J. Christopher Westland

# Audit Analytics

## Data Science for the Accounting Profession

Springer

# Contents

**Foreword by Erik Brynjolfsson** ..... v

**Preface** ..... vii

**Fundamentals of Auditing Financial Reports** ..... 1
Auditing ..... 1
Computers in Auditing and the Birth of Audit Analytics ..... 1
The Roots of Modern Financial Accounting and Auditing ..... 3
Al-Khwarizmi Algebra of Double-Entry ..... 4
The Renaissance ..... 6
The Industrial Revolution ..... 6
The Birth of Modern Auditing ..... 7
Public Accounting ..... 8
Emerging Technologies and Intangible Assets ..... 9
Financial Accounting ..... 12
The Products of Accounting: Financial Statements ..... 12
The Methodology of Accounting ..... 14
The Accounting Process and Major Document Files ..... 15
Code and Data Repositories for Audit Analytics ..... 16
R Packages Required for This Book ..... 17
References ..... 18

**Foundations of Audit Analytics** ..... 19
Business and Data Analytics ..... 19
Accounting Data Types ..... 24
Numerical vs. Categorical ..... 25
Categorical (Enums, Enumerated, Factors, Nominal, Polychotomous) Data ..... 26
Binary (Dichotomous, Logical, Indicator, Boolean) Data ..... 28
Ordinal (Ordered Factor) Data ..... 30
Data Storage and Retrieval ..... 33
Further Study ..... 38
*R* Packages Required for This Chapter ..... 39
References ..... 39

**Analysis of Accounting Transactions** ..... 41
Audit Procedures and Accounting Cycles ..... 41
The Origin of Accounting Transactions ..... 42
Audit Tests as Learning Models ..... 42
Working with Dates ..... 43
Accounting Transactions ..... 44
Couching Institutional Language in Statistical Terms ..... 47
Transaction Samples and Populations ..... 48
Accounting Cycles ..... 49

Substantive Testing ..... 50
Metrics and Estimates ..... 50
Important Concepts in Probability and Statistics ..... 52
Machine Learning Methods ..... 58
Statistical Perspectives on Audit Evidence and its Information Content ..... 60
Support and the Additivity of Evidence: The Log-Likelihood ..... 60
The "Score" ..... 61
Fisher Information ..... 61
Reference ..... 61

**Risk Assessment and Planning** ..... 63
Auditing ..... 63
Risk Assessment in Planning the Audit ..... 63
Accessing the SEC's EDGAR Database of Financial Information ..... 65
Caveats on accessing EDGAR information with R ..... 73
Audit Staffing and Budgets ..... 76
The Risk Assessment Matrix ..... 77
Using Shiny to create a Risk Assessment Matrix Dashboard ..... 77
Generating the Audit Budget from the Risk Assessment Matrix ..... 81
Technical Sampling Structure of the Audit Program ..... 81
Sample Sizes for Budgeting ..... 82
Notable Audit Failures and Why They Occurred ..... 84
Auditing: A Wicked Problem ..... 85
Final Thoughts on Audit Planning and Budgets ..... 86
References ..... 86

**Analytical Review: Technical Analysis** ..... 87
Analytical Review ..... 87
Institutional Context of Analytical Review ..... 87
Technical Measures of a Company's Financial Health ..... 90
Purpose and Types of Ratios ..... 90
Common Technical Metrics ..... 91
Accessing Financial Information from EDGAR (https://www.sec.gov/edgar/) ..... 92
Accessing Financial Information from EDGAR (https://www.sec.gov/edgar/) with the `finreportr` Package ..... 95
Computing Technical Metrics ..... 97
Visualization of Technical Metrics ..... 98
Internet Resources for Analytical Review ..... 100
US Census Data ..... 101
R and Application Programming Interfaces (API) ..... 104
Technical Analysis of Product and Customer News Sources on the Web ..... 105
Vocabulary-Based Vectorization ..... 106
References ..... 112

**Analytical Review: Intelligence Scanning** ..... 113
Intelligence Scanning of Internet Resources ..... 113
Sentiment Analysis with Tidy Data ..... 113
Scanning of Uncurated News Sources from Social Networks ..... 121
Example: Extracting Tweets About General Motors and the Auto Industry ..... 123
Intelligence Scanning of Curated News Streams ..... 125
Accessing General Web Content through Web Scraping ..... 131
Final Comments on Analytical Review with R ..... 136

**Design of Audit Programs** ..... 139
Audit Programs as Natural Experiments ..... 139
Collecting and Analyzing Audit Evidence: Sampling ..... 139
AICPA Guidelines on audit sampling ..... 140
Sampling for Interim Tests of Compliance ..... 141

Discovery Sampling ..... 141
Coefficient of Variation Formulas for Sample Size ..... 142
Rules of Threes to Calculate 95% Upper Confidence Bounds ..... 142
Attribute Sampling ..... 143
Acceptance Sampling ..... 143
Acceptance Sampling with Poisson Data ..... 146
A Statistical Decision Framework for Auditing ..... 146
Audit Tests as Learning Models ..... 147
Materiality ..... 149
Risk ..... 149
The AICPA on Sampling Approaches and Risks ..... 150
AICPA Pronouncements on Generally Accepted Auditing Standards ..... 151
Types of Sampling Allowed or Discussed by AICPA ..... 152
Judgmental Sampling ..... 153
Fixed-Interval Sampling ..... 153
Cell or Random-Interval Sampling ..... 153
Random Sampling ..... 153
Conditional Sampling ..... 153
Stratified Sampling ..... 153
Monetary Unit Sampling ..... 154
Transaction or Record Sampling ..... 155
Non-statistical Sampling ..... 155
Accounting Transaction Distributions ..... 156
The Audit Cycle ..... 157
The Context of Auditing and Information Technology ..... 158
Auditors' Opinion: The Product of an Audit ..... 159
References ..... 160

**Interim Compliance Tests** ..... 161
Interim Compliance Tests and the Management Letter ..... 161
The SAS 115 Letter to Management ..... 161
Three Methods of Sampling ..... 162
Discovery Sampling ..... 163
Statistics for Interim Tests ..... 165
Attribute Sampling with t-Tests ..... 168
Audit of Collection Transactions ..... 171
Machine Learning Models for Audit of Controls ..... 175
Autoencoders and Unbalanced Datasets ..... 178
Final Thoughts on Machine Learning Applications in Auditing ..... 187
References ..... 187

**Substantive Tests** ..... 189
Substantive Tests ..... 189
Objective of Substantive Tests ..... 189
Exploratory Substantive Tests ..... 189
Creating Trial Balance Figures in One Step ..... 194
Accounts Receivable Auditing ..... 198
Footing and Agreeing to the Trial Balance ..... 198
Tests of Supporting Evidence ..... 200
Acceptance Sampling ..... 202
Accounts Receivable Confirmation ..... 204
Confirmations and Experimental Design ..... 205
Audit Program Procedures for Accounts Receivable Confirmation ..... 205
Timing of Confirmation Request ..... 205
Confirming Prior to Year-End ..... 205
Steps in Confirmation Process ..... 206

Non-response to Confirmation Requests ..... 208
Confirmation Responses Not Expected ..... 208
Confirmation and Estimation of Error in Account ..... 209
Post-Confirmation Tests ..... 211
Probability Proportional to Size (PPS) Sampling ..... 211
Estimation and Accrual of the Allowance for Doubtful Accounts ..... 212
GLM-Exponential Regression ..... 215
Time-Series Forecasting ..... 216
Forecasting Accounts Receivable Collections ..... 219
Calculating Allowance for Uncollectable Accounts ..... 222
Stratified Samples and Monetary Unit Sampling ..... 223
PPS ..... 224
Stringer's Perspective on Monetary Unit Sampling ..... 225
"Taintings" and the Poisson–Poisson Compound Distribution ..... 226
The Audit of Physical Inventory ..... 227
Periodic Inventory Systems ..... 227
Perpetual Inventory Systems ..... 227
Counting Inventory When Preparing Financial Statements ..... 227
Inventory Systems ..... 228
Physical Inventory (Counting) Process ..... 228
Physical Inventory Count Versus Cycle Counts ..... 229
Inventory Audit Procedures ..... 229
Computer Analytic Workpaper Support ..... 230
Footing, Reconcile the Inventory Count to the General Ledger ..... 232
Cutoff Analysis ..... 233
Duplicates and Omissions ..... 234
Physical Count Exceptions to Perpetual Inventory Ledger ..... 236
Test for Lower of Cost or Market (LOCOM) ..... 237
Audit a Subset of High-Value Items ..... 238
Inventory Allowances ..... 239
Other Inventory Tests ..... 240
References ..... 241

**Sarbanes–Oxley Engagements** ..... 243
The Sarbanes–Oxley Act: Security, Privacy, and Fraud Threats to Firm Systems ..... 243
Academic Research on SOX Effectiveness ..... 244
Evidence from Industry on SOX Effectiveness ..... 245
Using R to Assess SOX Effectiveness in Predicting Breaches, and Identifying Control Weaknesses ..... 246
Exploratory Analysis of the SOX-Privacy Clearninghouse Dataset ..... 249
Using an Autoencoder to Detect Control Weaknesses ..... 252
Preprocessing ..... 253
Tensorflow Implementation of the Autoencoder ..... 255
The H2O Implementation of Autoencoders and Anomaly Detection for Fraud Analytics ..... 257
Anomaly Detection ..... 265
Pre-trained Supervised Model ..... 266
Measuring Model Performance on Highly Unbalanced Data ..... 268
Fama–French Risk Measures and Grid Search of Machine Learning Models to Augment Sarbanes–Oxley Information 271
Fama–French Risk Factors ..... 272
Final Thoughts on Sarbanes–Oxley Reports ..... 277
References ..... 278

**Blockchains, Cybercrime, and Forensics** ..... 279
Blockchains for Securing Transactions ..... 279
The "Block" ..... 279
Hashing ..... 279
Proof-of-Work (PoW) ..... 280

Adding Transactions (New Blocks) to the Blockchain ..... 281
Cybercrime and Forensics ..... 283
Forensic Analytics: Benford's Law ..... 287
References ..... 290

**Special Engagements: Forecasts and Valuation** ..... 291
Special Engagements for Assurance and Valuation ..... 291
The Role of Valuations in the Market ..... 291
Hi-Tech, High Risk ..... 292
Strategy Drivers and Figures of Merit ..... 292
Corporate Figures of Merit ..... 293
The ROI Figure of Merit ..... 293
The Profit Figure of Merit ..... 293
The Sales Revenue Figure of Merit ..... 294
Opportunity Costs ..... 294
Valuation Models ..... 295
The Behavioral (Historical) Model ..... 295
Data: Transaction Stream Time Series ..... 296
How Much Information Is in a Dataset? ..... 297
Forecast Models ..... 298
Discount Model ..... 302
Terminal Dividend ..... 303
Generating a Current Valuation ..... 303
Other Approaches to Valuation ..... 303
Real Options ..... 303
Scenario Analysis and Decision Trees ..... 304
Monte Carlo Simulations ..... 304
Further Study ..... 305
References ..... 305

**Simulated Transactions for Auditing Service Organizations** ..... 307
"Test Decks" and Their Progeny ..... 307
Service Organization Audits ..... 307
Accounting Cycles, Audit Tasks, and the Generation of Audit Data ..... 308
Generation of Sales and Procurement Cycle Databases for A/B Testing of Audit Procedures ..... 308
Setting Up the Simulation ..... 309
Code and Data Repositories for *Audit Analytics* ..... 310
Assumptions Used in the Generation of Simulated Data ..... 310
Document Generation ..... 311
Strategy for Document Generation ..... 313
Statistical Assumptions and the Distribution of Accounting Transactions ..... 313
General Parameters of the Simulation ..... 314
The Sales Journal ..... 315
Cash and Bank Deposits ..... 318
Inventory and Purchase Orders for Inventory Replenishment ..... 318
Inventory ..... 319
Perpetual Inventory, Accounts Payable, and Other Inventory Related Accounts ..... 323
Customer Credit Limits and Outstanding Accounts Receivable ..... 323
Accounts Receivable ..... 324
Accounts Receivable Aging ..... 325
Employee Expenditures ..... 326
Omissions, Duplications, and Monetary Errors in Transactions ..... 327
Audit Tasks: Inventory and Accounts Receivable ..... 332
Accounts Receivable Confirmations ..... 333
Accounting Files for Audit ..... 334
Auditing with Simulated Accounting Transactions ..... 335

References ..... 335

**Index** ..... 337