Wiley Corporate F&A

FRAUD DATA ANALYTICS METHODOLOGY

The Fraud Scenario Approach to Uncovering Fraud in Core Business Systems

LEONARD W. VONA

WILEY

Fraud Data Analytics Methodology

The Fraud Scenario Approach to Uncovering Fraud in Core Business Systems

LEONARD W. VONA

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:
Names: Vona, Leonard W., 1955- author.
Title: Fraud data analytics methodology : the fraud scenario approach to uncovering fraud in core business systems / Leonard W. Vona.
Description: Hoboken, New Jersey : John Wiley & Sons, [2017] | Includes index.
Identifiers: LCCN 2016036161 | ISBN 9781119186793 (cloth) | ISBN 9781119270348 (ePDF) | ISBN 9781119270355 (epub)
Subjects: LCSH: Auditing. | Forensic accounting. | Fraud—Prevention. | Auditing, Internal.
Classification: LCC HF5667 .V659 2017 | DDC 658.4/73—dc23
LC record available at https://lccn.loc.gov/2016036161

Cover design: Wiley
Cover image: © kentoh/Shutterstock

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

This book is dedicated to my family, Patricia, Amy, David, and Jeffrey, for supporting me in my quest to explain fraud auditing. In the memory of my dad, who told me to go to college, and the memory of the women who shaped my life.

Preface

Even the world's best auditor using the world's best audit program cannot detect fraud unless their sample includes a fraudulent transaction. That is why fraud data analytics is so essential to the auditing profession.

Fraud auditing is a methodology tool used to respond to the risk of fraud in core business systems. The methodology must start with the fraud risk identification. Fraud data analytics is about searching for a fraud scenario versus a data anomaly. I have often referred to fraud data analytics as code breaking. The fraud auditor is studying millions of transactions in the attempt to find the needle in the haystack, called the *fraud scenario*. It is my hope that my years of professional experience in using fraud data analytics will move the auditing profession to become the number-one reason for fraud detection.

This book is about the science of fraud data analytics. It is a systematic study of fraud scenarios and their relationship to data. Like all scientific principles, the continual study of the science and the practical application of the science are both necessary for success in the discovery of fraud scenarios that are hiding in all core business systems.

The methodology described in the book is intended to provide a step-by-step process for building the fraud data analytics plan for your company. The first five chapters explain each phase of the process. Later chapters illustrate how to implement the methodology in asset misappropriation schemes, corruption schemes, and financial reporting schemes.

The practitioner will learn that fraud data analytics is both a science and an art. In baseball, there is a science to hitting a baseball. The mechanics of swinging a bat is taught to players of all ages. However, you can read all the books in the world about swinging a bat, but unless you actually stand in the batting box and swing the bat, you will never truly learn the art of hitting a baseball. Likewise, the fraud auditor needs to learn to analyze data and to employ the tools to do so in order to be able to find fraud scenarios hiding in your data systems.

Acknowledgments

To my friends at Audimation Services: Carolyn Newman, Jill Davies, and Carol Ursell. It is because of working with you that I developed the art of fraud data analytics.

To Sheck Cho (Executive Editor), who encouraged me to write my books, and to the editors at Wiley, without you I could not have written this book.

To Nicki Hindes, who keeps my office going while I travel the world.

To all those people who have inspired me. Thank you!

Contents

CHAPTER ONE

Introduction to Fraud Data Analytics

The world's best auditor using the world's best audit program cannot detect fraud unless their sample includes a fraudulent transaction. This is why fraud data analytics (FDA) is so critical to the auditing profession.

How we use fraud data analytics largely depends on the purpose of the audit project. If the fraud data analytics is used in a whistle blower allegation, then the fraud data analytics plan is designed to refute or corroborate the allegation. If the fraud data analytics plan is used in a control audit, then the fraud data analytics would search for internal control compliance or internal control avoidance. If the fraud data analytics is used for fraud testing, then the fraud data analytics is used to search for a specific fraud scenario that is hidden in your database. This book is written for fraud auditors who want to integrate fraud testing into their audit program. The concepts are the same for fraud investigation and internal control avoidance—what changes is the scope and context of the audit project.

Interestingly, two of the most common questions heard in the profession are, "Which fraud data analytic routines should I use in my audit?" and, "What are the three fraud data analytics tests I should use in payroll or disbursements?" In one sense, there really is no way to answer these questions

because they assume the fraud auditor knows what fraud scenario someone might be committing. In reality, we search for patterns commonly associated with a fraud scenario or we search for all the logical fraud scenario permutations associated with the applicable business system. In truth, real fraud data analytics is exhausting work.

I have always referred to fraud data analytics as code breaking. It is the auditor's job to search the database using a comprehensive approach consistent with the audit scope. So, the common question of which fraud data analytics routines should I use can only be answered when you have defined your audit objective and audit scope. A key element of the book is the concept that while the fraud auditor might not know what fraud scenario a perpetrator is committing, the fraud auditor can identify and search for all the fraud scenario permutations. Therefore, the perpetrator will not escape the long arm of the fraud data analytics plan.

Once again, the question arises as to which fraud data analytic routines I should use in my next audit. Using the fraud risk assessment approach, the fraud data analytics plan could focus on those fraud risks with a high residual rating. The auditor could select those fraud risks that are often associated with the particular industry or with fraud scenarios previously uncovered within the organization—or the auditor might simply limit the scope to three fraud scenarios. Within this text, we plan to explain the methodology for building your fraud data analytics plan; readers will need to determine how comprehensive to make their plan.

WHAT IS FRAUD DATA ANALYTICS?

Fraud data analytics is the process of using data mining to analyze data for red flags that correlate to a specific fraud scenario. The process starts with a fraud data analytics plan and concludes with the audit examination of documents, internal controls, and interviews to determine if the transaction has red flags of a specific fraud scenario or if the transaction simply contains data errors.

Fraud data analytics is not about identifying fraud but rather, identifying red flags in transactions that require an auditor to examine and formulate a decision. The distinction between identifying transactions and examining the transaction is important to understand. Fraud data analytics is about creating a sample; the audit program is about gathering evidence to support a conclusion regarding the transaction. The final questions in the fraud audit process:

Is there credible evidence that a fraud scenario is occurring? Should we perform an investigation?

It is critical to understand that fraud data analytics is driven by the fraud scenario versus the mining of data errors. Based on the scenario, it might be one red flag or a combination of red flags. Yes, some red flags are so overpowering that the likelihood of fraud is higher. Yes, some red flags simply correlate to errors. The process still needs the auditor to examine the documents and formulate a conclusion regarding the need for a fraud investigation. It is important to understand the end product of data analytics is a sample of transactions that have a higher probability of containing one fraudulent transaction versus a random sample of transactions used to test control effectiveness. One could argue that fraud data analytics has an element of Las Vegas. Gamblers try to improve their odds of winning. Auditors try to improve their odds of detecting fraud. Figure 1.1 illustrates the concept of improving your odds by reducing the size of the population for sample selection.

Within most literature, a vendor with no street address is a red flag fraud. But a red flag of what? Is a blank street address field indicative of a shell company? How many vendors have no address in the accounts payable file because all payments are EFT? If a vendor receives payment through the EFT process, then is the absence of a street address in your database a red flag? Should a street address be considered a red flag of a shell company? Is the street address linked to a mailbox service company? What are the indicators of a mailbox service company? Do real companies use mailbox service companies? Fraud examiners understand that locating and identifying fraudulent transactions is a matter of sorting out all these questions. A properly developed fraud data-mining plan is the tool for sorting out the locating question.

To start your journey of building your fraud data analytics plan, we will need to explain a few concepts that will be used through the book.

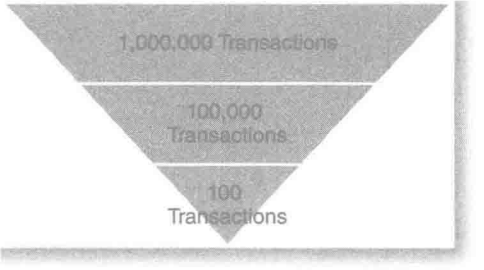

FIGURE 1.1 Improving Your Odds of Selecting One Fraudulent Transaction

What Is Fraud Auditing?

Fraud auditing is a methodology to respond to the risk of fraud in core business systems. It is a combination of risk assessment, data mining, and audit procedures designed to locate and identify fraud scenarios. It is based on the theory of fraud that recognizes that fraud is committed with intent to conceal the truth. It incorporates into the audit process the concept of red flags linked to the fraud scenario concealment strategy associated with data, documents, internal controls, and behavior.

It may be integrated into audit of internal controls or the entire audit may focus on detecting fraud. It may also be performed because of an allegation or the desire to detect fraudulent activity in core business systems. For our discussion purposes, this book will focus on the detection of fraud when there is no specific allegation of fraud.

Fraud auditing is the application of audit procedures designed to increase the chances of detecting fraud in core business systems. The four steps of the fraud audit process are:

1. *Fraud risk identification.* The process starts with identifying the inherent fraud schemes and customizing the inherent fraud scheme into a fraud scenario. Fraud scenarios in this context will be discussed in Chapter 2.
2. *Fraud risk assessment.* In the traditional audit methodology the fraud risk assessment is the process of linking of internal controls to the fraud scenario to determine the extent of residual risk. In this book, fraud data analytics is used as an assessment tool through the use of data-mining search routines to determine if transactions exist that are consistent with the fraud scenario data profile.
3. *Fraud audit procedure.* The audit procedure focuses on gathering audit evidence that is outside the point of the fraud opportunity (person committing the fraud scenario). The general standard is to gather evidence that is externally created and externally stored from the fraud opportunity point.
4. *Fraud conclusion.* The conclusion is an either/or outcome, either requiring the transaction to be referred to investigation or leading to the determination that no relevant red flags exist. Chapters 6 through 15 contain relevant discussion of fraud data analytics in the core business systems.

What Is a Fraud Scenario?

A fraud scenario is a statement as to how an inherent scheme will occur in a business system. The concept of an inherent fraud scheme and the fraud risk

structure is discussed in Chapter 2. A properly written fraud scenario becomes the basis for developing the fraud data analytics plan for each fraud scenario within the audit scope. Each fraud scenario needs to identify the person committing the scenario, type of entity, and the fraudulent action to develop a fraud data analytics plan. The auditing standards also suggest identifying the impact the fraud scenario has on the company.

While all fraud scenarios have the same components, we can group the fraud scenarios into five categories. The groupings are important to help develop our audit scope. The groupings also create context for the fraud scenario. Is the fraud scenario common to all businesses or is the fraud scenario unique to our industry or our company? There are five categories of fraud scenarios:

1. *The common fraud scenario.* Every business system has the same listing of common fraud scenarios. I do not need to understand your business process, conduct interviews of management, or prepare a flow chart to identify the common fraud scenarios.
2. *The company-specific fraud scenario.* The company-specific fraud scenario in a business cycle because of business practices, design of a business system, and control environment issues. I do need to understand your business process, conduct interviews of management, or prepare a flow chart to identify the common fraud scenarios.
3. *The industry-specific fraud scenario.* The industry specific fraud scenarios are similar to the common fraud scenario, except the fraud scenario only relates to an industry. To illustrate the concept, mortgage fraud is an issue for the banking industry. This category of fraud scenarios requires the fraud auditor to be knowledgeable regarding their industry. However, using the methodology in Chapter 2, a nonindustry person could create a credible list of fraud scenarios.
4. *The unauthorized fraud scenario.* The unauthorized fraud scenario occurs when an individual, either internal or external to the company, commits an act by overriding company access procedures.
5. *The internal control inhibitor fraud scenario.* The concept of internal control inhibitor is to identify those acts or practices that inhibit the internal control procedures from operating as designed by management. The common internal control inhibitors are collusion and management override.

Chapter 2 will explain the concept of the fraud risk structure and how to write a fraud scenario that drives the entire fraud audit program. Chapter 2

will also cover the concept of fraud nomenclature. In the professional literature, we use various fraud words interchangeably, which I believe creates confusion within the profession. Words like *fraud risk statement*, *fraud risk*, and *inherent fraud schemes*, *fraud scenario*, *fraud schemes*, and *inherent fraud risk* are used to describe how fraud occurs for the purpose of building a fraud risk assessment or fraud audit program. Within this book, I will use the phrase *fraud scenario* as the words that drive our fraud data analytic plan.

What Is Fraud Concealment?

Fraud concealment is the general or specific conditions that hide the true nature of a fraudulent transaction. A general condition is the sheer size of database, whereas a specific condition is something that the perpetrator does knowingly or unknowingly to cause the business transaction to be processed in the business system and hide the true nature of the business transaction.

To illustrate the concept, all vendors need an address or a bank account to receive payment. On a simple basis, the perpetrator uses his or her home address in the master file. On a more sophisticated level, the perpetrator uses an address for which the linkage to the perpetrator is not visible within the data—for example, a post office box in a city, state, or country that is different from where the perpetrator resides. The fraud data analytics plan must be calibrated to the level of fraud sophistication that correlates to the specific condition of the person committing the fraud scenario. In Chapter 3, the sophistication model will describe the concepts of low, medium, and high fraud concealment strategies. The calibration concept of low, medium, and high defines whether the fraud scenario can be detected through the master file or the transaction file. It also is a key concept of defining the audit scope.

It is important to distinguish between a fraud scenario and the associated concealment strategies. Simply stated, the fraud scenario is the fraudulent act and concealment is how the fraudulent act is hidden. From an investigation process, concealment is referred to as the intent factor. From a fraud audit process, the concealment is referred to as the fraud concealment sophistication factor.

What Is a Red Flag?

A red flag is an observable condition within the audit process that links to the concealment strategy that is associated with a specific fraud scenario. A red flag exists in data, documents, internal controls, behavior, and public records.

Fraud data analytics is the search for red flags that exist in data that links to documents, public records, persons, and eventually to a fraud scenario.

The red flag is the inverse of the concealment strategy. The concealment strategy is associated with the person committing the fraud scenario and the red flag is how the fraud auditor observes the fraud scenario.

The red flag theory becomes the basis of developing the fraud data profile, which is the starting point of developing the fraud data analytics plan. The red flags directly link to the fraud concealment strategy. The guidelines for using the red flag theory are discussed in Chapter 3.

What Is a False Positive?

A false positive is a transaction that matches the red flags identified in the fraud data profile but the transaction is not a fraudulent transaction. It is neither bad nor good. It simply is what it is. What is important is that the fraud data analytics plan has identified a strategy for addressing false positives. Fundamentally, the plan has two strategies: Attempt to reduce the number of false positives through the fraud data analytics plan or allow the fraud auditor to resolve the false positive through audit procedure. There may be no correct answer to the question; however, ignoring the question is a major mistake in building your plan.

What Is a False Negative?

A false negative is a transaction that does not match the red flags in the fraud data profile but the transaction is a fraudulent transaction. From a fraud data analytics perspective, false negatives occur due to not understanding the sophistication of concealment as it related to building your fraud data analytics plan. Other common reasons for a false negative are: data integrity issues, poorly designed data interrogation procedures, the lack of data, and the list goes on.

While false positives create unnecessary audit work for the fraud auditor, false negatives are the real critical issue facing the audit profession because the fraud scenario was not detected.

The false positive conundrum: Refine the fraud data analytics or resolve the false positive through audit work.

There is no real correct answer to the question. The fraud data analytics should attempt to provide the fraud auditor with transactions that have a higher probability of a person committing a fraud scenario. The fraud data interrogation routines should be designed to find a specific fraud scenario. That is the purpose of fraud data analytics. However, by the nature of data and fraud, false positives will occur. Deal with it. The real question is how to minimize the number of false positives consistent with the fraud data analytics strategy selected for the fraud audit.

Remember, fraud data analytics is designed to identify transactions that are consistent with a fraud data profile that links to a specific fraud scenario. There needs to be a methodology in designing the data interrogation routines. The methodology needs to be based on a set of rules and an understanding of the impact the strategy will have on the number of false positives and the success of fraud scenario identification.

The reality of fraud data analytics is the process will have false positives; said another way, there are transactions that will have all the attributes of a fraud scenario, but turn out to be valid business transactions. That is the reality of the red flag theory. Unfortunately, the reality of fraud data analytics is that there will also be false negatives based on the strategy selected. This is why before the data interrogation process starts, there must be a defined plan that documents the auditor judgment. Senior audit management must understand what the plan is designed to accomplish and why the plan is designed to fail. Yes, based on the correlation of audit strategy and sophistication of fraud concealment, you can design a plan to fail to detect a fraud scenario. At this point in the book, do not read this as a bad or good; Chapter 3 will explain how to calibrate your data interrogation routines consistent with the sophistication of concealment.

To provide a real-life example, in one project involving a large vendor database, our fraud data analytics identified 200 vendors meeting the profile of a shell company. At the conclusion, we referred five vendors for fraud investigation. In one sense, the project was a success; in another sense, we had 195 false positives.

If I could provide one suggestion based on my personal experience, the person using the software and the fraud auditor need to be in the same room at the same time. As reports are created, someone needs to look at the report and refine the report based on the reality of the data in your database. Fraud data analytics is a defined process and with a set of rules. However, the process is not like the equation $1 + 1 = 2$. It is an evolving process of inclusion and exclusion based on a methodology and fraud audit experience. So, do not worry about the

false positive, which simply creates unnecessary audit work. Worry about the false negative.

FRAUD DATA ANALYTICS METHODOLOGY

I commonly hear auditors talk about the need to play with the data. This is one approach to fraud detection. The problem with the approach is that it relies on the experience of the auditor rather than on a defined methodology. I am not discounting audit experience, I would suggest that auditor experience is enhanced with a methodology designed to search for fraud scenarios. In fact, the data interpretation strategy explained in Chapter 3 is a combination of professional experience and methodology.

The fraud data analytics methodology is a circular approach to analyzing data to select transactions for audit examination (Figure 1.2).

- *Fraud scenario.* The starting point for building a fraud data analytics plan is to understand how the fraud risk structure links to the audit scope. The process of identifying the fraud scenarios within the fraud risk structure and how to write the fraud scenario is discussed in Chapter 2.
- *Strategy.* The strategy used to write data interrogation routines needs to be linked to the level of sophistication of concealment. For purposes of this book there are four general strategies, which are explained in Chapter 3.

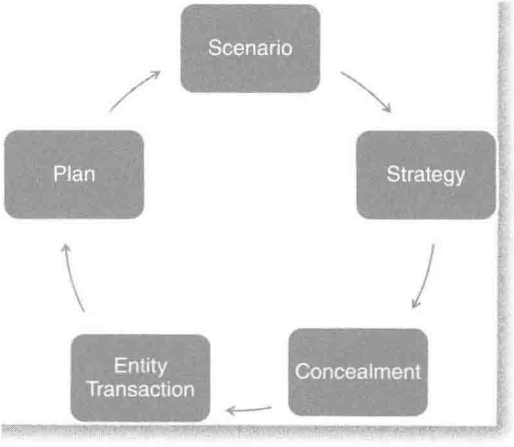

FIGURE 1.2 Circular View of Data Profile