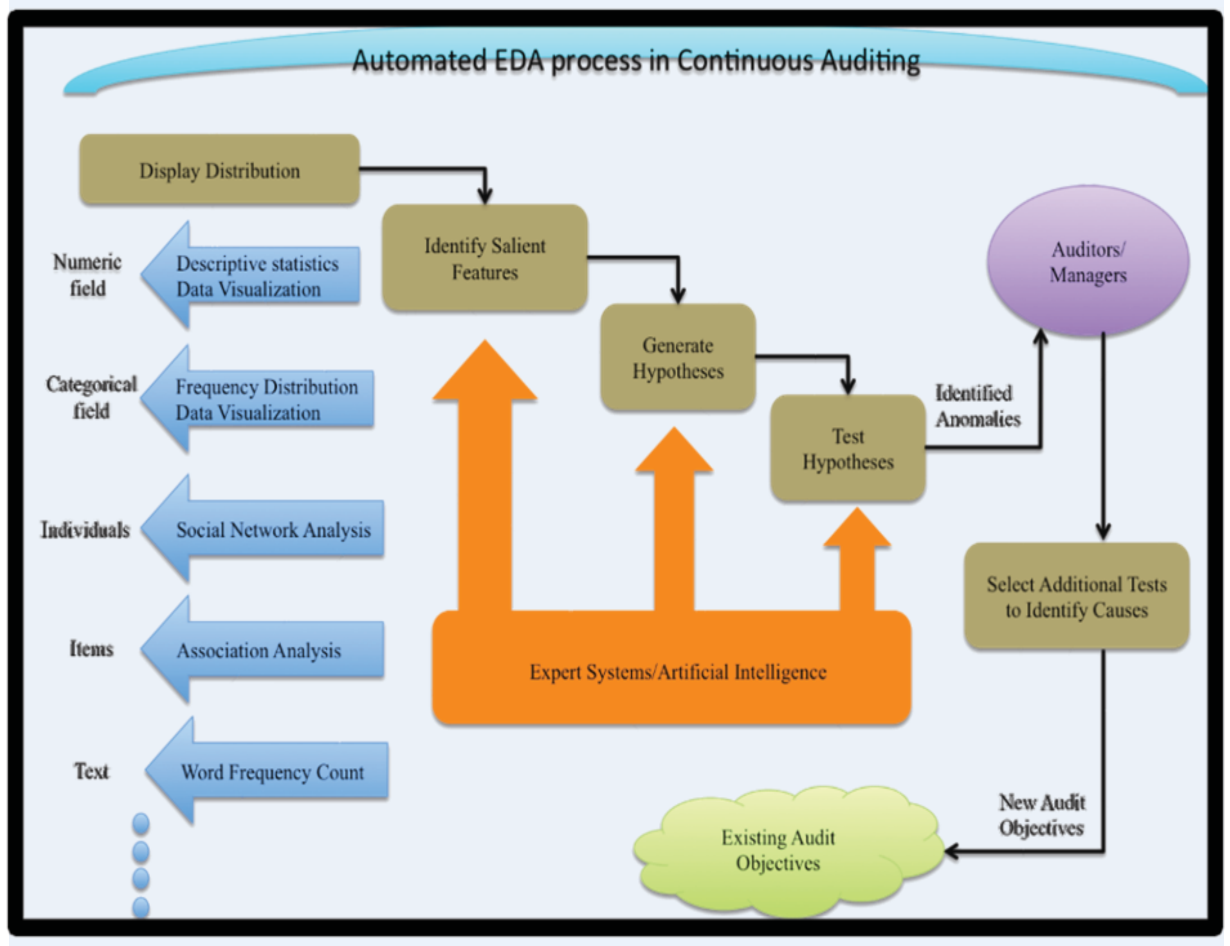
Evolving Database Structures and Their Audit (Expanded from Vasarhelyi & Halper 1991)

System Characteristic	Audit Complexity (level 1)	Audit Complexity (level 2)
Database	Documentation	Data dictionary query
Database size	User query	Auditor query
Transaction flows	Examine levels	Capture sample transactions
Duplicates	Sorting and listing	Logical analysis and indexes
Field analysis	Paper oriented	Software based
Security issues	Physical	Access hierarchies
Restart & Recovery	Plan analysis	Direct access
Database interfaces	Reconciliation	Reconciliation and transaction follow-through
Unstructured data	Linkage to know database elements	Establishment of stochastic relationships between data elements and unstructured data
Cloud storage	Access and privacy evaluation	Tests of system integrity and business continuity
Big Data	Selection of validating parameters	Linkage to data streams and extraction of meaning
		Creation of new forms of evidence
		Integration of new evidence into the traditional audit theory (Hoogduin, Yoon, and Zhang, 2015)

Envisaged within An Audit Ecosystem (Vasarhelyi and Kozlovski 2014)

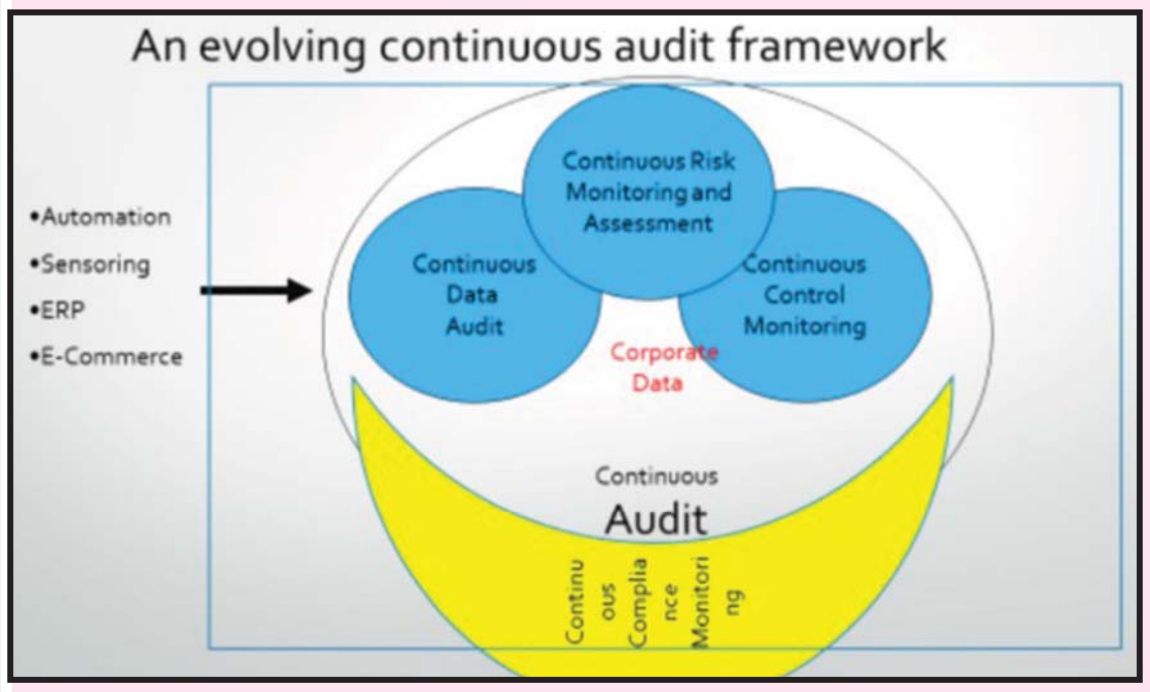


Incorporating EDA (Liu, 2014)

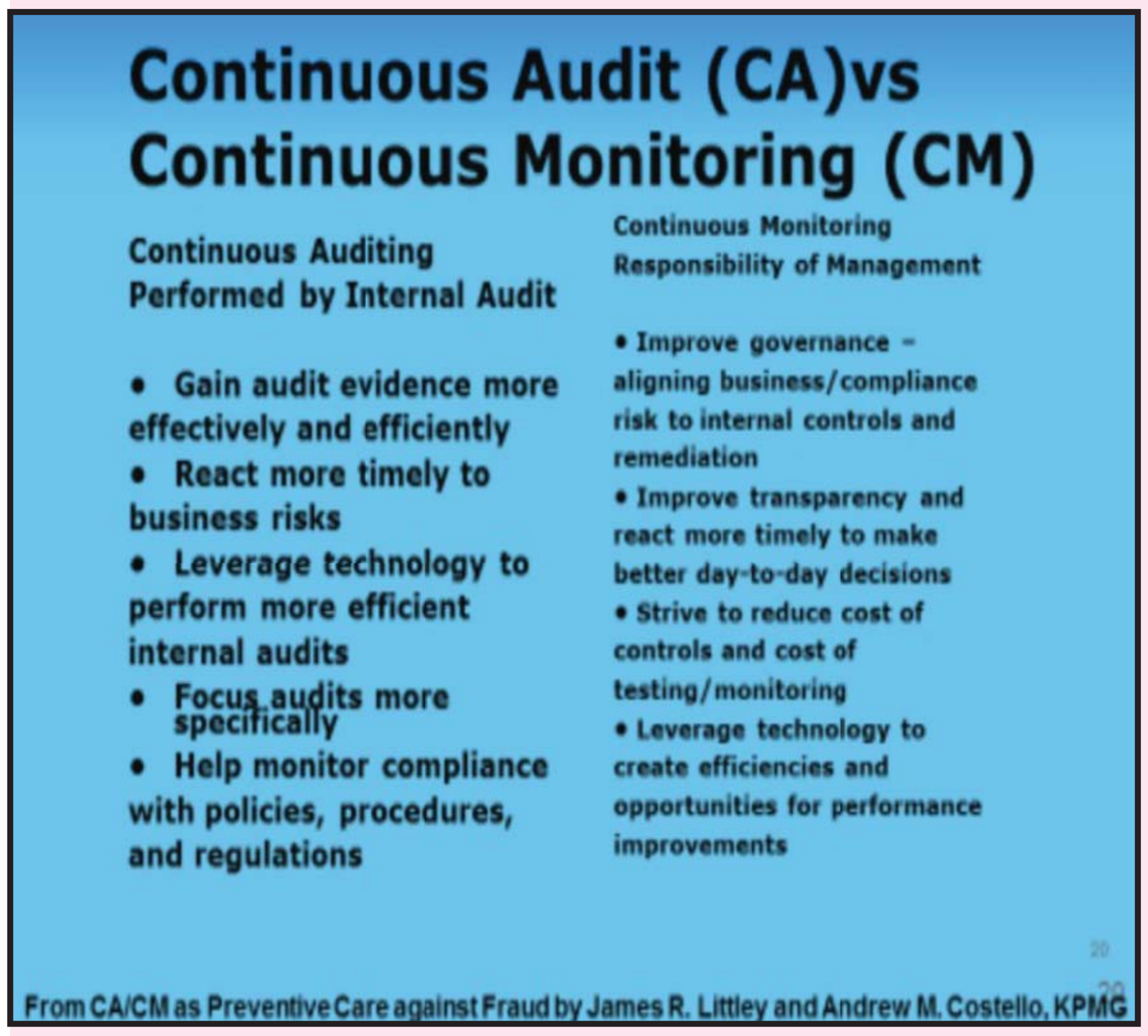

Automated EDA process in Continuous Auditing
Display Distribution
Numeric field
Descriptive statistics Data Visualization
Categorical field
Frequency Distribution Data Visualization
Individuals
Social Network Analysis
Items
Association Analysis
Text
Word Frequency Count
Identify Salient Features
Generate Hypotheses
Test Hypotheses
Identified Anomalies
Auditors/ Managers
Select Additional Tests to Identify Causes
Expert Systems/Artificial Intelligence
New Audit Objectives
Existing Audit Objectives

Several Elements of Continuous Assurance in A Real-Time Economy

CA= CDA + CCM + CRMA + COMO


An evolving continuous audit framework
•Automation
•Sensoring
•ERP
•E-Commerce
Continuous Risk Monitoring and Assessment
Continuous Data Audit
Continuous Control Monitoring
Corporate Data
Continuous Audit
Continuous Compliance Monitoring

To Which We Added Management Continuous Monitoring



Continuous Audit (CA)vs Continuous Monitoring (CM)

Continuous Auditing
Performed by Internal Audit

- Gain audit evidence more effectively and efficiently
- React more timely to business risks
- Leverage technology to perform more efficient internal audits
- Focus audits more specifically
- Help monitor compliance with policies, procedures, and regulations

Continuous Monitoring
Responsibility of Management

- Improve governance – aligning business/compliance risk to internal controls and remediation
- Improve transparency and react more timely to make better day-to-day decisions
- Strive to reduce cost of controls and cost of testing/monitoring
- Leverage technology to create efficiencies and opportunities for performance improvements

20

From CA/CM as Preventive Care against Fraud by James R. Littley and Andrew M. Costello, KPMG

Usage, Purpose, and Execution Along The Five Elements

	Data Assurance	Controls	Compliance	Risk Monitoring and Assessment	Operations (Monitoring)
Who uses					
-Management	X	X	X	X	X
-Audit (internal or external)	X	X			
-Investors	X				
-Regulators	X	X	X		
Purpose					
-Diagnostic		X	X	X	X
-Predictive				X	X
-Historic	X	X	X	X	X
Primarily performed by					
-Automation	X	X	X	X	X
-Manual		X		X	X

RUTGERS
Rutgers Business School
Newark and New Brunswick

The application of an Audit Ecosystem concept to an enterprise system

Stephen Kozlowski

Introduction

An audit ecosystem provides a technology-driven, self-sustaining audit function for firms and organizations of varying sizes and configurations. The ecosystem will leverage the digital capabilities in place at the firm. Many firms have implemented computer-based accounting systems ranging in size from PC-based packages to tailored ERP systems.

The advent of the internet provides a platform that allows for the collection of varied types and large amounts of data. Although not necessarily sourced from within the firm itself, certain forms of this data may provide insights to the firm's operations that can complement the audit function.

This current research investigates the applicability of an ecosystem approach in conjunction with an ERP system.

Continuous Audit Model

Data sources:

- Client ERP system consisting of transactional data and logs
- Other client automated systems
- Client's manual systems
- External data that will require design of an appropriate data receptacle

Financial data will be standardized to comply with Audit Data Standards

Analytic tools will analyze system logs to identify data paths to develop meta evidence to address audit risk

A tailored audit plan will be developed that considers:

- Industry
- Audit experience
- Auditor characteristics
- Incorporates analytic results of system logs and exogenous data

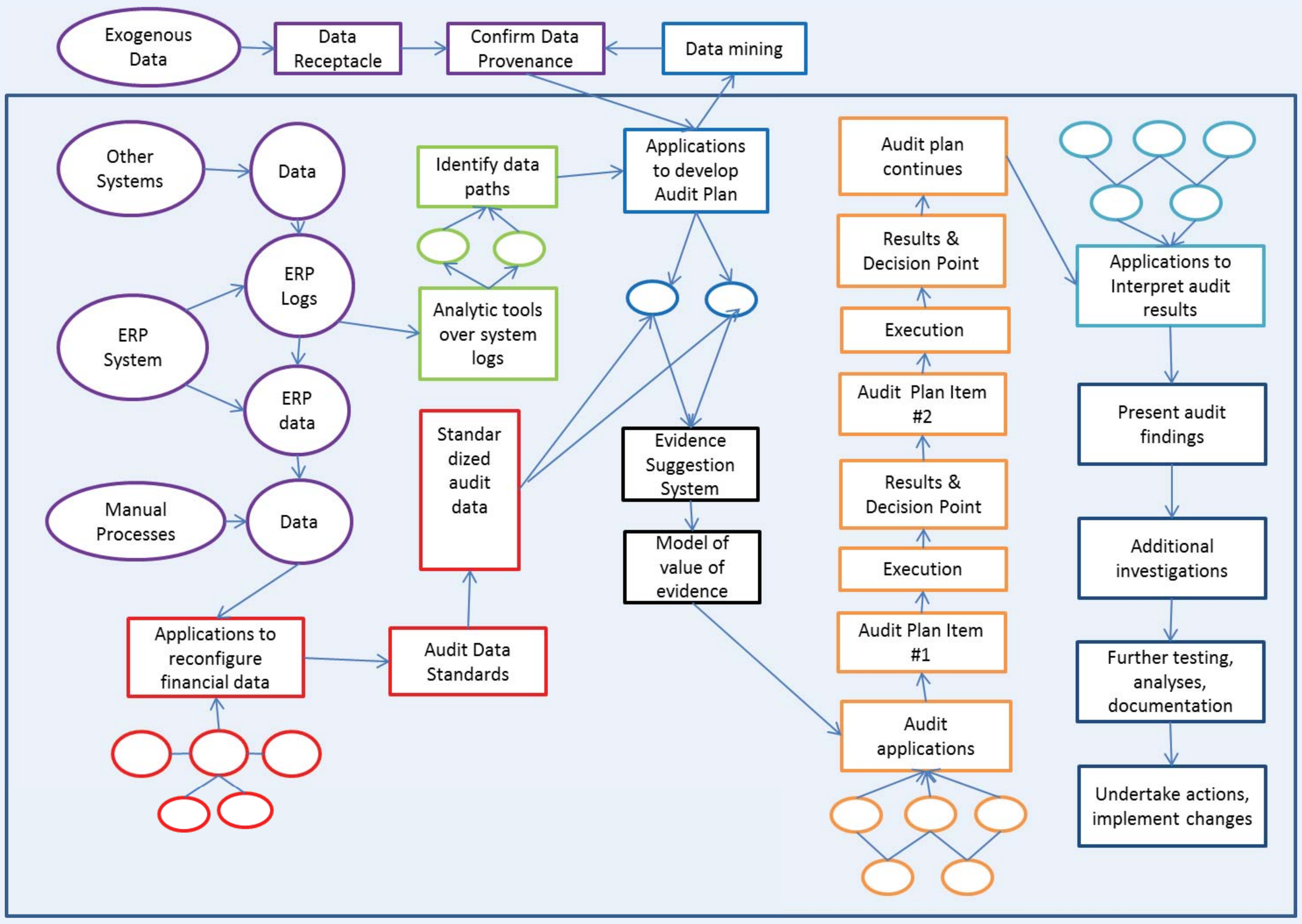
An evidence suggestion system will be developed to model the value of audit evidence

Appropriate audit applications will be identified and launched

Applications will be launched to interpret the results

Audit findings will be presented and further actions will be indicated

Proposed Design



Exogenous Data

Data Receptacle

Confirm Data Provenance

Data mining

Other Systems

Data

Identify data paths

Applications to develop Audit Plan

Audit plan continues

Results & Decision Point

Execution

Audit Plan Item #2

Results & Decision Point

Execution

Audit Plan Item #1

Audit applications

ERP Logs

ERP System

Analytic tools over system logs

ERP data

Manual Processes

Data

Standar dized audit data

Evidence Suggestion System

Model of value of evidence

Applications to reconfigure financial data

Audit Data Standards

Applications to Interpret audit results

Present audit findings

Additional investigations

Further testing, analyses, documentation

Undertake actions, implement changes

Related Activities

–ERP data for A/R, A/P, and G/L was obtained from one NFP client, typical CA/CM techniques were applied, and the results were provided to the client

-Payroll and H/R data was provided from a second client, appropriate audit tests as requested by the client were applied, and results provide to the client using spreadsheets and dashboards

An automated testing routine was developed and implantation is underway by the client

-Project planning is underway with a third client who has indicated they will provide Payroll, G/L, and A/P data for analytical purposes

Rutgers Business School
Newark and New Brunswick

k

Auditing Analytical Procedure Techniques:
Does Process Mining Complement or Substitute Data Mining?

Tiffany Chiu and Miklos Vasarhelyi

Introduction

Unlike traditional auditing analytical procedure, process mining of event logs provides a new aspect for audit in the way that this technique analyzes and processes transaction data for each and every business event instead of relying on only a sample of the population. Prior literature indicated that both process mining and data mining techniques can add value and improve the performance of analytical procedures in auditing. However, it is still not clear whether process mining of event logs and data mining techniques should be applied together in a complementary fashion or process mining of event logs could replace data mining techniques.

This study aims at analyzing and comparing the performance of process mining and data mining techniques in auditing analytical procedure using Volvo IT dataset, and distinguish whether process mining complements or substitutes data mining technique.

Process Mining of Event Logs

Process mining refers to the usage of event logs to analyze business processes. There are four characteristics that must be extracted from each event in the system in order to analyze the data:

Characteristics of Event	
(1) Activity	The ***activity*** taking place during the event (e.g. sign)
(2) Process Instance	The ***process instance*** of the event (e.g. invoice)
(3) Originator	The ***originator***, or party responsible for the event (e.g. action owner)
(4) Timestamp	The ***timestamp*** of the event or the date/time of the event (e.g. 2006-11-07T10:00:36)

Prior studies proposed that when utilizing process mining techniques to analyze the information from event logs, five different types of analysis can be performed in process mining:

Process discovery	Exploring the business process to see if there are any anomalies or unusual transactions
Conformance check	Conducting a confirmation as to whether the process reality matches the expectation or standard
Performance analysis	Measuring business process performance (KPI's)
Social network analysis	Utilizing information contained in the event log to identify which authorized user entered each transaction to detect whether anomalous relationships and/or collusive fraud exist
Decision mining and verification	Focusing on decision points in a discovered process model and using them to test assertions on a case by case basis

Literature Review

❖Application of Process Mining in Audit

- Jans et al. (2009) proposed a framework for reducing internal fraud risk based on process mining event logs.
- Jans et al. (2013) discovered that process mining can add value to audit as it enhances the effectiveness of fraud prevention, especially when auditees are made aware of event logs.
- Jans et al. (2014) applied process mining of event logs in auditing analytical procedures, and successfully detected anomalous transactions that traditional auditing analytical procedures may fail to discover.

❖Application of Cluster Analysis in Audit

- Thiprungsri (2010) applied cluster analysis to group transactions of transitory accounts; results indicated that cluster analysis is useful for detecting anomalous transactions in audit.
- Thiprungsri and Vasarhelyi (2011) examined life insurance claims using clustering and proposed that cluster analysis is a promising technique that can be integrated into the concepts of continuous system monitoring and assurance.

Methodology and Dataset

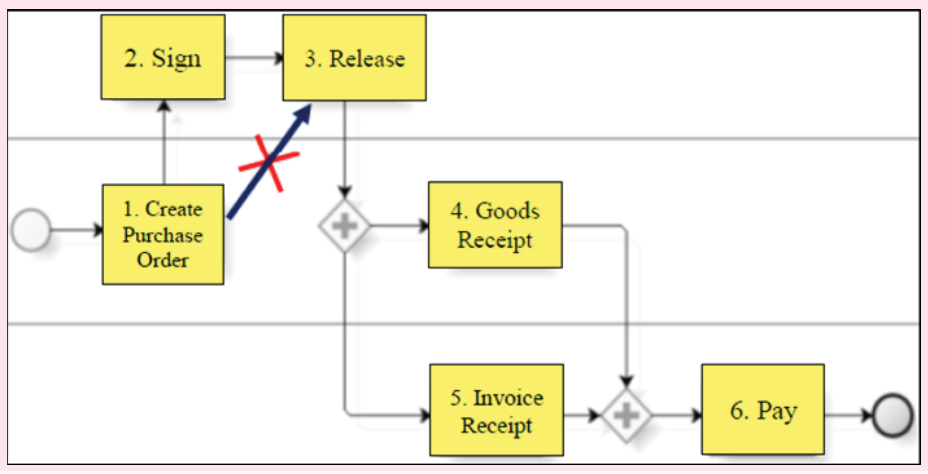
This study applied process mining and data mining techniques, respectively, to analyze a real life Volvo IT dataset. The unsupervised learning algorithm (cluster analysis) – K-mean and Fuzzy Miner technique in process mining are employed to analyze and compare the data.

Volvo IT Problem Management	
Total Number of Process Instances (cases)	819
Total Number of Events	2,351
Problem Status	3
Problem Sub-Status	5
Problem Involved Action Owner	240

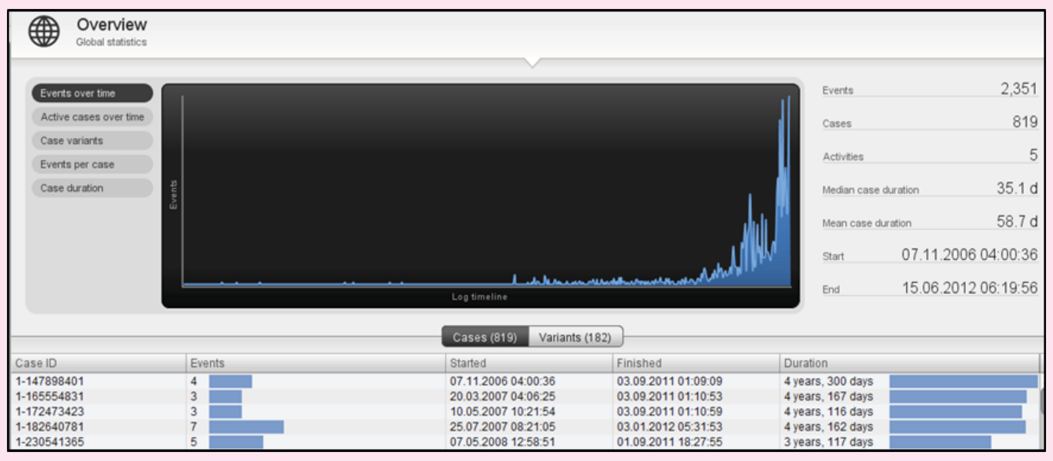
Future Research

Process mining may enhance the performance of "Audit by Exception" concept proposed by Vasarhelyi and Halper (1991). Audit by exception refers to the usage of CPAM in audit procedure so that the audit works will be focused on the alarm of exception gathered by the system on a continuous basis.

Application of Process Mining with "Audit by Exception": An alarm will arise when purchase order is ***released*** without proper ***sign***. The Figure below shows the example process; the flow chart is a procurement process extracted and revised from Jans et al. (2014).

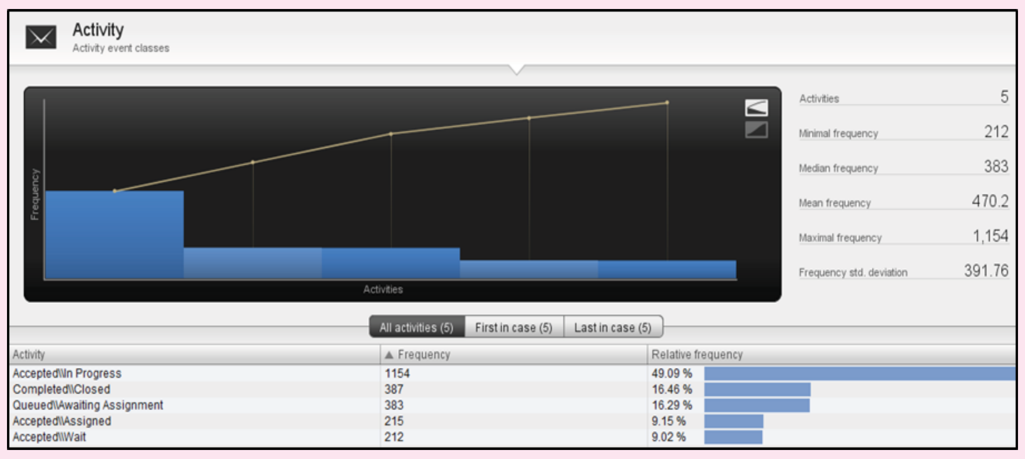


Preliminary Analysis and Expected Results



Events 2,351
Cases 819
Activities 5
Median case duration 35.1 d
Mean case duration 58.7 d
Start 07.11.2006 04:00:36
End 15.06.2012 06:19:56

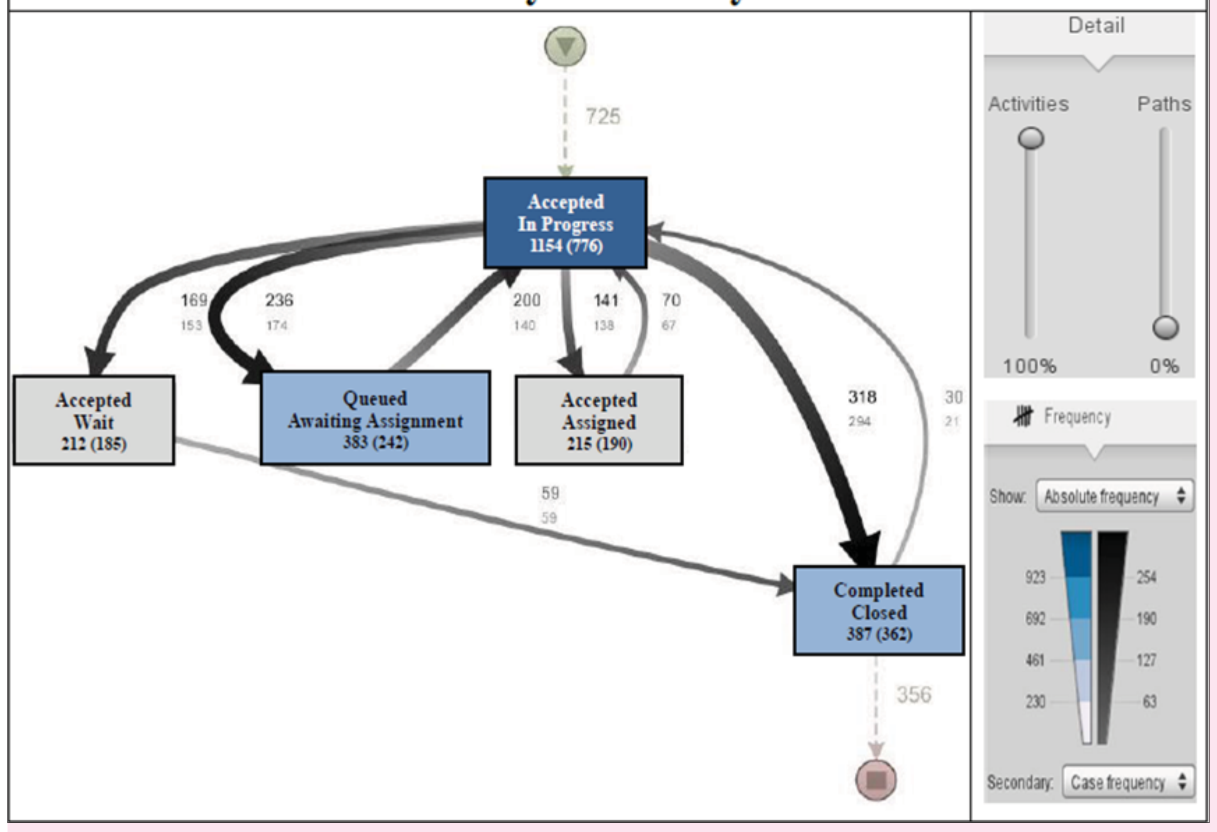
Case ID	Events	Started	Finished	Duration
1-147898401	4	07.11.2006 04:00:36	03.09.2011 01:09:09	4 years, 300 days
1-165554831	3	20.03.2007 04:06:25	03.09.2011 01:10:53	4 years, 167 days
1-172473423	3	10.05.2007 10:21:54	03.09.2011 01:10:59	4 years, 116 days
1-182640781	7	25.07.2007 08:21:05	03.01.2012 05:31:53	4 years, 162 days
1-230541365	5	07.05.2008 12:58:51	01.09.2011 18:27:55	3 years, 117 days



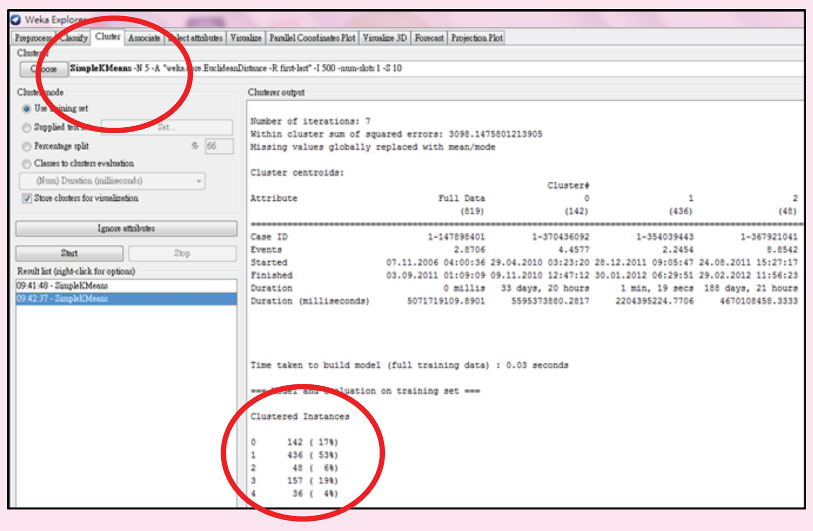
Activities 5
Minimal frequency 212
Median frequency 383
Mean frequency 470.2
Maximal frequency 1,154
Frequency std. deviation 391.76

Activity	▲ Frequency	Relative frequency
Accepted\\In Progress	1154	49.09 %
Completed\\Closed	387	16.46 %
Queued\\Awaiting Assignment	383	16.29 %
Accepted\\Assigned	215	9.15 %
Accepted\\Wait	212	9.02 %

Data Analysis – Fuzzy Miner



This study will conduct audit analytical procedures and compare results from process mining techniques and cluster analysis using the Volvo IT dataset. The comparison of process mining and cluster analysis can be done by analyzing (1) frequent patterns (Variants) from process mining techniques, and (2) cases that have been grouped together through cluster analysis. For example, the study can compare results from the two different techniques and determine which method discovers more anomalous transactions.

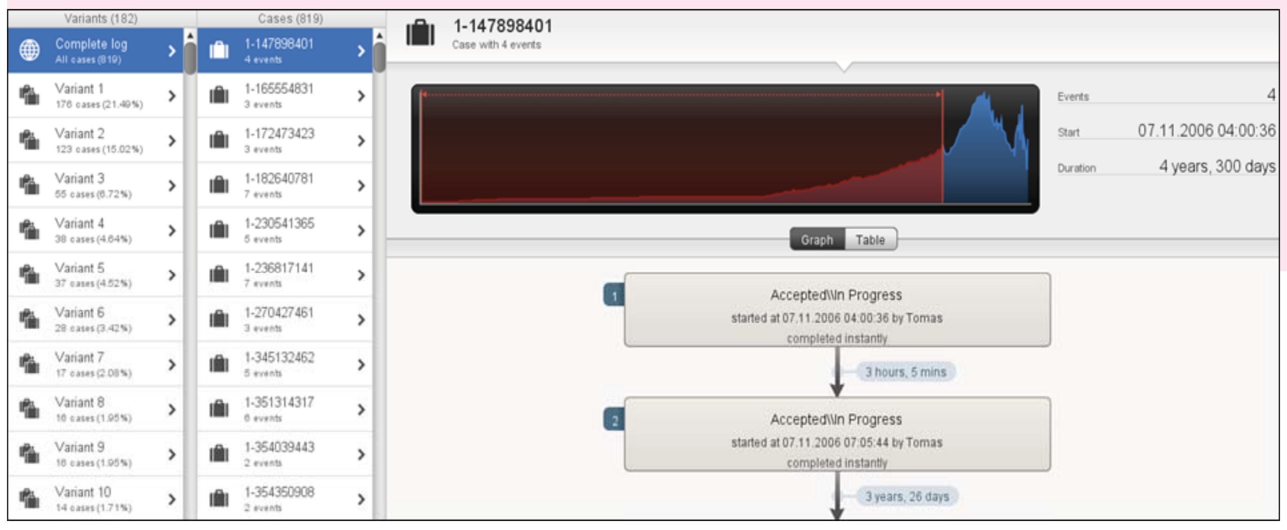


SimpleKMeans -N 5 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

○ Use training set
○ Percentage split % 66
☑ Store clusters for visualization

09:41:48 - SimpleKMeans
09:42:37 - SimpleKMeans

```
Number of iterations: 7
Within cluster sum of squared errors: 3098.1475801213905
Missing values globally replaced with mean/mode

Cluster centroids:
                                                  Cluster#
Attribute                     Full Data              0                  1                  2
                                  (819)            (142)              (436)               (48)
============================================================================================
Case ID                     1-147898401        1-370436092        1-354039443        1-367921041
Events                           2.8706             4.4577             2.2454             8.8542
Started             07.11.2006 04:00:36 29.04.2010 03:23:20 28.12.2011 09:05:47 24.08.2011 15:27:17
Finished            03.09.2011 01:09:09 09.11.2010 12:47:12 30.01.2012 06:29:51 29.02.2012 11:56:23
Duration                       0 millis   33 days, 20 hours    1 min, 19 secs  188 days, 21 hours
Duration (milliseconds)  5071719109.8901    5595373880.2817    2204395224.7706    4670108458.3333

Time taken to build model (full training data) : 0.03 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      142 ( 17%)
1      436 ( 53%)
2       48 (  6%)
3      157 ( 19%)
4       36 (  4%)
```



Complete log — All cases (819)
Variant 1 — 176 cases (21.49%)
Variant 2 — 123 cases (15.02%)
Variant 3 — 55 cases (6.72%)
Variant 4 — 38 cases (4.64%)
Variant 5 — 37 cases (4.52%)
Variant 6 — 28 cases (3.42%)
Variant 7 — 17 cases (2.08%)
Variant 8 — 16 cases (1.95%)
Variant 9 — 16 cases (1.95%)
Variant 10 — 14 cases (1.71%)

1-147898401 — 4 events
1-165554831 — 3 events
1-172473423 — 3 events
1-182640781 — 7 events
1-230541365 — 5 events
1-236817141 — 7 events
1-270427461 — 3 events
1-345132462 — 5 events
1-351314317 — 6 events
1-354039443 — 2 events
1-354350908 — 2 events

1-147898401
Case with 4 events

Events 4
Start 07.11.2006 04:00:36
Duration 4 years, 300 days

1 Accepted\\In Progress
started at 07.11.2006 04:00:36 by Tomas
completed instantly

3 hours, 5 mins

2 Accepted\\In Progress
started at 07.11.2006 07:05:44 by Tomas
completed instantly

3 years, 26 days

RUTGERS
Rutgers Business School
Newark and New Brunswick

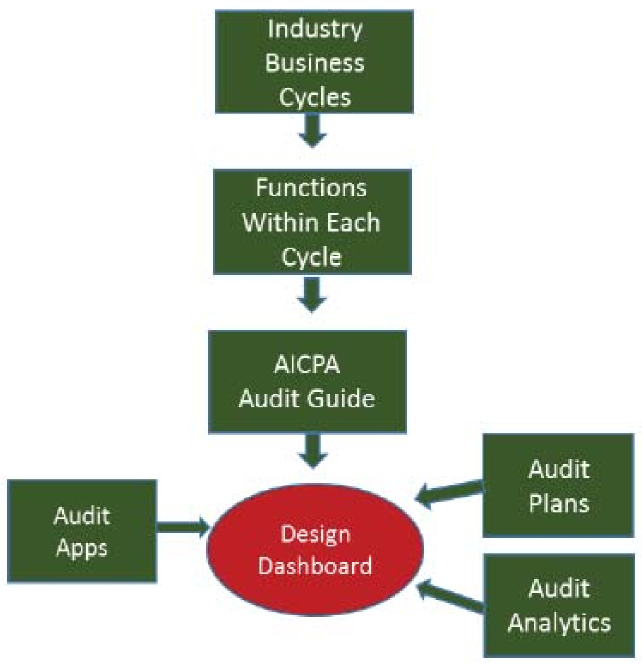
Interactive Auditor Dashboard:

Application On Life Insurance

Basma Moharram and Miklos Vasarhelyi

What to Dashboard?

Our objective is to create an auditor dashboard to assist the auditor in designing and performing his audit plan. The first question we had to ask ourselves was what to dashboard. To answer this question we followed this approach; We start with a specific industry (Insurance). We break down into its main business cycles. We then break each business cycle into its main functions. For each main function we think of the possible assertions the auditor would want to test. In deciding the assertions we use the AICPA audit guide, audit plans, audit analytics, and audit apps.



Industry Business Cycles

Functions Within Each Cycle

AICPA Audit Guide

Audit Apps

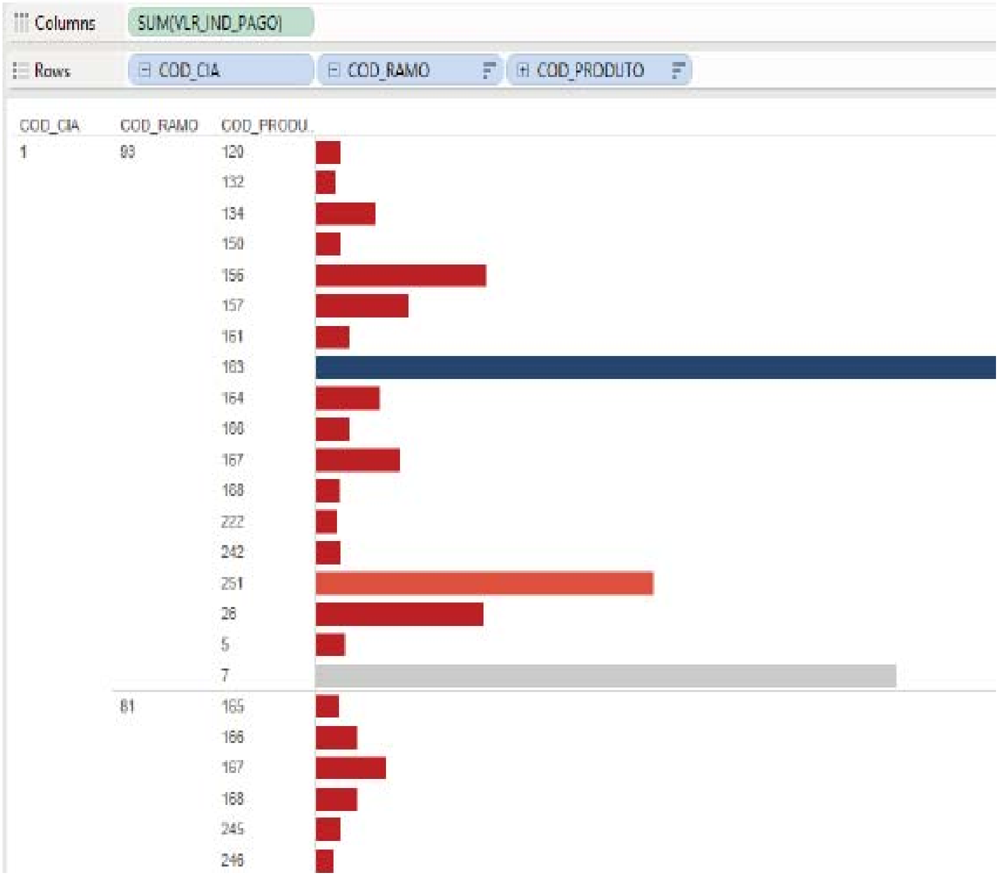
Design Dashboard

Audit Plans

Audit Analytics

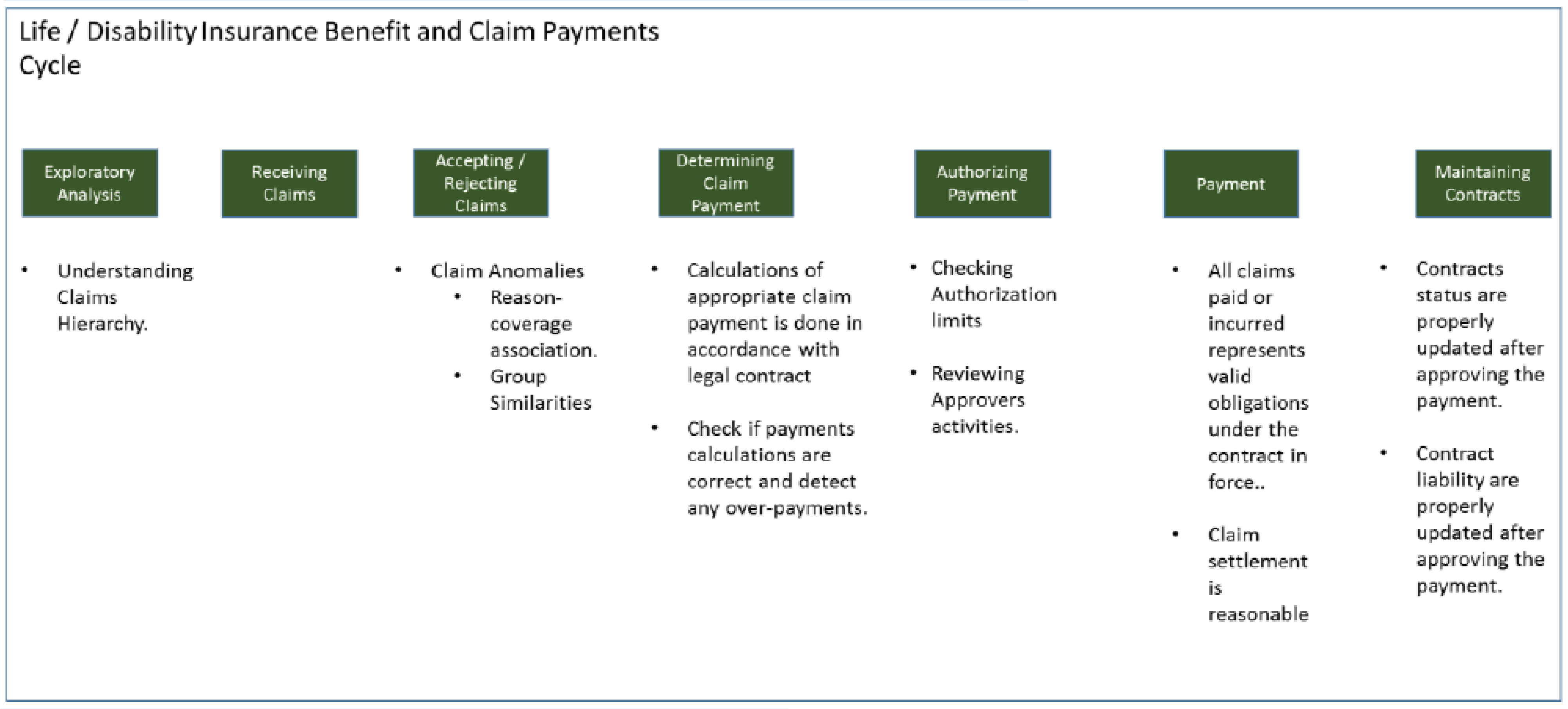
Exploratory Analysis

The Chart shows the amount of Claim payments made to clients by company code. The auditor can filter for a specific range of payments. He can drill down from Company level, down to type of insurance, to type of product, until we go down to each single claim. Using this graph, an auditor will gain an understanding of the claim payments made by different companies.



Columns: SUM(VLR_IND_PAGO)

Rows: COD_CIA, COD_RAMO, COD_PRODUTO



Life / Disability Insurance Benefit and Claim Payments Cycle

Exploratory Analysis
- Understanding Claims Hierarchy.

Receiving Claims

Accepting / Rejecting Claims
- Claim Anomalies
  - Reason-coverage association.
  - Group Similarities

Determining Claim Payment
- Calculations of appropriate claim payment is done in accordance with legal contract
- Check if payments calculations are correct and detect any over-payments.

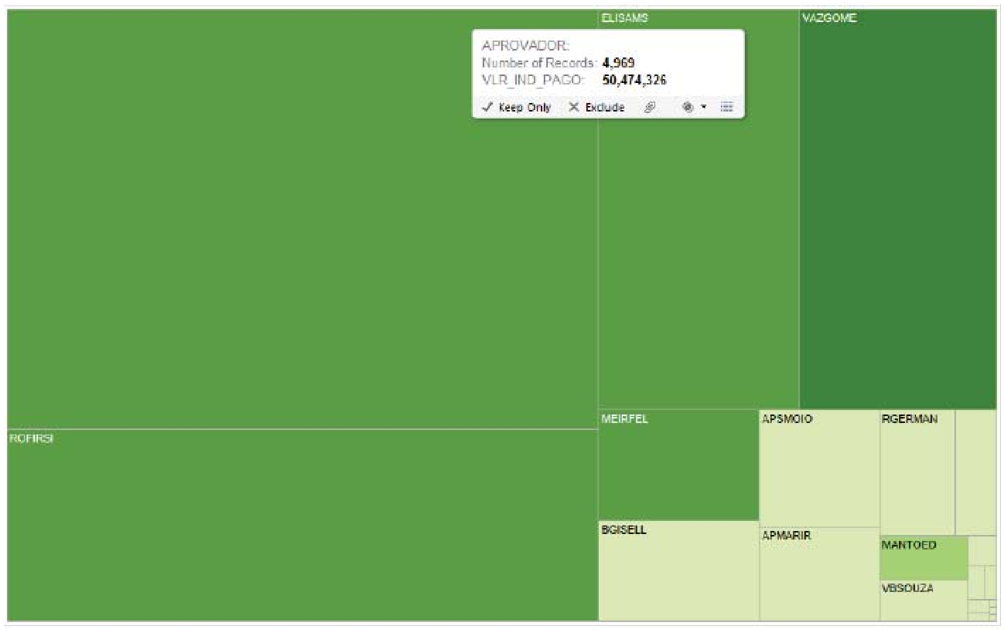
Authorizing Payment
- Checking Authorization limits
- Reviewing Approvers activities.

Payment
- All claims paid or incurred represents valid obligations under the contract in force..
- Claim settlement is reasonable

Maintaining Contracts
- Contracts status are properly updated after approving the payment.
- Contract liability are properly updated after approving the payment.

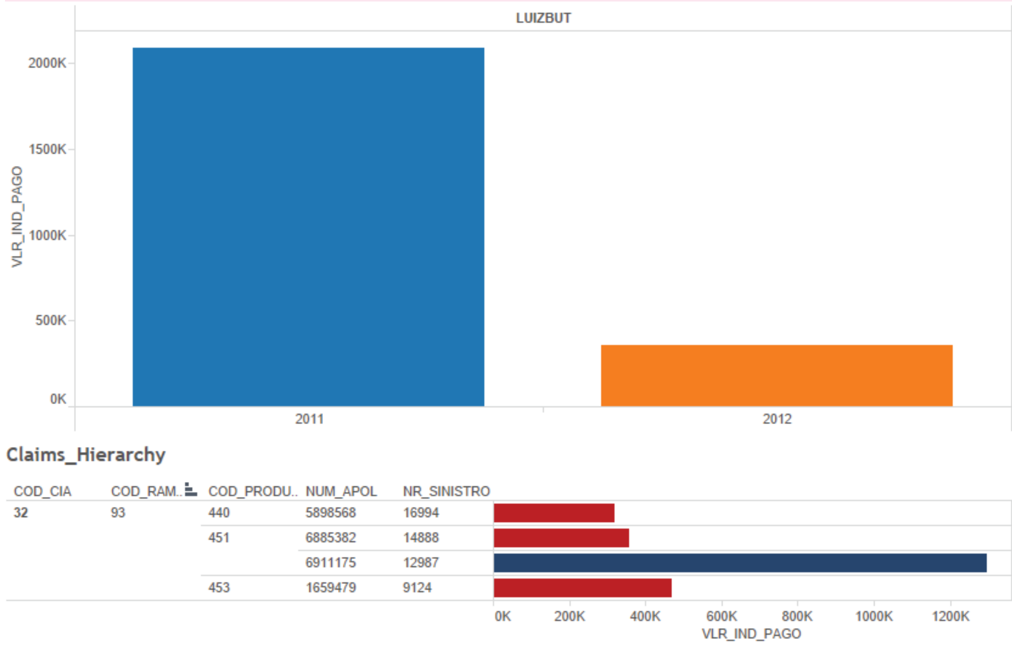
Approvers Activities

A chart showing both the number of transactions authorized by a specific approver (The higher the number of claims, the bigger the size of square) and the total monetary value he authorized as claim payments (the higher the value, the darker the green). An auditor using this graph might be interested in the approver who approved the highest monetary value, or he might be interested in the approver who only authorized one single transaction (smallest square on the lower right corner). The auditor can right click any square to see the actual data.



APROVADOR:
Number of Records: 4,969
VLR_IND_PAGO: 50,474,326

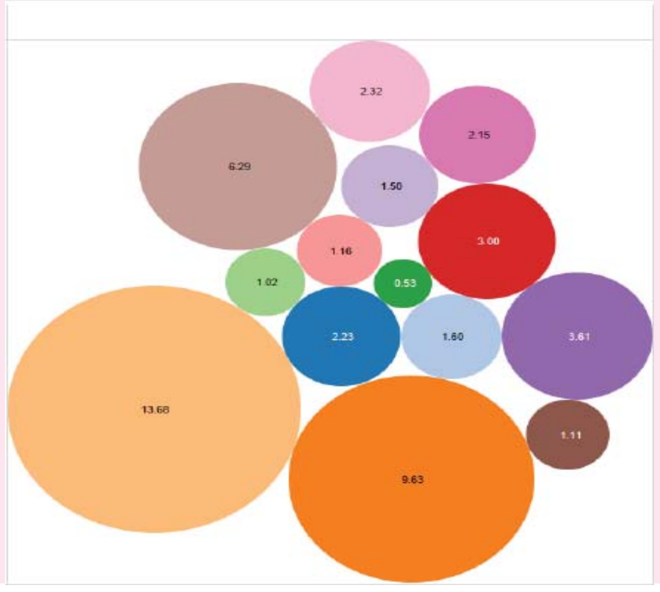
The top chart shows the approvers' total claim payments for each year. The bottom Chart shows the claim Hierarchy. When an auditor select a specific approver's activity from the top chart, the bottom chart will automatically shows only the claims approved by this approver.



Claims_Hierarchy

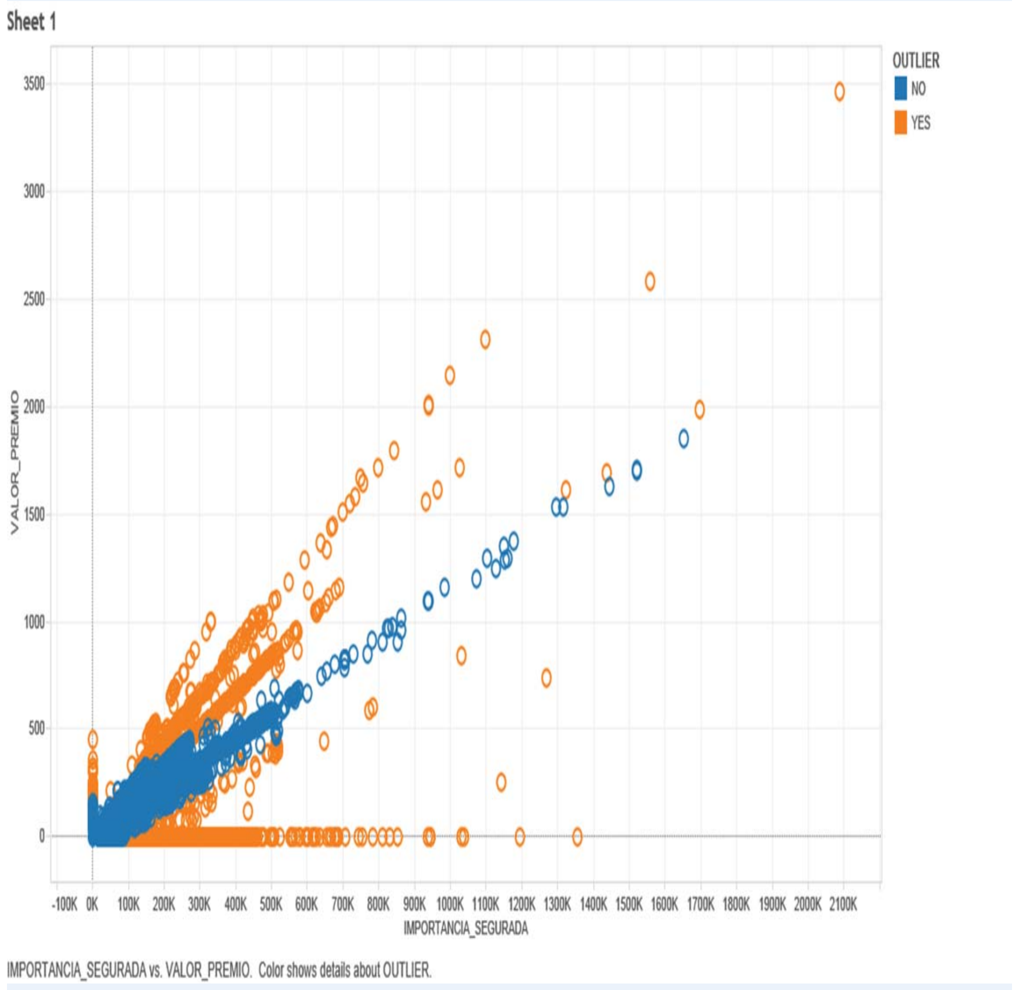
COD_CIA	COD_RAM..	COD_PRODU..	NUM_APOL	NR_SINISTRO
32	93	440	5898568	16994
		451	6885382	14888
			6911175	12987
		453	1659479	9124

Approvers and their average approved interest rates on the payments of the claims.



Premium Outliers

Based on a RobustReg SAS model, the chart shows potential premium outliers in orange. The ones under the blue line is specially important as it shows that the company is collecting less premium than it should.



IMPORTANCIA_SEGURADA vs. VALOR_PREMIO. Color shows details about OUTLIER.

RUTGERS
Rutgers Business School
Newark and New Brunswick

Analytical Procedure and Non-Financial Information: A Case of Multi-location Data

Kyunghee Yoon, Alexander Kogan, Miklos A. Vasarhelyi, and Tim Pearce (KPMG)

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 (macro-level) on APs. Knechel 1988; Dzeng 1994; Allen et al. 1999).

H1: Firm-wide sales expectations developed from disaggregated individual location model produce more accurate and more precise expectation than firm-wide sales expectation derived from aggregated firm-level models.

H2: Firm-wide sales expectations developed from daily disaggregated individual location model produce more accurate and more precise expectation than firm-wide sales expectation derived from weekly disaggregated individual location models.

APs and NFI

SAS No 56 (AICPA 1988) suggests Non-financial information (NFI) should be considered when performing APs, and also it can be used to evaluate risks and detect material misstatements (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.

H3: The model with both financial and non-financial information produces more accurate and more precise prediction than the model with only 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. NFI

Weather information such as daily precipitation and maximum temperature is utilized as non-financial 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 Description	Model Specifications
Panel A: Models Without NFI		
Weekly	Vector Auto-regression	$X_{wt,1} = \alpha + \beta_1 X_{wt-1,1} \cdots$
Daily		$X_{dt,1} = \alpha + \beta_1 X_{dt-1,1} \cdots$
Panel B: Models With NFI in a Firm-Wide Level		
Weekly	Multivariate Regression	$X_{wt,1} = \alpha + \beta_1 A_{t,1}$
Daily		$X_{dt,1} = \alpha + \beta_1 A_{t,1}$
Weekly	Vector Auto-regression	$X_{wt,1} = \alpha + \beta_1 X_{wt-1} + \cdots + \beta_3 A_t$
Daily		$X_{dt,1} = \alpha + \beta_1 X_{dt-1} + \cdots + \beta_3 A_t$
Panel C: Models in Store level Data		
Weekly	Multivariate Regression	$X_{wt,1} = \alpha + \beta_1 X_{wt,2} + \cdots$
Daily		$X_{dt,1} = \alpha + \beta_1 X_{dt,2} + \cdots$
Weekly	Vector Auto-regression	$X_{dt,1} = \alpha + \beta_1 X_{wt,2-1} + \cdots$
Daily		$X_{dt,1} = \alpha + \beta_1 X_{dt,2-1} + \cdots$
Panel D: Models With NFI in Store level Data		
Weekly	Multivariate Regression	$X_{wt,1} = \alpha + \beta_1 X_{wt,2} + \cdots + \beta_3 A_t$
Daily		$X_{dt,1} = \alpha + \beta_1 X_{dt,2} + \cdots + \beta_3 A_t$
Weekly	Vector Auto-regression	$X_{dt,1} = \alpha + \beta_1 X_{wt-1,2} + \cdots + \beta_3 A_t$
Daily		$X_{dt,1} = \alpha + \beta_1 X_{dt-1,2} + \cdots + \beta_3 A_t$

Preliminary Results

Level	Model Description	Adj. R square	MAPE
Panel A: Models Without NFI in Store level Data (Single store test result)			
Daily	Multivariate Regression	0.6910	0.1548
Daily	Vector Auto-regression	0.7490	0.1495
Panel B: Models With NFI in Store level Data (Single store test result)			
Daily	Vector Auto-regression	0.7512	0.1450

As far of empirical works with one of the stores located in Gastonia, North Carolina, the weather information plays important roles in explaining the sales account but doesn't improve the accuracy of expectation significantly.

RUTGERS
Rutgers Business School
Newark and New Brunswick

Log4Audit

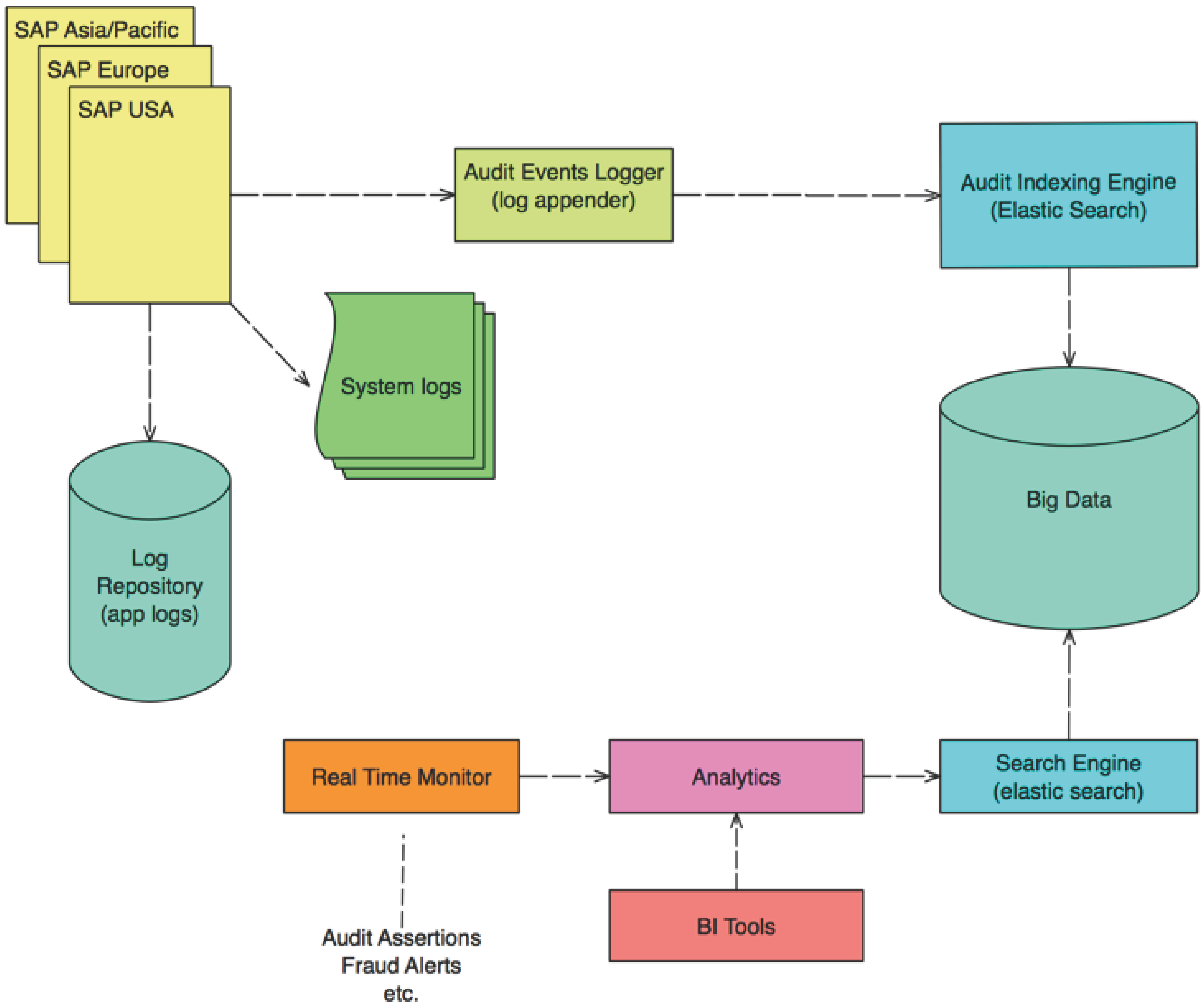
Tatiana Gershberg and Miklos Vasarhelyi

Traditional Audit Evidence

AU Section 326 pertaining Audit Evidence specifies that auditors obtain audit evidence by "testing the accounting records". The testing may include analysis, review, reproduction of "procedures followed in the financial reporting process, and reconciling related types and applications of the same information." Accounting records do not suffice as audit evidence; thus, auditors seek other information to explain how this data was compounded. Knowingly, financial data reported for auditing purposes is consolidated data gathered from various ERP systems within the organization.

Log4Audit

Reaching "conclusions through valid reasoning" by auditors can be supplemented with artificial intelligence providing exact set of events that led to an accounting record being examined. Predictive and, furthermore, preventive audit (Kuenkaikaew and Vasarhelyi, 2013) implementation here is essential. The model also allows for customization of analytics: search engine and Index engine, for example, enable tuning of search for keywords in certain proximity, or adjusting verbosity or severity of logging.


SAP Asia/Pacific
SAP Europe
SAP USA
Audit Events Logger (log appender)
Audit Indexing Engine (Elastic Search)
System logs
Log Repository (app logs)
Big Data
Real Time Monitor
Analytics
Search Engine (elastic search)
BI Tools
Audit Assertions
Fraud Alerts
etc.

Model

- *ERP systems already utilize a logging framework.*
- *Audit Events Logger's main function is to accept messages, accompanied by a date and time stamp, its verbosity and severity.*
- *Audit Indexing Engine serializes the data by indexing it. By tagging the data processed within Audit Indexing Engine (AIE), we speculate that AIE assists with structuring the data and producing higher quality analytics.*
- *Audit Big Data becomes a repository of indexed unstructured data that is accessed by a search engine in order to produce analytics that satisfy the needs of Business Intelligence tools and Real Time Monitoring to generate meaningful output that is further investigated by auditors.*
- *Implementing methodologies that lead to diagnostics, prioritization and evaluating of anomalies would streamline the auditing process cycles, leaving the only exceptional cases for human judgment.*

RUTGERS
Rutgers Business School
Newark and New Brunswick