Using Data Analytics to Detect Fraud

Introduction to Data Analytics and the Data Analysis Process

Jeremy R. Clopton, CFE, CPA, ACDA, CIDA
Director, Upstream Academy
Course Objectives

▪ How data analytics can be used to detect fraud
▪ Different tools to perform data analytics
▪ How to walk through the full data analytics process
▪ How to analyze non-numeric data, such as text and timelines, for signs of fraud
▪ Red flags of fraud that appear in the data
▪ Data analytics tests that can be used to detect fraud
Introduction to Data Analytics

- Data analytics, as it applies to fraud examination, refers to the use of analytics software to identify trends, patterns, anomalies, and exceptions in data.
Introduction to Data Analytics

- Especially useful when fraud is hidden in large data volumes and manual checks are insufficient
- Can be used reactively or proactively
Introduction to Data Analytics

- Effective data analysis requires:
  - Translating knowledge of the organization and common fraud indicators into analytics tests
  - Effectively using technological tools
  - Resolving errors in data output due to incorrect logic or scripts
  - Applying fraud investigation skills to the data analysis results to detect potential instances of fraud
Introduction to Data Analytics

- Data analysis techniques alone are unlikely to detect fraud; human judgment is needed to decipher results.
Types of Data That Can Be Analyzed

- Relevant data comes from numerous sources and takes numerous forms:
  - Accounting and financial data
  - Human resources data
  - Customer data
  - Vendor data
  - Internal communications and documents
  - External benchmarking data
## Types of Data That Can Be Analyzed

<table>
<thead>
<tr>
<th>Structured data</th>
<th>Unstructured data</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Sales records</td>
<td>• Email and instant messages</td>
</tr>
<tr>
<td>• Payment or expense details</td>
<td>• Payment text descriptions</td>
</tr>
<tr>
<td>• Payroll details</td>
<td>• Social media activity</td>
</tr>
<tr>
<td>• Inventory records</td>
<td>• Corporate document repositories</td>
</tr>
<tr>
<td>• Financial reports</td>
<td>• News feeds</td>
</tr>
<tr>
<td>• Found in accounting software, databases, spreadsheets, etc.</td>
<td></td>
</tr>
</tbody>
</table>
Types of Data That Can Be Analyzed

- Big data:
  - Gartner: “High volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision-making, insight discovery, and process optimization”
  - Information of extreme size, diversity, and complexity that eventually tells a story once all the data sets and pieces from different sources are connected
Population Analytics

- Although testing a sample of data is a valid *audit* approach, it is not as effective for *fraud detection* purposes.
- To detect fraud, data analysis techniques must be performed on the full data population.
- Need to define population boundaries, including the amount of historical data to include.
Benefits of Data Analytics

- Increase efficiency and effectiveness.
- Boost productivity and profitability.
- Reduce sampling errors.
- Assess and improve internal controls.
- Revise or reinforce policies.
- Monitor trends.
Benefits of Data Analytics

- Identify fraud before it becomes material.
- Focus detection efforts on suspicious transactions.
- Gain insight into how well internal controls are operating.
- Compare data from diverse sources to identify instances of fraud or noncompliance.
Challenges in Using Data Analytics

- Poorly defined scope
- Data acquisition
  - Manually maintained data
- False positives
- Lack of familiarity
  - Data storage systems
  - Software systems
  - Organizational processes
Challenges in Using Data Analytics

- Data security and integrity concerns
- Privacy and confidentiality concerns
- Evolving business processes and activities
- Learning curve
- Cost of data analytics software
- Evolution of fraud schemes
- Element of concealment
Data Analytics Software

- Commercial data analysis programs
- Spreadsheets and databases
- Customized data analysis programs
- Shareware and freeware data analysis programs
Spectrum of Analysis

- Ad hoc testing
- Repetitive testing
- Continuous testing
The Need for a Formal Process

- The ability of data analysis tests to help detect fraud depends greatly on what is done before and after the actual data analysis techniques are applied.
The Data Analysis Process

- Planning phase
- Preparation phase
- Testing and interpretation phase
- Post-analysis phase
Planning Phase

- Planning is essential and can help avoid:
  - Inefficient data analysis
  - A lack of focus or direction for the engagement
  - Avoidable technical difficulties
  - Overlooking key areas for exploration
Planning: Understand the Data

- Review database schema and technical documentation.
- Consult with the data administrator.
- Learn what fields and records exist.
- Learn what tables house data and how tables are linked together.
Planning: Articulate the Objectives and Scope

Consider:

• The purpose of the engagement
• The structure and size of the business
• The target area of examination and any barriers to retrieving the data
• The resources available for the engagement
• Whether any predication of fraud exists
• Any existing materiality thresholds or preferences
Planning: Build a Profile of Potential Frauds

- Identify:
  - The organization’s risk areas
  - The types of frauds possible in those risk areas
  - The resulting exposure to those frauds
- Refer to previous fraud risk assessments
Planning: Build a Profile of Potential Frauds

General focus of attorneys (opportunity for internal auditors and investigators)

**The Fraud Tree**

**Corruption**
- Conflicts of interest
- Bribery and corruption/FCPA
- Illegal gratuities
- Extortion

**Financial Statement Fraud**
- Revenue recognition
- Asset valuation
- Understated liabilities/expenses
- Improper disclosures

**Asset Misappropriation**
- Theft of cash receipts
- Check tampering schemes
- Billing schemes
- Payroll fraud
- T&E fraud
- Theft of noncash assets

General focus of internal auditors
Preparation Phase

- The results of a data analysis test are only as good as the data used for the analysis.
Preparation: Identify Relevant Data

- Use profile of potential frauds as a guide.
- For each fraud scenario, identify which fields and records would be affected.
Preparation: Identify Relevant Data

- Determine:
  - What specific data is available
  - Who generates and maintains the data
  - Where the data is stored
  - Timing of the data extraction
  - How the examiner will receive and store the data
  - Control totals needed for verification
  - Potential corroborating sources of data
## Common Data Sources: Asset Misappropriation Schemes

<table>
<thead>
<tr>
<th>NAME</th>
<th>FUNCTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vendor master</td>
<td>Lists all approved vendors</td>
</tr>
<tr>
<td>Employee master</td>
<td>Lists all employees</td>
</tr>
<tr>
<td>Accounts payable ledger</td>
<td>Tracks when and to whom payments are due</td>
</tr>
<tr>
<td>Cash disbursements journal</td>
<td>Tracks all cash disbursements</td>
</tr>
<tr>
<td>Purchases journal</td>
<td>Tracks requests for purchases</td>
</tr>
</tbody>
</table>

Depending on the case, certain general ledger accounts can also be selected.
## Common Data Sources: Bribery and Corruption Schemes

<table>
<thead>
<tr>
<th>NAME</th>
<th>FUNCTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vendor master</td>
<td>Lists all approved vendors</td>
</tr>
<tr>
<td>Accounts payable ledger</td>
<td>Tracks when and to whom payments are due</td>
</tr>
<tr>
<td>Cash disbursements journal</td>
<td>Tracks all cash disbursements</td>
</tr>
<tr>
<td>Purchases journal</td>
<td>Tracks requests for purchases</td>
</tr>
<tr>
<td>Selected GL accounts</td>
<td>Identifies accounts where a bribe payment could be hidden</td>
</tr>
<tr>
<td>• Charity/donations</td>
<td></td>
</tr>
<tr>
<td>• Agent payments</td>
<td></td>
</tr>
<tr>
<td>• Marketing expenses</td>
<td></td>
</tr>
<tr>
<td>Travel and entertainment</td>
<td>Item detail of T&amp;E submissions</td>
</tr>
</tbody>
</table>
Common Data Sources: Financial Statement Fraud

<table>
<thead>
<tr>
<th>NAME</th>
<th>FUNCTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales journal</td>
<td>Sales by product, date, customer</td>
</tr>
<tr>
<td>Accounts receivable</td>
<td>Tracks amounts due to company by customer, over time</td>
</tr>
<tr>
<td>Customer master</td>
<td>Lists all customers</td>
</tr>
<tr>
<td>Various sub-ledgers</td>
<td>Might include inventory, capital expenses, outstanding loans, etc.</td>
</tr>
</tbody>
</table>
Preparation: Obtain the Data

- Prepare and submit a formal request for the desired data.
- Receive a file containing data or access to data within the system.
Preparation: Verify the Data

- Ensure data analysis software can open and read the data as provided.
- Validate that data received contains all requested fields and records.
- Confirm control totals.
- Confirm period covered by data.
- Sort the file to test for leading or lagging errors.
Preparation: Verify the Data

- Check for gaps in applicable fields.
- Confirm the format of data in specific fields.
- Check for blank fields where information should be.
- Check for inappropriate duplicate fields or records.
- Test logical relationships in the data.
Preparation: Cleanse and Normalize the Data

- Cleanse and convert data to a format suitable for analysis before executing any tests.
- Normalize the data so that all data can be analyzed consistently.
Preparation: Cleanse and Normalize the Data

- Address inconsistencies in the data:
  - Known errors
  - Blanks or missing data
  - Duplicated data
  - Special or unreadable characters in the data
  - Other unusable entries
Testing and Interpretation Phase

- Data is now ready to be analyzed to uncover patterns consistent with specific fraud scenarios previously identified.
Testing and Interpretation: Organize the Data

- Organize the data in a way designed to uncover patterns consistent with the specific fraud scenarios identified during the planning stage.
- Group the data into homogenous groups to facilitate the ability to spot outliers:
  - Geographical location
  - Business division or unit
  - Time period
  - Salesperson
Testing and Interpretation: Analyze the Data

- High-level tests versus targeted tests
- The role of concealment
- Addressing false positives
Post-Analysis Phase

- Respond to analysis findings.
- Monitor the data.