

# **Audit Analytics in the Financial Industry**

*This page intentionally left blank*

# **RUTGERS STUDIES IN ACCOUNTING ANALYTICS**

## **Audit Analytics in the Financial Industry**

**BY**

**JUN DAI**

*Southwestern University of Finance and Economics, China*

**MIKLOS A. VASARHELYI**

*Rutgers University, USA*

**and**

**ANN F. MEDINETS**

*Rutgers University, USA*



United Kingdom – North America – Japan – India – Malaysia – China

Emerald Publishing Limited  
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2019

Copyright © 2019 Emerald Publishing Limited

**Reprints and permissions service**

Contact: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)

No part of this book may be reproduced, stored in a retrieval system, transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without either the prior written permission of the publisher or a licence permitting restricted copying issued in the UK by The Copyright Licensing Agency and in the USA by The Copyright Clearance Center. Any opinions expressed in the chapters are those of the authors. Whilst Emerald makes every effort to ensure the quality and accuracy of its content, Emerald makes no representation implied or otherwise, as to the chapters' suitability and application and disclaims any warranties, express or implied, to their use.

**British Library Cataloguing in Publication Data**

A catalogue record for this book is available from the British Library

ISBN: 978-1-78743-086-0 (Print)

ISBN: 978-1-78743-085-3 (Online)

ISBN: 978-1-78743-173-7 (Epub)



ISOQAR certified  
Management System,  
awarded to Emerald  
for adherence to  
Environmental  
standard  
ISO 14001:2004.

Certificate Number 1985  
ISO 14001



INVESTOR IN PEOPLE

# Contents

Introduction: What is Audit Analytics? <i>Jun Dai and Miklos Vasarhelyi</i>	vii
<b>Part I: Audit Analytics Procedures</b>	
<b>Chapter 1 An Application of Exploratory Data Analysis in Auditing – Credit Card Retention Case</b> <i>Qi Liu</i>	3
<b>Chapter 2 Audit Analytics: A Field Study of Credit Card After-sale Service Problem Detection at a Major Bank</b> <i>Jun Dai, Paul Byrnes, Qi Liu and Miklos Vasarhelyi</i>	17
<b>Part II: Analytics in Credit Card Audits</b>	
<b>Chapter 3 Automated Clustering: From Concept to Reality</b> <i>Paul Byrnes</i>	37
<b>Chapter 4 A Multi-faceted Outlier Detection Scheme for Use in Clustering</b> <i>Paul Byrnes</i>	47
<b>Chapter 5 Are Customers Offered Appropriate Discounts? An Exploratory Study of Using Clustering Techniques in Internal Auditing</b> <i>Jun Dai, Paul Byrnes and Miklos Vasarhelyi</i>	59
<b>Chapter 6 Predicting Credit Card Delinquency: An Application of the Decision Tree Technique</b> <i>Ting Sun and Miklos Vasarhelyi</i>	71

**Part III: Analytics in Insurance Audits**

<b>Chapter 7 Cluster Analysis for Anomaly Detection in Accounting</b> <i>Sutapat Thiprungsri</i>	87
-----------------------------------------------------------------------------------------------------	----

<b>Chapter 8 Multi-dimensional Approaches to Anomaly Detection: A Study of Insurance Claims</b> <i>Basma Moharram</i>	111
--------------------------------------------------------------------------------------------------------------------------	-----

**Part IV: Audit Analytics in Transitory Systems**

<b>Chapter 9 Development of an Anomaly Detection Model for a Bank's Transitory Account System</b> <i>Yongbum Kim</i>	147
-------------------------------------------------------------------------------------------------------------------------	-----

<b>Chapter 10 Development of an Anomaly Detection Model for an Insurance Company's Wire Transfer System</b> <i>Yongbum Kim</i>	165
-----------------------------------------------------------------------------------------------------------------------------------	-----

**Part V: Audit Analytics for Lawsuit Risk Detection**

<b>Chapter 11 A Legal Risk Prediction Model for Credit Cards</b> <i>Feiqi Huang, Qi Liu and Miklos Vasarhelyi</i>	203
----------------------------------------------------------------------------------------------------------------------	-----

**Part VI: Audit Analytics in the Payment Process**

<b>Chapter 12 Analyzing Payment Data and Its Process: A Bank Case</b> <i>Karina Chandia and Miklos Vasarhelyi</i>	217
----------------------------------------------------------------------------------------------------------------------	-----

# Introduction: What is Audit Analytics?

*Jun Dai and Miklos Vasarhelyi*

The spate of accounting scandals and corporate failures since 2001 has brought unprecedented attention to the importance of corporate governance. The Enron scandal, revealed in October 2001, resulted in a loss of about \$80 billion in market capitalization for investors ([The Washington Post, 2002](#)), and a year later, an audit team unearthed \$3.8 billion in fraud at WorldCom ([Pulliam & Solomon, 2002](#)). Since then, both professional auditors and audit researchers have devoted significant effort to improving the capabilities of auditing, internal control, and continuous monitoring ([Alles, Brennan, Kogan, & Vasarhelyi, 2006](#); [Byrnes, 2015](#); [Chan & Vasarhelyi, 2011](#); [Jans, Alles, & Vasarhelyi, 2014](#); [Vasarhelyi, Alles, & Williams, 2010](#)).

“Big data” is receiving increased attention from accounting practitioners. Organizations have collected more data in 2 years than in the previous 2,000 years ([Syed, Gillela, & Venugopal, 2013](#)). For example, Walmart collects more than 1 million customer transactions every hour, and Facebook collects more than 200 gigabytes of data per night ([Cao, Chychyla, & Stewart, 2015](#)). In addition to data stored in traditional accounting systems, auditors are also able to acquire evidence from vast amounts of other complex data, such as non-financial data extracted from modern enterprise resource planning (ERP) systems or online databases, radio frequency identification trackers and networked sensors, social media, and even closed-circuit television videos in stores ([Moffitt & Vasarhelyi, 2013](#)). In addition, many countries now permit some of their government administrative information and data collected from their citizens and businesses to be open to the public, which provides auditors with even more data for monitoring and investigations ([Dai & Li, 2016](#); [O’Leary 2015](#); [Schneider, Dai, Janvrin, Ajayi, & Raschke, 2015](#)).

To extract and process data from a variety of sources to be used for identifying risks, collecting evidence, and ultimately supporting decisions, auditors are utilizing the emerging technology of audit analytics (AA). AA is defined as a science of

discovering and analyzing patterns, identifying anomalies, and extracting other useful information in data underlying or related to the subject matter of an audit through analysis, modeling, and visualization for the purpose of planning or performing the audit. ([AICPA, 2015](#))

The predecessor of AA is the analytical procedure, which has long been used as one of external auditors’ techniques in the planning, substantive testing, and

completion phases of audits (AICPA, 2015). Since analytical procedures performed in the planning phase typically “use data aggregated at a high level” (AICPA, 2012), “the results of those analytical procedures provide only a broad initial indication about whether a material misstatement may exist” (AICPA, 2012). AA techniques can be applied to transaction-level data because such techniques generally maintain good performance even when used on huge and high-dimensionality data sets. As a result, AA can enhance the accuracy of risk assessment and improve the quality of planning.

Traditional analytical procedures usually rely heavily on sampling of audit-related data (AICPA, 2015). However, as large-scale ERP systems are rapidly growing in popularity among businesses, sufficient evidence can no longer be collected from only a sample of data. AA increases the tested population from limited samples (judgmental or statistical) to millions of transactions in full population testing, which enlarges the audit coverage from a small percent of overall transactions to the entire population (AICPA, 2015). Besides data recorded by a client firm’s ERP system, auditors also have access to public data, such as social media postings (Moon, 2016), open government data (Dai & Li, 2016; Kozłowski, 2016), and weather data (Yoon, 2016). Emerging data analytics technologies have the capability to explore vast amounts of data in various structures and formats, which cannot be handled by traditional analytical procedures.

AA offers several advantages over traditional approaches. First, audit data analytics are more cost-effective in terms of evidence collection. On average, AA costs \$0.01 compared to \$4 for a standard audit of the same evidence.<sup>1</sup> In addition, many data analytics techniques are scalable in that they can generally maintain good performance when handling huge and high-dimensionality data-sets (Alpaydin, 2010). Some AA techniques also have the ability to identify data patterns in an unsupervised learning paradigm in which the training data sets for building detection models do not need to contain class label information (Byrnes, 2015; Thiprungsri & Vasarhelyi, 2011).

Part One of this book presents two articles illustrating the process of applying AA to solving audit problems. Part Two contains four studies that use various AA techniques to discover fraud risks and potential frauds in the credit card sector. Part Three focuses on the insurance sector and uses two articles to show the application of clustering techniques in auditing. Part Four includes two chapters on how to employ AA in the transitory system for fraud/anomaly detection. Parts Five and Six illustrate the use of AA to assess risks in the lawsuit and payment processes.

Auditing researchers have been devoting significant efforts to integrating AA techniques into existing audit programs. AA can facilitate various stages of the audit process with simple or complex tests. Chapter 1 summarizes exploratory data analysis (EDA) techniques and the audit stages in which they could be employed for both internal and external audits. This research also conceptualizes the process of implementing EDA in audit procedures. Similarly, Chapter

---

<sup>1</sup><http://raw.rutgers.edu/node/89.html>



2 provides guidance for auditors to apply these new technologies in actual audit work.

A variety of AA technologies can be employed to facilitate risk discovery, anomaly identification, and fraud detection. Chapters 3–5 explore the use of clustering methodologies to identify risky customer groups for a bank's credit card department. After grouping customers with similar characteristics and purchase/payment behaviors into clusters, the bank can manage each group differently and take actions for high-risk credit card holders.

Similar approaches are also employed to identify abnormal life insurance claims. Chapter 7 uses a simple *K*-means clustering model to group claims with similar characteristics and to flag unusually small clusters for further investigation. Chapter 8 explores the attributes to be used to identify outliers, and then uses clustering to assess whether life/disability insurance claim settlements are reasonable and whether the claims themselves are legitimate.

Decision tree is an AA technique that is easy to understand and can facilitate risk and error identification effectively by learning the characteristics and behavior patterns in the data. Chapter 6 shows the potential of Decision Trees for helping internal auditors to identify credit card delinquency, and Chapter 11 applies the Decision Tree methodology to the risk of lawsuits for credit card customers.

Fraud detection is another domain that can benefit from AA techniques. By analyzing transaction-level data, AA can capture unusual data flows and abnormal patterns. Chapters 9 and 10 illustrate how rule-based systems can facilitate fraud detection by incorporating expert knowledge into models. Chapter 9 illustrates the development and testing of a model to detect anomalous transactions in a bank's transitory accounts. Chapter 10 detects fraudulent transactions in the payment process for wire transfers by identifying potential fraud indicators, each of which is assigned a risk score based on perceived severity. Payments with total scores that exceed a threshold would be considered potentially fraudulent transactions that can be recommended for further investigation. Internal control is another important and complex area that could benefit from AA. In Chapter 12, two methods are presented. One of them is fuzzy logic which is used to create a generic risk model for assessing internal controls over payments and the other is the use of statistical tools to detect outliers and anomalies on the data.

The goal of this book is to provide insights for academics, auditors, and business professionals on potential applications of AA in the financial industry. Real-life data and audit problems are used to demonstrate how AA can facilitate the discovery of audit concerns that would be difficult or time consuming if traditional approaches were used.

## References

- AICPA. (2012). Understanding the entity and its environment and assessing the risks of material misstatement, AU-C Section 315. Retrieved from [www.aicpa.org/Research/Standards/AuditAttest/DownloadableDocuments/AU-C-00315.pdf](http://www.aicpa.org/Research/Standards/AuditAttest/DownloadableDocuments/AU-C-00315.pdf)

- AICPA. (2015). Audit analytics and continuous audit: Looking toward the future. Retrieved from [https://www.aicpa.org/interestareas/frc/assuranceadvisoryservices/downloadabledocuments/auditanalytics\\_lookingtowardfuture.pdf](https://www.aicpa.org/interestareas/frc/assuranceadvisoryservices/downloadabledocuments/auditanalytics_lookingtowardfuture.pdf)
- Alles, M. G., Brennan, G., Kogan, A., & Vasarhelyi, M. A. (2006). Continuous monitoring of business process controls: A pilot implementation of a continuous auditing system at Siemens. *International Journal of Accounting Information Systems*, 7(2), 137–161.
- Alpaydin, E. (2010). *Introduction to machine learning* (2nd ed.). Boston, MA: MIT Press.
- Byrnes, P. E. (2015). *Developing automated applications for clustering and outlier detection: Data mining implications for auditing practice*. Doctoral dissertation, Rutgers University Graduate School, Newark, NJ.
- Cao, M., Chychyla, R., & Stewart, T. (2015). Big data analytics in financial statement audits. *Accounting Horizons*, 29(2), 423–429.
- Chan, D. Y., & Vasarhelyi, M. A. (2011). Innovation and practice of continuous auditing. *International Journal of Accounting Information Systems*, 12(2), 152–160.
- Dai, J., & Li, Q. (2016). Designing audit apps for armchair auditors to analyze government procurement contracts. *Journal of Emerging Technologies in Accounting*, 13(2), 71–88.
- Jans, M., Alles, M., & Vasarhelyi, M. (2014). A field study on the use of process mining of event logs as an analytical procedure in auditing. *The Accounting Review*, 89(5), 1751–1773.
- Kozlowski, S. (2016). *A vision of an ENHanced ANalytic constituent environment: ENHANCE*. Doctoral dissertation, Rutgers University Graduate School, Newark, NJ.
- Moffitt, K. C., & Vasarhelyi, M. A. (2013). AIS in an age of Big Data. *Journal of Information Systems*, 27(2), 1–19.
- Moon, D. (2016). *Crowdsourcing for social sciences researchers: Data gathering, teaching, learning and research dissemination from a single project*. London School of Economics and Political Science, LSE Impact Blog. Retrieved from <https://blogs.lse.ac.uk/impactofsocialsciences/2016/10/03/crowdsourcing-for-social-sciences-researchers-data-gathering-teaching-learning-and-research-dissemination-from-a-single-project/>
- O’Leary, D.E. (2015). Armchair auditors: Crowdsourcing analysis of government expenditures. *Journal of Emerging Technologies in Accounting*, 12(1), 71–91.
- Pulliam, S., & Solomon, D. (2002). How three unlikely sleuths exposed fraud at WorldCom. *The Wall Street Journal*, October 30. Retrieved from <https://www.wsj.com/articles/SB1035929943494003751>
- Schneider, G. P., Dai, J., Janvrin, D. J., Ajayi, K., & Raschke, R. L. (2015). Infer, predict, and assure: Accounting opportunities in data analytics. *Accounting Horizons*, 29(3), 719–742.
- Syed, A., Gillela, K., & Venugopal, C. (2013). The future revolution on Big Data. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(6), 2446–2451.
- The Washington Post. (2002). Timeline of Enron’s collapse. *The Washington Post*, February 25.
- Thiprungrsri, S., & Vasarhelyi, M. A. (2011). Cluster analysis for anomaly detection in accounting data: An audit approach. *International Journal of Digital Accounting Research*, 11(17), 69–84.
- Vasarhelyi, M. A., Alles, M. G., & Williams, K. T. (2010). *Continuous Assurance for the now economy. A thought leadership paper for the Institute of Chartered Accountants in Australia*. Queensland, Australia: Institute of Chartered Accountants.
- Yoon, K. (2016). *Three essays on unorthodox audit evidence*. Doctoral dissertation, Rutgers University Graduate School, Newark, NJ.

Part I

**Audit Analytics Procedures**

*This page intentionally left blank*

## Chapter 1

# An Application of Exploratory Data Analysis in Auditing – Credit Card Retention Case\*

*Qi Liu*

### 1. Introduction

Over the last several decades, operational risks in banking systems has attracted both regulatory and academic attention due to the devastating losses experienced by some banks. For example, Allied Irish Banks lost \$750 million due to rogue trading,<sup>1</sup> and Prudential Insurance entered into a \$4 billion class action settlement with regard to fraudulent sales practices over 13 years<sup>2</sup> (Muermann & Oktem, 2002).

An operational audit focuses on evaluating the efficiency, effectiveness, and economy of organizational activities to reduce operational risks and improve future performance (Lane, 1983). It plays an important role in ensuring that organizations achieve their strategies and objectives. This chapter demonstrates an application of exploratory data analysis (EDA) (Tukey, 1977) in a real operational audit setting and illustrates how internal auditors can benefit from this approach.

This case study in this chapter analyzes a data set of credit card annual fee discounts from an international bank in Brazil. In this case study, the EDA process is mainly applied to test three pre-defined audit objectives. The results of the EDA process are compared with the results of conventional audit procedures. The outcomes of this comparison demonstrate that EDA permits the auditor to obtain comprehensive findings easily with simple statistics and visualization techniques.

---

\*This chapter is based on the third chapter of the author's dissertation (Liu, 2014).

<sup>1</sup><http://online.wsj.com/news/articles/SB1012991042190203640>

<sup>2</sup><http://caselaw.findlaw.com/us-3rd-circuit/1362355.html>

The chapter begins with a description of the audit problems facing this bank and then discusses the data and specific methods used in this case. The results of both conventional audit procedures and the EDA process are then presented. Finally, implications and limitations of this case study are discussed.

## **2. The Audit Problem**

### **2.1. Scenario**

This study investigates the credit card division of a large international bank in Brazil. Most of the credit cards issued by this bank have annual fees. Clients who do not want to pay these fees may call the bank and ask for a fee cancellation or a fee reduction. In these circumstances, bank representatives negotiate with the clients about the fees. Based on clients' backgrounds, representatives can then offer appropriate discounts. During the discount negotiation process, bank representatives should follow the bank's policy; they cannot offer discounts higher than they are authorized to give, and they should give top priority to the benefit of the bank. In other words, they should offer the lowest discounts acceptable to the clients.

### **2.2. Audit Objectives**

The bank suggested that the initial audit scope for this study is to identify the bank representatives whose behavior, in the course of the annual fee negotiations, may cause the loss of bank revenue. Risky behaviors include (1) offering higher discounts than allowed; (2) offering high discounts without making an effort to negotiate lower discounts; and (3) offering discounts without any client negotiation. Based on these behaviors, three audit objectives are developed:

- (1) All bank representatives obeyed bank policy when offering discounts.
- (2) Bank representatives offered the lowest possible discounts to retain clients.
- (3) Bank representatives negotiated with clients for lower discounts before offering final discounts.

In addition to these issues, the audit scope is extended to discovering potential operational risks in the annual fee-offering process. Non-behavioral factors, such as lack of effective internal controls, can also lead to loss of revenue. Even though some cases are not directly related to current revenue losses, business process risks may cause future revenue loss.

To achieve these audit objectives, all related fields need to be thoroughly explored for irregularities, making this topic a suitable scenario for EDA. As in a traditional audit, the auditors must gain an understanding of the process and then identify the risks and problems related to this process and its associated internal control system before testing to determine whether any policies have been violated.

### 3. Methodology

#### 3.1. Data

Two data sets are used in this case: the retention data and the account master data. The retention data includes information on customer phone calls made in January 2012. The data set consists of 195,694 records in total. Each record represents a customer's phone call and contains 162 fields.

The account master data is a large data set with 60,309,524 records and 504 fields. Each record represents a credit card account. All accounts opened in the bank from July 1980 to March 2012 are included in the data set. The fields in the account master data cover a wide variety of information relevant to the accounts and accounts holders: account information, such as account type and account status; demographic information, such as account holders' age and gender; and financial information, such as credit limits and late pay amounts. Account master data is updated by the bank on a continuous basis.

This case study uses eight attributes: call length, bank representative ID, supervisor ID, customer service center location, original fee, actual fee, sequence number of the account, and number of cards. Most of these attributes, such as call length, annual fee, and output annual fee, are necessary to test the original audit objectives. Other attributes are newly added during the EDA process, such as supervisor number and number of cards. The names, source database, and descriptions of these attributes are listed in [Table 1](#).

Among these fields, call length, original fees, actual fees, and number of cards are continuous variables. Representative's ID, supervisor's ID, client's ID, account sequential number, and customer service center location are nominal variables. To protect clients' privacy, the account sequential numbers and clients IDs are

Table 1. Description of Attributes Included in This Study.

Attribute Name (Source Database)	Description
Call length (retention)	The duration of each call in seconds
Call location (retention)	The location of the customer service center
Agent number (retention)	ID of the bank representative answering the call
Supervisor number (retention)	ID of the representative's supervisor
Sequential number (retention and account master)	Sequence Number of an account
Annual fee (retention)	Original annual fees of a credit card
Output annual fee (retention)	Actual annual fees paid by each client
Number of cards (account master)	Number of cards associated with each account

encrypted in the data set. The encryption method preserves the integrity of the original data; each original value corresponds to a unique cipher text.

### **3.2. Data Preprocessing**

Discounts offered by bank representatives play an important role in the process of analyzing loss of revenue. However, there is no field that directly reflects discounts in the raw retention data. Two existing fields that relate to discounts are original fees before negotiation and actual fees after negotiation. The difference represents the discount, which is needed to conduct EDA. Specifically, the discount is the difference between original fees and actual fees divided by the original fees. The formula used to calculate discounts is:

$$\text{Discount} = \frac{(\text{Original fee} - \text{Actual fee})}{\text{Original fee}} \times 100\%$$

EDA analyses may require account master data. Therefore, retention data and customer master data need to be joined so that related data elements can be matched. For example, while each client exists only once in the customer master data, each phone call to negotiate discounts creates another item in the retention data set. These many-to-one data sets can be joined based on this relationship. The joining process uses the account sequential number field as it exists in both data sets and is the unique identifier in the Visual Basic for Applications (VBA) data.

### **3.3. Applied EDA Techniques**

In this case study, traditional EDA techniques, such as descriptive statistics, data transformation, and data visualization techniques are mainly used to explore the data. Descriptive statistics used in this study include frequency distribution, summary statistics (mean and standard deviation), and categorical summarization. Data transformation is achieved by the logarithm function. Applied data visualization techniques involve pie charts, bar charts, linear charts, and scatter plots.

## **4. Results and Discussion**

### **4.1. Policy-violating Bank Representatives and Negative Discounts**

**4.1.1. Conventional Audit Procedures.** To determine whether bank representatives are violating bank policy, the maximum discount that each bank representative is allowed to offer according to bank policy must be determined. The bank policy allows bank representatives to offer discounts up to 100% of the annuity to retain the customer, so the conventional audit procedure to test this audit objective is to check whether any bank representatives offered more than 100% discounts. Internal auditors can perform this test simply by applying a filter to select all of the records with discounts greater than 100%. In this case, this filter returned



Table 2. Descriptive Statistics of Discounts.

Field Name	Mean	Median	Minimum	Maximum	Std. Deviation
Discount	-2,326.04%	60%	-27,944,522.22%	100.00%	219,933.88%

no records, indicating that no bank representative violated bank policy. Thus, this audit objective is confirmed by a conventional audit procedure. Auditors can check this box on their checklist and move to the next one.

**4.1.2. EDA Process.** The first step of EDA is to display the distribution of related fields. As bank representatives' discount-offering behaviors are the main concern of the bank, the analysis begins with some descriptive statistics: mean, median, minimum value, maximum value, and standard deviation of the discounts offered by the representatives. The results are shown in Table 2.

According to these results, the maximum discount offered by the bank representatives is 100% of the annual fee. Using this number, the same conclusion can be drawn: No bank representatives offered more than 100% discount. Thus, no bank representative violated bank policy.

This table also shows that the minimum discount is a large negative value (-27,944,522.22%). The mean is also negative (-2,326.04%), which implies that negative discounts overwhelm positive discounts. In addition, the median discount amount is positive (60%) indicating that half of the discounts are larger than 60% and half of the discounts are smaller than 60%. These statistics imply the existence of a few extremely large negative discounts. The frequency distribution of discounts, shown in Fig. 1, also reveals that only 0.15% (286) discounts are negative.

According to the formula for discount, a negative discount means that the actual fee after negotiation is higher than the original annual fee. A negative discount, especially a large one, is counterintuitive.

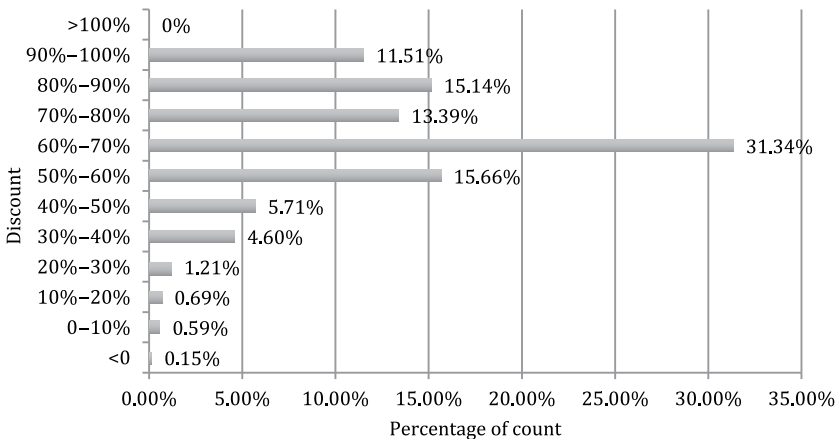


Fig. 1. Frequency Distribution of Discounts.

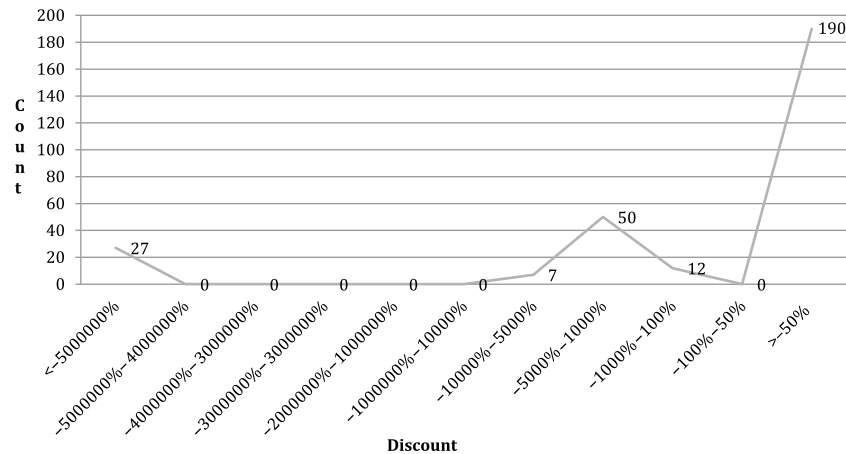


Fig. 2. Distribution of Negative Discounts.

A discussion with the bank’s internal auditors yielded a potential explanation: In some cases, a group of people (e.g., a family) has the same credit card account in the form of primary cards and additional cards. If one of these customers called to negotiate the prices for the whole group, the actual fee may reflect the total actual fees of the group. As the actual fee for all cards may surpass the original fee for one card, the negative discounts may be due to group discounts offered to clients with more than one credit card.

To gain insight into negative discounts, the frequency distribution of negative discounts is calculated and displayed in Fig. 2.

Fig. 2 demonstrates a multimodal and discontinuous distribution of negative discounts featuring three separate clusters. The first cluster contains 27 records (10%) with extreme discounts (lower than  $-5,000,000\%$ ). The second cluster includes 69 data points (24%) associated with relatively significant discounts (between  $-10,000\%$  and  $-100\%$ ). The third and largest cluster involves 190 records with small discounts (less than  $-50\%$ ). The negative discounts in this cluster may due to group discounts. However, this explanation cannot apply to the negative discounts in the other two clusters because of their extreme values. Therefore, these 96 records in the first and second clusters are considered as suspicious cases that may be attributable to errors or frauds.

Even though the remaining 190 cases have reasonable discounts, they are not necessarily group discounts. An easy verification for these data points is to determine whether a given client has multiple cards. The results of this verification are shown in Fig. 3.

Fig. 3 shows that 39 of these 190 clients (20.5%) have only one credit card, so their negative discounts cannot be explained by the effect of group discounts. Therefore, these 39 accounts are also considered suspicious.

As original fees and actual fees are the two determining factors in calculating discounts, the relationship between negative discounts and these two figures is

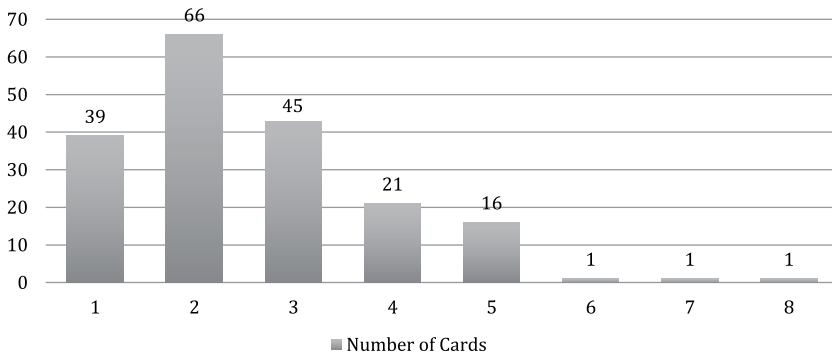


Fig. 3. Frequency Distribution of Number of Cards for the 190 Cases with Reasonable Negative Discounts.

examined to investigate the cause of this distribution. As the ranges of the variables are very wide, the values need to be transformed to another scale to display the data satisfactorily. Specifically, the values of original and actual fees are transformed to their logarithmic values. Due to the negative values, the logarithmic value of the absolute value of negative discounts is calculated. Scatterplots are then used to display the relationship between discounts and actual and original fees, shown in Fig. 4.

Fig. 4 reveals three clusters of negative discounts that are evenly distribute among the original fees. The same three clusters can also be observed in the scatterplot of discounts and actual fees. Hence, the new hypothesis is that these large negative discounts are caused by irregular actual fees.

As the number of extreme negative discounts is manageable, a substantive test is performed to investigate the specific reason for these extreme negative discounts. Among these 96 cases, 27 negative discounts are due to obvious input errors (e.g., dates are mistakenly input as the actual fees). The other 69 negative

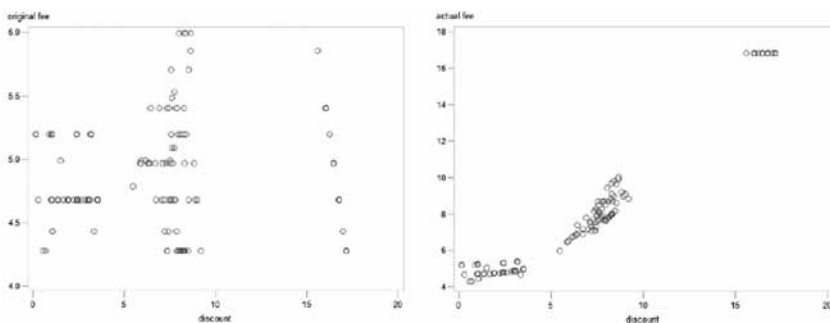


Fig. 4. Relationships between Negative Discounts and Original and Actual Fees.

discounts are caused by round, unreasonably large actual fees. These records may also include input errors, such as incorrect placement of a decimal point.

The analysis of extreme negative discounts points out some risks in the bank’s internal control system. For example, the system should have a control to restrict the input format of each variable so that date format cannot be input into the actual fee field. By setting the upper and lower boundary of each field, the risk of unreasonable, extreme values can be moderated as well. These recommendations and the results of this analysis were reported to the bank, and a new audit objective was developed: Actual fees were recorded correctly.

Based on the EDA analysis, the 39 suspicious cases with reasonable negative discounts were reported to internal auditors for further investigation, and another new audit objective was proposed: Negative discounts have been offered to clients with multiple cards. Overall, using the EDA process to test this audit objective identified 135 abnormal cases, whereas no anomaly can be identified using conventional audit procedures. In addition to identifying these exceptional cases, EDA helped to generate two new audit objectives and suggest two new internal control functions.

**4.2. Lazy and Inactive Bank Representatives**

**4.2.1. Conventional Audit Procedures.** In addition to identifying representatives who violate policy, the bank also wants to identify representatives who make no effort to reduce the discount offered below 100% (i.e., “lazy” representatives). Internal auditors can use conventional audit procedures to calculate the ratio of 100% discounts to all discounts offered by each bank representative. The distribution of this ratio is shown in Fig. 5. Internal auditors can identify lazy representatives by setting a ratio threshold of acceptability. For example, if bank representatives who offer 100% discounts in more than half of their total phone calls are defined as lazy, Fig. 5 shows that internal auditors would identify 59 such representatives who have ratios greater than 0.5.

**4.2.2. EDA Process.** In the EDA process, the representatives who offered 100% discounts are first identified as they are the main concern of this audit

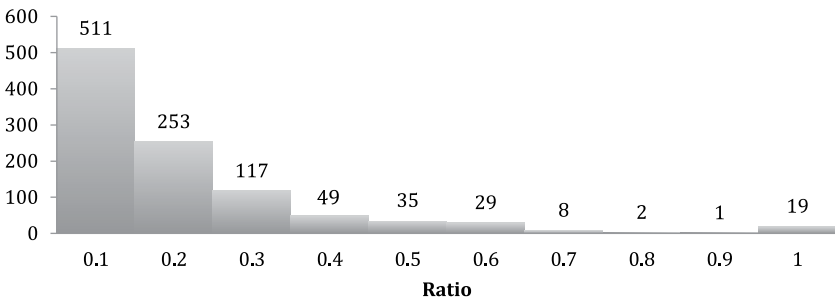


Fig. 5. Frequency Distribution of the Ratio of 100% Discounts to All Discounts Offered by Each Bank Representative.