The Application of Exploratory Data Analysis in Auditing

Dissertation Defense
Qi Liu

Dissertation Committee
Advisor: Dr. Miklos A. Vasarhelyi
Dr. Alex Kogan
Dr. Michael Alles
Dr. Ingrid Fisher

August 11, 2014
Main Outline

- Introduction

- A Conceptual Framework to Apply Exploratory Data Analysis in Audit Practice

- An Application in Operational Audit – Credit Card Retention Case

- An application in Fraud Risk Assessment -- Healthcare Fraud Detection

- Conclusion and Future Research
Introduction

❖ Contribution

➢ Identify the value that EDA can add to auditing
➢ Develop a conceptual framework to guide auditors application of EDA
➢ Illustrate applications to demonstrate how auditors can benefit from the framework

❖ Main Findings

➢ EDA allows auditors to identify the emerging risks and the existing risks that have not been realized before.

➢ By combining EDA with CDA, auditors are capable of creating new audit objectives when they are testing the existing ones, thus continuous improve audit quality and efficiency.

➢ Newly developed EDA and CDA techniques provide more analytical power for auditors to not only provide high-level assurance but also identify new business opportunities or directions for the organization.
Exploratory data analysis (EDA) is a statistical data analysis approach emphasizing on pattern recognition and hypothesis generation (Tukey, 1977).

Confirmatory Data Analysis (CDA) is a widely used data analysis approach emphasizing on experimental design, significance testing, estimation, and prediction (Good, 1983).

### Comparison of EDA and CDA

<table>
<thead>
<tr>
<th>Reasoning Type</th>
<th>Exploratory Data Analysis (EDA)</th>
<th>Confirmatory Data Analysis (CDA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal</td>
<td>Pattern Recognition and Hypothesis generation</td>
<td>Estimation, Modeling, Hypothesis testing</td>
</tr>
<tr>
<td>Applied Data</td>
<td>Observation Data (data collected without well-defined hypothesis)</td>
<td>Experimental data (data collected through formally designed experiments)</td>
</tr>
<tr>
<td>Tools</td>
<td>Descriptive Statistics, Data Visualization</td>
<td>Traditional statistical tools of inference, significance, and confidence</td>
</tr>
</tbody>
</table>
| Advantages     | • No assumptions needed  
                 • Promotes deeper understanding of the data | • Precise  
                 • Well-established theory and methods |
| Disadvantages  | • No conclusive answers,  
                 • Difficult to avoid bias produced by overfitting | • Required unrealistic assumptions,  
                 • Difficult to notice unexpected results |

*Comparison of EDA and CDA*
Overview of Exploratory Data Analysis (2/2)

- Traditional definition of EDA encounters some challenges in the current big data era.

- With the emergence of these data analysis methods the nature of EDA changes. EDA is converging with other methodologies such as data mining. (Yu, 2010)

- Empiricist give a profoundly misguided depiction of EDA, ignoring the role played in discovery by mental models. (De Mast and Kemper, 2009)

<table>
<thead>
<tr>
<th></th>
<th>Traditional EDA</th>
<th>Modern EDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Means-oriented</td>
<td>Goal-oriented</td>
</tr>
<tr>
<td>Scope</td>
<td>Visualized Exploratory Analysis</td>
<td>General Exploratory Analysis</td>
</tr>
<tr>
<td>Tools</td>
<td>Simple Arithmetic, Easy-to-draw Pictures</td>
<td>Descriptive Statistics, Advanced Data Visualization Techniques, Data Mining Techniques</td>
</tr>
<tr>
<td>Key Features</td>
<td>• No assumptions needed, • No conclusive answers</td>
<td>• No assumptions needed, • May provide conclusive answers</td>
</tr>
</tbody>
</table>

Comparison between Traditional EDA and Modern EDA
Exploratory Data Analysis Techniques

- Traditional Exploratory Data Analysis Techniques
  - Descriptive Statistics
  - Data Visualization
  - Data Transformation

- Advanced Exploratory Data Analysis Techniques
  - Advanced Data Visualization
  - Feature Selection
  - Cluster Analysis
  - Association Analysis
  - Process Mining
  - Text Mining
  - Social Network Analysis

Introduction
Motivation and Research Questions

 Motivation
    Currently widely used data analysis approach in auditing is CDA.
    CDA has limited power to identify hidden/emerging risks in organization.
    Systematically apply EDA in auditing can effectively identify and assess emerging and unaware risks.

 Research Questions
  – What value can EDA add to audit process?
  – When should auditors use EDA in audit process?
  – How specifically should auditors perform EDA in auditing?
  – What EDA techniques can be applied in different audit stages?
  – How can EDA be implemented in continuous auditing system?
  – How to demonstrate and evaluate the application of EDA in audit practice?
Current Applications of EDA in Other Disciplines

- Since 1980s, EDA has been applied to diversified disciplines such as interior design, marketing, industry engineering, and geography (Chen et al., 2011; Nayaka and Yano, 2010; Koschat and Sabavala, 1994; Wesley et al., 2006; De Mast and Trip, 2007, 2009).

- A framework to apply EDA in practical problem solving issues include: (1) display the data; (2) identify salient features; (3) interpret salient features (De Mast and Kemper, 2009).
# Current Applications of EDA in Auditing

<table>
<thead>
<tr>
<th>EDA technique</th>
<th>Application areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive Statistics</td>
<td>Fraud Risk Analysis (Sokol et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>Continuous Auditing (Rezaee et al., 2002)</td>
</tr>
<tr>
<td>Data Visualization</td>
<td>Fraud Risk Analysis (Cox et al., 1997; Sokol et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>Continuous Auditing (Rezaee et al., 2002)</td>
</tr>
<tr>
<td>Feature Selection</td>
<td>Fraud Risk Analysis (Brockett et al., 2002)</td>
</tr>
<tr>
<td>Cluster Analysis</td>
<td>Understand auditees’ business (Thiprungsri, 2012)</td>
</tr>
<tr>
<td></td>
<td>Fraud Risk Analysis (Zaslavsky and Strizhak 2006, Quah and Sriganesh 2008, Thiprungsri, 2011)</td>
</tr>
<tr>
<td></td>
<td>Continuous Auditing (Rezaee et al., 2002)</td>
</tr>
<tr>
<td>Text Mining</td>
<td>Fraud Risk Analysis (Holton, 2009)</td>
</tr>
<tr>
<td>Process Mining</td>
<td>Understand auditees’ business (Jens et al., 2013)</td>
</tr>
<tr>
<td></td>
<td>Business Risk Assessment (Jens et al., 2013)</td>
</tr>
<tr>
<td></td>
<td>Internal Control Assessment (Jens et al., 2013)</td>
</tr>
<tr>
<td></td>
<td>Fraud Risk Analysis (Yang and Hwang, 2006; Jens et al., 2009)</td>
</tr>
<tr>
<td>Social Network Analysis</td>
<td>Fraud Risk Analysis (Debreceny and Gray, 2011; Jens et al., 2014)</td>
</tr>
</tbody>
</table>
EDA Application Framework in Auditing

Conceptual Framework

Process
- Display distribution
- Identify salient features
- Generate explanations
- Test explanations
- Identify suspicious cases
- Explore the causes
- Confirm the relationships
- Report results and add new audit objectives

Assess Engagement Risk
- Understand client’s business and assess client’s business risk
- Understand internal control and control risk
- Assess fraud risk
- Perform analytical procedures
- Review subsequent events
- Assess engagement quality

Means
- Descriptive statistics
- Data visualization
- Data transformation
- Feature Selection
- Data Mining
- Process Mining
- Text Mining
- Social Network Analysis

Timing
EDA Application Framework in Auditing -- Timing

- Assess and respond to engagement risk
- Understand client’s business
- Assess client business risk
- Perform preliminary analytical procedures

- Set materiality and assess acceptable audit risk and inherent risk
- Understand internal control and assess control risk
- Gather information to assess fraud risks
- Develop overall audit plan and audit program

- Perform Subsequent events review
- Issue audit report
- Assess engagement quality

- Perform tests of controls
- Substantive tests of transactions
- Perform analytical procedures
- Test of details of balances
## Audit standards related to EDA

<table>
<thead>
<tr>
<th>Standard</th>
<th>Description</th>
</tr>
</thead>
</table>
| **AU-C 240** | - Performing substantive analytical procedures relating to revenue using disaggregated data; for example, computer-assisted audit techniques may be useful in identifying **unusual or unexpected** revenue relationships or transactions.  
- Performing a detailed review of the entity's quarter-end or year-end adjusting entries and investigating any that appear to have **an unusual nature or amount**.  
- Analyzing sales discounts and returns for **unusual patterns or trends**.  
- Reviewing the propriety of **large and unusual expenses**. |
| **AU-C 315** | - Analytical procedures performed as risk assessment procedures may identify aspects of the entity of which the auditor was **unaware**... |
| **AU-C 520** | - The results of analytical procedures designed and performed near the end of the audit may identify a **previously unrecognized** risk of material misstatement. |
| **AU-C 550** | - **...an unusually high turnover** of senior management or professional advisors may suggest unethical or fraudulent business practices that serve the related party's purposes.  
- In evaluating the business rationale of a significant related party transaction... the auditor may consider... Whether the transaction (1) **has unusual terms** of trade, such as unusual prices, interest rates, guarantees, and repayment terms (2) is processed in **an unusual manner**. |
<table>
<thead>
<tr>
<th>Audit Tasks</th>
<th>EDA Technologies</th>
<th>Audit Tasks</th>
<th>EDA Technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assess Engagement Risk</td>
<td>Data Visualization, Text Mining</td>
<td>Assess Fraud Risk</td>
<td>Descriptive Statistics, Data Visualization</td>
</tr>
<tr>
<td>Understand Clients Business and Assess Client’s Business Risk</td>
<td>Descriptive Statistics, Data Visualization, Process Mining, Social Network Analysis, Text Mining</td>
<td>Perform Analytical Procedures</td>
<td>Descriptive Statistics, Data Visualization, Data Transformation, Data Mining, Social Network Analysis, Text Mining, Process Mining</td>
</tr>
<tr>
<td>Understand Internal Control and Assess Control Risk</td>
<td>Process Mining</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review Subsequent Events</td>
<td>Descriptive Statistics, Data Visualization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assess Engagement Quality</td>
<td>Descriptive Statistics, Data Visualization</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
EDA Application Framework in Auditing -- Process

1. Display the distribution of fields related to audit objects
2. Identify salient features from the distribution
3. Generate possible explanations
4. Check whether the explanation can explain all the exceptional cases
   - Yes: End analysis
   - No: Identify suspicious case
     - Explore the causes of abnormal cases
       - Perform CDA or substantive tests to confirm the relationship
       - Add new audit objectives & Report the results
The application of EDA in Continuous Auditing

- Performing EDA in the continuous auditing system design and development phase
- Integrating EDA function in the continuous auditing system as part of the continuous risk monitoring and assessment module.