Marek Gruszczyński

Financial Microeconometrics

A Research Methodology in Corporate Finance and Accounting

Springer

Financial Microeconometrics

Marek Gruszczyński

Financial Microeconometrics

A Research Methodology in Corporate Finance and Accounting

Springer

Marek Gruszczyński
Institute of Econometrics
SGH Warsaw School of Economics
Warsaw, Poland

ISBN 978-3-030-34218-0 ISBN 978-3-030-34219-7 (eBook)
https://doi.org/10.1007/978-3-030-34219-7

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book explores topics of modern research in corporate finance and accounting. Emphasis is placed on both methodology and applications, jointly labeled as financial microeconometrics. How can financial distress in companies be predicted? Is corporate governance associated with financial results? What are the determinants of accounting disclosures? These and other questions may be reliably examined utilizing the methods explored in this book. The clear presentation of methods is buttressed by numerous examples from corporate finance and accounting literature.

Reportedly, this is the first English-language monograph on financial microeconometrics, the term first coined by the author in a publication in 2008. Financial microeconometrics covers topics known as "empirical corporate finance" and "applied accounting" especially in the subjects connected with the econometric analysis of microdata. "Financial microeconometrics emerges as a natural consequence of applying statistical and econometric methods to corporate finance, accounting and other fields of finance. The *applied* research in accounting and corporate finance is inevitably linked with the use of notions like statistical sample, population, and the operation on sets of microdata" (Gruszczyński 2018a).

This book demonstrates perhaps in a somewhat simplified form how quantitative (statistical, econometric) methods may be applied in corporate finance and accounting research. It is intended for practitioners, researchers, and students who are not yet familiar with the variety of approaches extant in data analysis/microeconometrics.

Textbooks and monographs on financial econometrics typically do not include analyses of financial microdata (Wang 2009 being a notable exception). Two recent textbooks on financial econometrics are time series oriented: Bofetti and Urga (2016) and Fan and Yao (2017). "On the other hand, microeconometrics tools are justifiably included in modern classes on empirical corporate finance (e.g., Ph.D. class of prof. Da Rin at Bocconi, 2016–2017) and in some basic textbooks, like Damodaran (2014) or Berk and DeMarzo (2017, Chap. 16)" (Gruszczyński 2018a).

The backstory of this book begins in the early transformation period in Poland. After the 1989 roundtable accord between Solidarity and the authorities, the new government—with Leszek Balcerowicz at the helm of the country's finances—

introduced Western-style management and opened the country to investors. In the 1990s, I was involved in numerous projects of privatization and restructuring of state-owned companies. As an academic econometrician and lecturer at SGH Warsaw School of Economics, I worked hard to expand my knowledge base and add depth to my expertise. This period—challenging for all Poles—was professionally fruitful for many of us academics. We learned—and immediately applied in the real world—the basics of Western-style corporate finance, accounting, corporate law, management, etc. Obviously, the constantly changing laws governing business in Poland had to be monitored and implemented.

Simultaneously, the field of econometrics, as always, was expanding into new frontiers. My research interest in microeconometrics was enhanced by major developments in data quality and accessibility. In 2000, two microeconometricians Daniel McFadden and James Heckman were awarded the Nobel Prize in Economics. The fields of microeconometrics and data analysis were continuing to advance in applications, both operationally and as vehicles for expanding scientific research.

After returning full-time to academia around 2000, I employed my new corporate finance and accounting experience in research grounded in statistical-econometric methodology. The results were research projects, books, and papers. My major theme has been the application of microeconometric methods in corporate finance and accounting. This book is about those applications as I see them from an econometrician's perspective and as I advocate their application (or not) to finance and accounting researchers and practitioners in those fields.

The author's earlier version of this book was published in Polish in 2012 and has found success among researchers, students, and practitioners in corporate finance and accounting within Poland (Gruszczyński 2012a). This new book, written in English, is rooted in the fundamental concept of the author's earlier Polish text. I have borrowed from it the structural framework for this book and the main elements of the presentation of financial microeconometrics. The major narratives on each topic have been rewritten, incorporating new developments. The set of examples—the core of this book—has been updated and expanded. Completely new are sections on the "new microeconometrics" in Chap. 2 as well as a fresh look at good practices in financial microeconometrics, also in Chap. 2. Chapters 3–6 also include substantial new parts. Overall, about 50% of this book contains new text (as compared with the 2012 Polish version).

The primary message of this undertaking is as follows: While making decisions or preparing opinions about specific companies—their finances, transactions, governance, etc.—do not forget the statistical (econometric) evidence about how those subjects fare or fared among other companies (in a sample). On a cautionary note, the book warns that evidence from statistical-econometric research is often flawed and, at a minimum, should not be taken as applicable to other time periods, to other sets of data, and to other industries.

My approach is to demonstrate ample research-based evidence to refer to in case it is needed. Also, I strive to present—perhaps in simplified terms—how research is being conducted and how (easily) its results can be challenged. This book attempts to show the current state of financial microeconometrics as presented by researchers and lecturers worldwide. Inevitably, the content does not encompass all possible topics, aspects, or solutions that comprise the expanding field of empirical corporate finance and applied accounting.

My gratitude goes to the participants of numerous conferences in Poland and worldwide for the discussions on my presentations pertaining to various topics mentioned in this book. Especially important for me were the conferences of the International Atlantic Economic Society in the USA, Canada, and Europe and the conferences on investment and finance organized by the University of Szczecin, Wrocław University of Economics, the University of Łódź, and other institutions. I also thank my MA and PhD students at SGH Warsaw School of Economics. This book includes several results presented in their theses. Finally, I would like to thank J. Richard Quigley for his thorough proofreading and useful suggestions that I incorporated into this book. Obviously, all remaining errors are mine.

Warsaw, Poland
September 2019

Marek Gruszczyński

References

Berk J, DeMarzo P (2017) Corporate finance, 4th edn. Pearson, Sydney

Bofetti S, Urga G (2016) Financial econometrics using Stata. Stata Press, College Station, TX

Damodaran A (2014) Applied corporate finance, 4th edn. Wiley, Hoboken, NJ

Fan J, Yao Q (2017) Elements of financial econometrics. Cambridge University Press, Cambridge

Gruszczyński M (2012a) Empiryczne finanse przedsiębiorstw. Mikroekonometria finansowa [Empirical corporate finance. Financial microeconometrics]. Difin, Warszawa

Gruszczyński M (2018a) Financial microeconometrics as research methodology in corporate finance and accounting. In: Dudycz T, Osbert-Pociecha G, Brycz B (eds) Efficiency in business and economics. Springer proceedings in business and economics. Springer, New York, pp 71–80

Wang P (2009) Financial econometrics, 2nd edn. Routledge, London

Contents

List of Abbreviations

Methodology

2SLS	Two-stage least squares
ATE	Average treatment effect
ATT	Treatment effect on the treated
AUC	Area under curve
CDF	Cumulative distribution function
CIA	Conditional dependence assumption
CPH	Cox proportional hazard
Diff-in-Diff	Difference-in-differences
DPS-GMM	Dynamic panel system GMM
ECF	Empirical corporate finance
EV1	Extreme value distribution type 1
FE	Financial econometrics
FM	Financial microeconometrics
GLS	Generalized least squares
GMM	Generalized method of moments
IID	Independent and identically distributed
IMR	Inverse Mills ratio
IV	Instrumental variable
LDA	Linear discriminant analysis
LPM	Linear probability model
LS	Least squares method
ME	Marginal effect
ML	Maximum likelihood method
OLS	Ordinary least squares
OR	Odds ratio
PSM	Propensity score matching
RDD	Regression discontinuity design

ROC	Receiver operating characteristic
ROWR	Random oversampling with replication
RU	Random undersampling
SMOTE	Synthetic minority oversampling technique
SVM	Support vector machine
TE	Treatment effect
WoE	Weights of evidence

Other

AIMD	Artificial intelligence measurement of disclosure
AIMR	Association for Investment Management and Research
ASA	American Statistical Association
ASX	Australian Stock Exchange
BCDI	Brazilian Corporate Disclosure Index
BV	Book value
CEO	Chief executive officer
CFA	Chartered financial analyst
CG	Corporate governance
CIFAR	The Center for International Financial Analysis and Research
CSR	Corporate social responsibility
EAA	European Accounting Association
EBIT	Earnings before interest and taxes
EBITDA	Earnings before interest, taxes, depreciation, and amortization
ECGI	European Corporate Governance Institute
EPS	Earnings per share
FCPA	Foreign Corrupt Practices Act
FD	Financial distress
GEM	Hong Kong Growth Enterprise Market
IAS	International Accounting Standard
IFRS	International Financial Reporting Standards
IPO	Initial public offering
IRRCI	Investor Responsibility Research Center Institute
M&A	Mergers and acquisitions
MBAR	Market-based accounting research
NYSE	New York Stock Exchange
OECD	Organisation for Economic Co-operation and Development
P/E	Price-earnings ratio
PCDI	Polish Corporate Disclosure Index
PFCG	Polish Corporate Governance Forum
PID	Polish Institute of Directors
R&D	Research and development
ROA	Return on assets

ROE	Return on equity
SDC	Securities Data Corporation
SOX	Sarbanes–Oxley Act
SRD II	Shareholder Rights Directive II
VR	Value relevance
WSE	Warsaw Stock Exchange

List of Examples

Chapter 1
Introduction

Financial microeconometrics is the application of microeconometric methods in corporate finance and accounting research. This introductory chapter focuses on the place of financial microeconometrics within the framework of corporate finance, accounting, corporate law, financial management, corporate governance, statistical methods, and financial econometrics. Reflections on the theory of corporate finance are complemented by reflections on its data-oriented counterpart—i.e., empirical corporate finance. The final part of the chapter is devoted to issues regarding the samples of microdata used in corporate finance and accounting research.

1.1 The Core of This Book: Microdata, Microeconometrics, Corporate Finance, and Accounting

This book focuses on microeconometrics as applied to corporate finance and accounting. It can be read as a text about these domains from the perspective of statistical samples of companies, transactions, and corporate events. This empirical perspective puts corporate finance and accounting in a statistical setting. We study companies and their financial decisions—but always many companies and many decisions. We do not reflect on when and how to carry out an IPO in a specific company. Instead, we attempt to specify what the determinants of an offering are and how an IPO can influence a company's prospects—by analyzing a large group of companies that have undergone the same experience in the past.

M. Gruszczyński, *Financial Microeconometrics*,
https://doi.org/10.1007/978-3-030-34219-7_1

Microeconometrics

Readers may like to know firstly what microeconometrics is. It obviously is econometrics (i.e., the modeling of economic, financial, and managerial relationships between particular variables) with the use of statistical data. Specifically, for microeconometrics we have microdata—data about single companies, transactions, and events. James Heckman in his Nobel Prize lecture in 2000 (Heckman 2000) states, "Microeconometrics is a scientific field within economics that links the theory of individual behavior to individual data where individuals may be firms, persons or households." As defined by Cameron and Trivedi (2005), "Microeconometric analysis is the analysis of individual-level data on the economic behavior of individuals or firms. . . . Usually regression methods are applied to cross-section or panel data."

Heckman (2001, 2004) summarizes the contributions of microeconometrics to economic knowledge in four main themes:

> (1) Microeconometricians developed new tools to respond to econometric problems raised by the analysis of the new sources of microdata produced after the Second World War. (2) Microeconometrics improved on aggregate time series methods by building models that linked economic models for individuals to data on individual behavior. (3) An important empirical regularity detected by the field is the diversity and heterogeneity of behavior. This heterogeneity has profound consequences for economic theory and for econometric practice. (4) Microeconometrics has contributed substantially to the scientific evaluation of public policy. (Heckman 2001, p. 255)

The need for microeconometric analyses matches the emerging profusion of microdata available for economic and social research. Equally plentiful is the demand for lower level economic models explaining individual behavior rather than making use of the average "representative agent." Again, Heckman (2000) states, "Microeconometrics extended the Cowles theory by building richer economic models where heterogeneity of agents plays a fundamental role and where the equations being estimated are more closely linked to individual data and individual choice models."

Vast sources of microdata are available operationally in today's business environment. From that perspective, microeconometric methods are a part of what is popularly labelled *data mining* or the identification of patterns in data. Microeconometrics itself, by providing tools for (economic) microdata analysis, may be considered as a platform for developing the fields in which it can be applied. Therefore, one can consider, for example, labor microeconometrics, social policy microeconometrics, and financial microeconometrics.

Microdata

Quantitative research in corporate finance and accounting commonly uses microdata on companies, their activities, legal facts, social behavior, etc. Typical examples are financial statements, legal registers, and statistical reports, always from a potentially large number of companies—i.e., hundreds or thousands.

Microdata are objective or subjective. For companies the data we collect are objective—in the sense that they present what the company choses to present, such as financial statements. Subjective data are obtained in various polls, surveys, etc. For example, a survey of companies by a central bank (or research institute) inquiring about business prediction will return answers that are the "subjective opinion of a company." Similar data are obtained from individuals in various social surveys.

Microdata for empirical corporate finance and accounting are usually cross-sectional or panel data. In the latter case, the time dimension is usually small. Microdata used in this type of research are called observational data since they are collected from official databases or from questionnaire surveys. Data of such type may be biased by selection error if the samples are not randomly selected.

Financial Microeconometrics

Financial microeconometrics (FM) emerges as a natural consequence of applying statistical and econometric methods to corporate finance, accounting, and other fields of finance. Since microdata are the daily outcome in finance, FM deals with the relationships that are the result of analyzing microdata on consumers, investors, companies, etc. The applied side of research in accounting and corporate finance is inevitably linked with the use of concepts like statistical sample, population, and operations on sets of microdata. In such research, sets of microdata on companies substitute for a single company, microdata on transactions substitute for a single transaction, events substitute for an event, etc. The tools of microeconometrics are commonly applied in research on corporate finance and accounting, as well as in research on nonfinancial managerial subjects, such as corporate governance. The term "financial microeconometrics" was first introduced by the author (Gruszczyński 2006, 2008, 2012a).

The target of FM is finance. The methods belong to microeconometrics. Therefore, FM is finance, not econometrics. Obviously, FM is a part of financial econometrics; the field is well advanced today, especially in the time series context. The cross-sectional perspective of financial econometrics is easily skipped in classical lectures. Here, we stress its importance.

Empirical Corporate Finance and Applied Accounting

FM is the application of microeconometric methods in corporate finance and accounting research. This methodology constitutes a major foundation of empirical corporate finance (ECF). To verify its research questions, ECF uses large datasets of

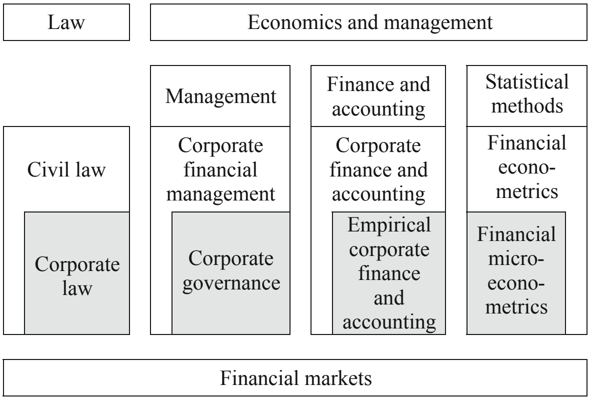


Fig. 1.1 Empirical corporate finance and financial microeconometrics (Source: Gruszczyński 2012a, 2018)

companies, their financials, decisions, etc. This is where the methods of FM are usually indispensable. This is shown in Fig. 1.1.

ECF and FM—the lower right-hand corner of Fig. 1.1—are of major interest in this exposition. However, in order to properly execute research in corporate finance and accounting, it is inevitable to simultaneously consider questions relating to law and management, corporate law, corporate governance, and—finally—financial markets. It is the synergy of expertise from all these fields, which appears necessary for meaningful research in corporate finance and accounting.

FM and ECF

Financial microeconometrics is primarily empirical corporate finance. It covers all econometric research based on individual financial data about companies:

- FM is Altman's discriminant analysis Z-score models and Fama–French's CAPM models, etc.
- FM is cross-section and panel data models constructed for listed as well as for unlisted companies.
- FM is model of IPO effects; models of behavioral corporate finance; statistical models of fundamental investment strategies; various quantitative accounting models (e.g., for analyzing disclosures).
- FM represents many empirical corporate governance studies.

Obviously, significant segments of empirical corporate finance and applied accounting do not use the methodology of microeconometrics. There are other data analytic methods that might be and are applied in ECF research. Then again, a strict distinction between methods appears today to be artificial: all methods seek the best sets of determinants and/or the best fit to the major (dependent) variable in question.

A Pragmatic Approach in Corporate Finance and Accounting

In order to properly assess the hypotheses concerning companies, their managers' decisions, and the reception of these from the outside—always from the perspective of larger datasets—we should approach them from various disciplines, including:

- Finance, especially corporate finance
- Statistics/econometrics, particularly microeconometrics
- Accounting, including corporate accounting
- Management, particularly financial management
- Corporate law, especially corporate governance provisions

The list above includes economics (encompassing finance and management) as well as law (civil and corporate law). We are involved in several fields of the economic sciences: management (including corporate governance), accounting, corporate finance, and finally the field representing quantitative (statistical-econometric) methods. The coexistence of these fields is important in specific research exercises that are discussed later in this book.

Subject areas of financial microeconometrics include all instances in corporate finance and accounting in which microdata may help to facilitate either research solutions or instant operational decisions. The complexity of these broad areas of finance necessitates narrowing the exposition in a book such as this. We, therefore, concentrate here on selected subjects and on selected techniques.

FM and ECF in the Classroom

Topics considered in this book may be of interest to readers who are studying corporate finance and accounting in courses with a research focus in the syllabi. Empirical corporate finance (i.e., financial microeconometrics) is the subject of several advanced courses at the Master and PhD levels, and classes of empirical corporate finance are increasingly popular. Universities offering such classes include Bocconi, Harvard Business School, Aarhus University, École Polytechnique Fédérale de Lausanne, Columbia Business School, University of New South Wales, NHH Norwegian School of Economics, Munich School of Management, and many others.

1.2 Corporate Finance, the Theory of Corporate Finance, and the Theory of Economics

Corporate finance is the major field of finance that, together with corporate accounting, is dedicated to explaining the connections related to the financial aspects of a company's operations. Corporate finance belongs to finance and finance belongs to the economic sciences. Theories of corporate finance are theories of economics. In fact, many representatives of finance are among the recipients of the Nobel Memorial Prize in Economic Sciences. Notable names are the following: Modigliani (Nobel Prize awarded in 1985); Markowitz, Miller, and Sharpe (in 1990); Merton and Scholes (in 1997); Akerlof, Spence, and Stiglitz (in 2001); Fama, Hansen, and Schiller (in 2013); and Tirole (in 2014).

Corporate finance deals with all the financial and monetary events or phenomena linked with the activity of a company. The theory of corporate finance is one of the most dynamic areas in the theory of economics. It is based on research that is relatively new—conducted over the last 30–40 years. The canon of corporate finance is being modified along with enterprises that encounter dynamic changes in the global world of today. The theory of corporate finance is comprehensively presented in few works, of which the leading example is that by Jean Tirole (2006).

Corporate finance books and textbooks usually concentrate on tools, with little emphasis on theory. In fact, most are more devoted to corporate financial management (i.e., the practical tools for managing company finances, with lots of examples and cases). Texts such as those aimed at practitioners present information about a single company, from the viewpoint of management, owners, and outside investors. The list of "classical" textbooks compiled by Fernandez (2017) has 150 titles. That many books were consulted by the author for the survey about how *risk premium* is presented in corporate finance literature.

Over the last 10–20 years, the contents of the major textbooks on corporate financial management have changed. Sections explaining the theory of finance have been shortened or made readable for a more general audience. Reflections on "why" have been replaced by sections explaining "how." The practical side of corporate finance is important and in more demand. Theories change, as does their practical value.

In the 13th edition of their seminal textbook *Intermediate Financial Management*, Brigham and Daves (2019) chose to present theory in short sections accompanying some chapters of the book, in contrast to an earlier version of the same text Brigham and Gapenski's 5th edition (1996), which introduced basic theories of finance in Chap. 1. In any case, all the important theories are presented in all editions. These are:

- Modigliani and Miller's capital structure theory (MM), and related theories such as modified MM theory, dividend MM theory
- Markowitz's portfolio theory and Sharpe's capital asset pricing model
- Black–Scholes' option pricing model
- Fama–French three-factor model
- Agency theory
- Pecking order, signaling, and informational asymmetry

The books of Aswath Damodaran present another example. This influential author is also concerned with conveying practical issues of corporate finance and valuation along with demonstrating their proper relevance to theory. In the 4th edition of *Applied Corporate Finance*, his definition of corporate finance is simple:

"Every decision made in a business has financial implications, and any decision that involves the use of money is a corporate financial decision" (Damodaran 2014). He states that there are three principles "that govern and guide everything that gets done" within corporate finance:

- *The investment principle*: Invest in assets and projects that yield a return greater than the minimum acceptable hurdle rate (a hurdle rate is the minimum acceptable rate of return for investing resources in a new investment; the hurdle rate should be higher for riskier projects and should reflect the financing mix).
- *The financing principle*: Choose a financing mix (debt and equity) that maximizes the value of the investments made and match the financing to the nature of the assets being financed.
- *The dividend principle*: If there are not enough investments that earn the hurdle rate, return the cash to the owners of the business.

Damodaran (2014) considers corporate finance in five areas: (1) the objective of the firm: maximization of the firm's value; (2) the investment principle; (3) the financing principle; (4) the dividend principle; and (5) the firm's value linked to the investment, and financing and dividend decisions.

This short list of textbooks in corporate finance concludes with *Corporate Finance* 11th edition by Ross et al. (2015). This edition of the book includes several sections on corporate governance and on the last financial crisis—in relation to corporate finance. Theories of finance are examined in various sections, usually with applicative focus. In the introductory (first) chapter, the authors state that

> Corporate finance has three main areas of concern: (1) capital budgeting: What long-term investments should the firm take? (2) capital structure: Where will the firm get the long-term financing to pay for its investments? Also, what mixture of debt and equity should it use to fund operations? (3) working capital management: How should the firm manage its everyday financial activities?

Ross et al. (2015) state firmly that the goal of financial management is maximization of the market value of equity (at least for a for-profit business) and that is enhanced by the existence of financial markets. The corporate organizational form has advantages and the significant disadvantage of double taxation. Also, agency problems (i.e., conflicts between stockholders and management) may cause possible distortion that, however, can be controlled and reduced. A notable feature of the book by Ross et al. (2015) is the inclusion of an up-to-date discussion of corporate financial distress as a separate chapter.

As can be seen from this brief discussion, corporate finance is a major component of economics that has resulted in Nobel prizes for several academics from the field. It also comprises very practical knowledge as shown in our brief review of textbooks on corporate financial management.

1.3 The Theory of Corporate Finance: Tirole

The theory of corporate finance, deeply rooted in economics, has a research counterpart that is empirical corporate finance. We now turn to both in the form of concise reviews of two works: *The Theory of Corporate Finance* by the Nobel Prize laureate Jean Tirole (2006) and the two volumes *Handbook of Corporate Finance. Empirical Corporate Finance* (2007, 2008) edited by B. Espen Eckbo.

In the 16 chapters of his book, Tirole (2006) considers and presents the following main branches of the modern theory of corporate finance embedded in microeconomics:

1. Corporate financing and agency costs—From the standpoint of incentives for the firm's insiders (managers, entrepreneurs, borrowers). The incentive scheme for insiders should best align the interests of insiders and outsiders (investors, lenders). These are the major topics in that area:

   - Financing from outside; the fixed investment; the moral-hazard model of credit rationing.
   - The determinants of borrowing capacity.
   - Liquidity management; the optimal design of debt maturity (multiperiod financing); optimal liquidity and risk management.
   - Asymmetric information between insiders and outsiders at the financing stage; negative stock price reaction associated with equity offerings; the "pecking order hypothesis" according to which issuers have a preference order for funding their investments (retained earnings, debt, hybrid securities, equity).
   - Interaction between corporate finance and product–market competition; insider incentive problems; "manipulation of performance measures."

2. Outsiders' and outsiders' incentives—Passive and active monitoring on the part of outsiders. Two subbranches are indicated here:

   - Exit and voice—Passive and active monitoring: monitoring of management by one or several securityholders (e.g., large owner and main bank); the social costs and benefits of passive monitoring; active monitoring (e.g., by lenders) for curbing a borrower's moral hazard.
   - Security design—The control rights view: the control rights approach to corporate finance; allocation of formal control between insiders and outsiders; the raider's ability to take over a firm (the "normative theory of takeovers," the classical theory of the tendering of shares, the free rider problem, poison pills, and dual class voting rules).

3. Consumer liquidity demand—i.e., the demand for liquidity on the part of investors ("consumers"): security holders differ in their preferences for state contingent returns; consumers face personal shocks and value the flexibility of being able to realize their assets when needed; consumers compete with corporations for the available stock of liquidity.

4. Implications of corporate finance for macroeconomic activity and policy. This area includes the following subjects:
   - Credit rationing and economic activity—Classic topics relating credit constraints with recessions and booms (“balance-sheet channel” and “lending channel”).
   - Capital reallocations (mergers and acquisitions, sales of property, plants and equipment); endogenization of the resale value of assets.
   - Existence of stores of value in the economy—Stores of value condition the corporate sector’s ability to meet liquidity shocks in the aggregate; shortage of this “inside liquidity” makes “outside liquidity” valuable.
   - Institutions, public policy, and the political economy of finance—There exists the “topsy-turvy principle” of policy preferences: For a widespread variety of public policies, the relative preference of heterogeneous borrowers switches over time: borrowers with weak balance sheets have, before they receive funding, the highest demand for investor friendly public policies, but they are the keenest to lobby to have these policies repudiated once they have secured financing. This principle affects the legal enforcement of collateral, income, and control rights pledges made by borrowers.

Topics considerately omitted by the author are specific questions on taxes, speculative bubbles, areas of behavioral finance (irrational managers/investors), financial innovations, and international finance. Also, the “empirics” of corporate finance is not much present in the book.

It is problematic to define precisely the theory of corporate finance as a separate discipline (or as a subdiscipline). This difficulty is common to many areas of finance and economics and, therefore, is destined to be more eclectic than rigorously uniform. In the introduction to his book, Tirole (2006) highlights that a unified theory of corporate finance has not emerged, despite enormous progress in the previous 20 years. He identifies the Arrow–Debreu general equilibrium model as the base for theoretical considerations, followed by Modigliani and Miller on financial structure and, recently, by Jensen–Meckling and others on agency problems and corporate structure.

Theoretical works in corporate finance are now emerging daily, as in other fields of finance and economics. Various streams are developing at different paces. More than ever before, it is expected that a theoretical consideration is followed by an empirical counterpart. Empirical corporate finance and accounting are increasingly taking over the field in terms of research interest and the number of publications. Since specific theories prove to be unstable over time and space—as shown by empirics—there is, perhaps, no need to strive for a unified theory of corporate finance. Instead, there is growing demand for more “operational” results based on statistical data on companies.

The book by Tirole (2006) on the theory of corporate finance may be complemented by other works like *Theoretical Foundations of Corporate Finance* by Joao Amaro de Matos (2001) and *Financial Markets and Corporate Strategy* by Grinblatt et al. (2011).

Contemporary publications on corporate finance may be categorized into three streams:

1. Theoretical
2. Empirical: Research oriented
3. Empirical: Practical (operational)

Each stream is valuable, and it is pointless to ponder how to unify them. Moreover, the classification is not precise: many works belong in more than one stream, and the streams may overlap as well.

Both this and the previous section introduced works from the first and the third streams. The second stream is discussed in the next section and, actually, in all the following chapters of the book.

1.4 Empirical Corporate Finance/Financial Microeconometrics: Eckbo

Eckbo

The data-oriented counterpart to the theory of corporate finance is empirical corporate finance (ECF). ECF validates this theory as well as opens new areas for research and verification in the corporate finance field. Contemporary research usually links the theory of corporate finance with the empirical approach. It is due to the demand for more practical results that may be translated into guidance for the parties that create a firm's value. These are insiders (managers, entrepreneurs, borrowers) and outsiders (investors, lenders).

The methodological part of ECF is financial microeconometrics (FM), as indicated in Sect. 1.1. Here we use the terms ECF and FM as equivalent.

The quantity of ECF literature is enormous and growing daily. In this section, the focus is on books, which are surveys or summaries of current discussions on specific topics. To date the most important texts for ECF are the two volumes of *Handbook of Corporate Finance. Empirical Corporate Finance* (2007, 2008) edited by Professor B. Espen Eckbo. The first volume appeared in 2007, the second in 2008. This work contains survey articles on topics of empirical corporate finance. The editor admits that this is just a snapshot of the then current state of ECF research and does not pretend to generalizations since the literature is vast and growing.

Another work worth mentioning is the four volume set entitled *Empirical Corporate Finance*, edited by Brennan (2001), containing reprints of papers from 1969 to 1999. We should also acknowledge a two volume text *Corporate Takeovers. Modern Empirical Developments* edited by Eckbo (2010) that includes reprints of papers on mergers and acquisitions from 1983 to 2009.

The volumes of *Handbook of Corporate Finance. Empirical Corporate Finance* (2007, 2008) contain chapters on methodology and on topics that are part of ECF.

What comprises ECF according to Eckbo? In terms of methodology, it is obviously microeconometrics. That is evident from the contents of both volumes. *Handbook* is divided into four main parts:

1. Econometric issues and methodological trends
2. Banking, public offerings, and private sources of capital
3. Dividends, capital structure, and financial distress
4. Takeovers, restructurings, and managerial incentives

The most significant areas included by Eckbo within empirical corporate finance are[1]

- The econometrics of event studies
- Self-selection models in corporate finance
- Auctions in corporate finance
- Behavioral corporate finance
- Banks in capital markets (the dual role of banks as creditors and data gathering entities)
- Security offerings (including valuation errors of IPOs)
- Conglomerate firms and internal capital markets (company diversification and its effect on valuation; conglomerate's discount)
- Financial institutions and corporate finance (commercial banks, investment banks, venture capital)
- Payout policy (dividend policy)
- Taxes and corporate finance (company's tax policy, capital structure, dividend strategy, risk management, revenues management, transfer prices, etc.)
- Trade-off and pecking order theories of debt
- Capital structure and corporate strategy
- Bankruptcy and the resolution of financial distress
- Corporate takeovers
- Corporate restructuring (financial and asset restructuring)
- Executive compensation and incentives
- Managing corporate risk

To a great extent, these topics are parallel to those considered by Tirole (2006) in the framework for the theory of corporate finance. The surveys presented in the volumes edited by Eckbo (2007, 2008) are primarily based on the experience and the results of US companies. Of significant value in these reviews are the concise lists of references, which may be evidence of the maturity of the authors' reflections. A notable shortcoming, however, is the absence of topics related to corporate governance, a failing noted and acknowledged by the editor.

[1]Gruszczyński (2009) offers a detailed review of the two volumes by Eckbo (2007, 2008) (in Polish).

Literature on ECF/FM

As evidenced in the previous section, empirical corporate finance encompasses a broad spectrum of topics, many of which are also topics of empirical accounting (presented in Chap. 4). Studies that belong to ECF are numerous and sometimes not straightforward as to the fields into which they should be classified. ECF literature is also literature on financial microeconometrics, although not specifically identified as such.

As in other social sciences, empirical studies in economics and finance (corporate finance included) often produce inconclusive results, very much dependent on the financial market, the characteristics of the sample, the period of investigation, etc. Major results are presented in leading journals, as well as in survey volumes.

The following is a subjective enumeration of journals dedicated primarily to publishing quality papers on empirical corporate finance/financial microeconometric topics: *Abacus*; *Advances in International Accounting*; *Accounting and Finance*; *Accounting Review*; *Corporate Governance: An International Review*; *Financial Analysts Journal*; *International Journal of Accounting*; *International Review of Financial Analysis*; *Journal of Accounting and Economics*; *Journal of Accounting and Public Policy*; *Journal of Accounting Research*; *Journal of Applied Corporate Finance*; *Journal of Banking and Finance*; *Journal of Business, Finance and Accounting*; *Journal of Corporate Finance*; *Journal of Economics and Business*; *Journal of Empirical Finance*; *Journal of Finance*; *Journal of Financial Economics*; *Journal of Financial and Quantitative Analysis*; *Journal of Law and Economics*; *Quarterly Journal of Finance and Accounting*; *Review of Economic Studies*; *Review of Financial Economics*; and *Review of Financial Studies*.

This list includes corporate finance, accounting, and economics journals. Included is only a single journal related to corporate governance, while there are no journals whose focus is financial law, corporate law, or management. ECF literature may be found in a vast number of good quality journals, some of which are included above. Moreover, papers on empirical corporate finance also appear on quality online journals like those on *SSRN (Social Science Research Network)*: *Corporate Finance eJournals*, *Accounting Research Network eJournals*, *Corporate Governance eJournals*, and *Econometrics and Financial Economics eJournals*. These include selected papers and working papers on subjects that may be classified within one or more eJournals.

In the following chapters, we will refer to several papers on ECF/FM that have been published in various journals.

1.5 Sample Considerations

Sources of Microdata for FM

A list of possible sources of company microdata gathered in various registers, data collections, commercial repositories, etc. is beyond the scope of this book. There are, obviously, well-known collections like Orbis/Amadeus in Europe, Ifo Economics and Business Data Center in Munich, Datastream (Thomson Reuters), Worldscope (also Thomson Reuters), and Compustat. Additional resources are the regionally operated databases in various national and international markets, including Infocredit in Poland.

Data on companies listed on stock exchanges are easily available due to the transparency that prevails across markets. Data for listed companies, collected in various bases, are, therefore, readily available to professionals and to researchers.

On a considerably less sanguine note, however, convenient use of data for purposes of microeconometric research is not easy. In order to collect large samples, the researcher must "clean" the data taking into account various aspects. The preprocessing phase is challenging and time consuming. For example, the comparison of fundamental data between countries requires considerable effort in explaining and catching the cross-country differences. Moreover, there are many errors in records, unlikely items, and missing data. To facilitate research, the financial data on a large group of companies in Poland are consistently checked by a research group composed of financial experts from academic and professional circles. Similar undertakings are implemented in many other countries.

Sampling

Research in empirical corporate finance and accounting uses samples of companies, statements, transactions, business events, etc. The best are random samples from a population. In this book, we sometimes refer to such samples as the ideal solution.

In simple random sampling, every element of a population has the same probability of being chosen for the sample. However, this sampling scheme is not convenient in the case of large populations. Nor is it correct that such sampling is best in terms of the adequacy of the results, as simple random sampling does not account for the features of the unit (individual, company, etc.). Instead of a pure random sample, stratified (multistage) sampling is advocated. This is usually performed by national statistical offices for purposes of supplying information on social issues.

Usually, we have samples that are nonrandom—i.e., from datasets that are not complete or are collected in a nonrandom manner. These are response-based samples (or choice-based samples). Such samples suffer from at least two biases: choice-based sample bias and sample selection bias. This topic is further elaborated in Chap. 3 on modeling financial distress and bankruptcy.

Conventional wisdom about sampling in empirical corporate finance and accounting is that we accept what we have (as the sample) and keep this in mind during the modeling stages.

Analyzing the Entire Population

Sometimes we have complete data on all units within the population. For example, the Central Statistical Office in Poland has data on the quarterly results of companies that are included in a certain population (e.g., companies that employ at least 50 people). Other examples are the financial statements of all companies listed on a stock exchange, and data on all the corporate clients of a bank. In those cases, the "sample" is, in fact, the "population."

Now, what can be analyzed if we already know everything (i.e., the entire population)? The answer is not straightforward. If we have the "entire population" and we estimate one of its parameters (e.g., the average value of some variable), then the only errors are measurement errors. However, even the simplest regression will include some "error in equation" represented by random disturbance (on the right-hand side of the model). Therefore, inference about the model (i.e., about its parameters) should take this into account.

During an Internet discussion about research on an entire population, the following example was presented. Assume that we are to verify the hypothesis that the average salaries of men and women are equal in a company. If we have data on all the employees, we simply calculate the two averages and state if one is higher than the other. However, researcher–statisticians eagerly assume that there is a data-generating process where men have the mean equal M and women have the mean equal W. They would be delighted to test whether M = W. In fact, such exercise would be incorrect since it would concern some larger population than the employees in the company in question and this (clearly) is not the case here.

What is the lesson for ECF from this example? Simply that we need to know and remember whether our data comprise a sample or represent an entire population. In both cases, the relationships between variables may and should be examined. The interpretations and statistical inference will be different. The model for the entire population tells only about this particular population—i.e., the insignificant variable in the regression equation may mean that, in fact, this particular variable may not be suitable for explaining the relationship in question for this population.

Purposive Samples

A purposive sample is the opposite of having an entire population for research purposes. A purposive sample (also referred to as a subjective or a judgmental sample) is selected in a completely non-probabilistic manner, usually in the social sciences (e.g., in sociological research). A purposive choice of elements for the sample is an expert's choice. The main feature of this choice is purpose (intention).

Elements are selected for such sample using expert (researcher) judgment. An expert believes that the elements chosen are the best, the most representative, and the most useful for the research goal. For example, researchers and the media administer polls of voters in counties that had previous election results very close to the national outcome. The choice of such counties is obviously "purposive." Other examples are studies of cases not typical in the population. Researchers target such cases in order to better explore their "sample." The choice of elements for a purposive "sample" is based on the subjective probability of the expert. Inference from such sample will be restricted to its elements.

It should be noted that, in the case of a purposive sample, statistical tools may include regression (not only descriptive statistics). Regression models assume randomness of the error term—independent of the randomness or the non-randomness of the sample. This means that, for example, inference about the significance of the explanatory variable in a regression is rightfully based on stochastic assumptions.

1.6 ECF and FM: An Introductory Resume

ECF = FM?

As posited in Sect. 1.1, empirical corporate finance and applied accounting are intertwined with financial microeconometrics. FM represents the connection between contemporary corporate finance and the methodology that is appropriate for verifying theories and hypotheses. FM emerges as the most relevant application of econometric–statistical methodology for financial microdata, principally in corporate finance.

The gap between "theory" and "application" or "between Tirole and Eckbo" depends on the statistical–econometric awareness of the researcher. We advocate extensive use of financial microeconometric methodology providing it is feasible in terms of the research strategy and the availability of data. FM is the microeconometrics for finance—i.e., the methods of analyzing microdata as applied to financial topics.

FM is applicable when we examine not just a single company, its financial position, its standing, and its future. FM is useful when we consider a group (sample) of companies and ask questions that are common to the entire sample. This methodology allows for generalization: we may not know how the planned IPO will affect a specific company, but we may assess this issue by using a sample of companies that have recently undergone the IPO experience.

Financial Econometrics and FM

This book is not devoted to microeconometrics per se; instead, the focus here is on its financial applications. It is important to note that microeconometrics—as is the case with other advanced data analytic methodologies—is currently experiencing strong demand from practitioners. This is partly due to the expanding accessibility of large datasets.

The microeconometric model, as presented in Chap. 2, is a regression-type model (equation) that is estimated using microdata. Texts fully dedicated to presenting microeconometric methodology are still rare.[2] Cameron and Trivedi have published two such books: *Microeconometrics* (2005) and *Microeconometrics Using Stata* 2nd ed. (2010). We should also cite *Analysis of Microdata* by Winkelman and Boes (2006) and *Econometric Analysis of Cross-section and Panel Data* by Wooldridge 2nd ed. (2010). Another category are two important works by Angrist and Pischke *Mostly Harmless Econometrics* (2009) and *Mastering 'Metrics* (2015). Microeconometric topics are now included in major econometric textbooks such as Stock and Watson 4th ed. (2019) and Wooldridge 7th ed. (2019).

Financial microeconometrics (FM) is naturally a part of financial econometrics (FE), that is, econometrics dedicated to finance. Historically, FE has specialized in time series, however, some textbooks on financial econometrics rightfully recognize the necessity of including topics on financial microdata analysis (e.g., Wang 2009). Today, a major methodological stream in financial econometrics is still devoted to time series, as in recent textbooks by Bofetti and Urga (2016) and Fan and Yao (2017). Also, the latest handbooks on financial econometrics published by Elsevier (Ait-Sahalia and Hansen 2010) and by Springer (Lee and Lee 2015) contain sets of papers with a focus on time series methodology.

Examples of FM Models

Chapter 2 presents financial microeconometric models from the standpoint of econometric methodology. But let us first complete the discussion in this chapter by presenting the grouping of typical FM models as seen from corporate finance and accounting.

A broader view of FM includes topics from other areas of finance, not only corporate finance and accounting. The models shown in Table 1.1 relate to the fields of corporate finance, corporate governance, accounting, household finance, banking, and investment banking. While the list in Table 1.1 is not complete, it serves to

[2]The *Mikroekonometria* book ed. by Gruszczyński was published in 2010 in Polish (Gruszczyński 2010) 2nd edition in 2012 (Gruszczyński 2012b).

Table 1.1 Examples of models in financial microeconometrics

Aspect of finance	Examples of microeconometric models
Financial management and corporate strategy	Capital and debt structure Choice of financing, strategy, and stakeholders Diversification and value Financial distress models Bankruptcy models Company duration (survival) models Mergers and acquisition models Financial restructuring models
Equity valuation	Regression models for comparative valuation Event studies—market reaction to information regarding – Mergers, splits, dividends – Financial results Models of profit forecasts Models of analytics recommendations
IPO valuation	Event studies for IPO Market reaction to IPO news Models of IPO undervaluation
Financial decisions of companies	Choice models – Choice of debt financing – Company bonds (issue, underwriters) – Company tax policy versus results and debt – Capital structure models Dividend models – Dividend decisions and company value – Dividend policy and stakeholder conflict
Corporate governance	Models of corporate governance constituents and firm results – Ownership structure, institutional/bank ownership – Managerial ownership, family ownership – Composition of the supervisory board and its committees – CEO change – Auditor category, auditor's opinion Models of managerial compensation and incentives Models of corporate governance index Models of M&A as viewed in corporate governance setup
Accounting	Models of external audit decision – Auditor's opinion – Choice of auditor class, auditor independence – Auditor changes in a company Models of accounting standards – For example, choice of R&D expenditure – Types of reporting Financial disclosure models – Models of corporate disclosure index – Compulsory versus voluntary disclosure

(continued)

Table 1.1 (continued)

Aspect of finance	Examples of microeconometric models
Banking and investment banking	Models of lender–borrower relationship – Duration, scope, number, intensity Credit risk models Credit scoring models for individual clients Underwriting costs (shares, bonds)
Financial decisions of households	Models of household decisions – Savings, investment, etc. Models of financial self-assessment

Source: Gruszczyński (2012a) revised

demonstrate the range of microeconometric applications in finance. Each model uses sets (samples) of microdata on companies, transactions, customers, households, etc.

Table 1.1 presents the author's subjective choice and classification of selected microeconometric models in finance. The subjects in the right-hand column differ substantially in terms of the scope and the importance of the topics. It should be noted that the microeconometric model is just one of the methodological approaches that may be of use in resolving a given research problem.

The examples of models listed in Table 1.1 are indeed only examples—the tools of microeconometrics are of use in finance to a much broader extent. In fact, all research hypotheses about the relationship between one category (variable) in finance with others require the use of a statistical–econometric approach for verification, with a regression-type model as the leading choice. Therefore, most research reports in empirical finance contain elements of microeconometrics of time series econometrics.

What Comes Next and What Has Been Omitted?

This introductory chapter has focused on placing financial microeconometrics within the framework of corporate finance, accounting, corporate law, financial management, corporate governance, statistical methods, and financial econometrics. Additional insight on financial microeconometric models is presented in the chapters that follow, beginning with a discussion of issues of methodology. The next chapters also contain surveys of financial microeconometrics in specific areas of corporate finance and accounting.

Chapter 2 is devoted to methodological questions of financial microeconometrics. Various models and methodologies are presented that are effective in appropriate applications in corporate finance and accounting. A distinctive feature of this and other chapters is the inclusion of examples of published models. The final section of Chap. 2 presents a new look at good practices in financial microeconometrics.

The subsequent chapters are dedicated to the choice of specific models presented in Table 1.1 as follows:

- Topics of corporate governance modeling, especially models with corporate governance indices—in Chap. 5.
- Models of financial distress and bankruptcy—in Chap. 3.
- Applications of microeconometrics in accounting research, especially in modeling financial disclosure—in Chap. 4.
- A selection of models of financial microeconometrics applied to equity valuation and value relevance of accounting statements—in Chap. 6.

This text is by no means complete in terms of topics that define all aspects of financial microeconometrics. Some important exclusions are:

- The models of mergers and acquisitions that were, for example, presented in 44 papers published in 1983–2009 and reprinted in two volumes edited by B.E. Eckbo (2010).
- The models of microeconometrics of banking shown, for example, in the book by Degryse et al. (2009).
- The panel models in corporate finance and accounting presented, for example, in papers by Flannery and Hankins (2013) and de Jager (2008).
- The models incorporating event study analysis presented, for example, in Khotari and Warner (2007).

References

Ait-Sahalia Y, Hansen LP (eds) (2010) Handbook of financial econometrics: tools and techniques, Handbooks in finance, vol 1–2. Elsevier, Amsterdam

Amaro de Matos J (2001) Theoretical foundations of corporate finance. Princeton University Press, Princeton, NJ

Angrist JD, Pischke J-S (2009) Mostly harmless econometrics. Princeton University Press, Princeton, NJ

Angrist JD, Pischke J-S (2015) Mastering ‘metrics: the path from cause to effect. Princeton University Press, Princeton, NJ

Bofetti S, Urga G (2016) Financial econometrics using Stata. Stata Press, College Station, TX

Brennan MJ (ed) (2001) Empirical corporate finance, vol I, II, III, IV. Edward Elgar, Cheltenham

Brigham EF, Daves PR (2019) Intermediate financial management, 13th edn. Cengage, Boston, MA

Brigham EF, Gapenski LC (1996) Intermediate financial management, 5th edn. Thomson/South Western, Ohio

Cameron AC, Trivedi PK (2005) Microeconometrics: methods and applications. Cambridge University Press, New York

Cameron AC, Trivedi PK (2010) Microeconometrics using Stata. Revised edition, Stata Press, College Station, TX

Damodaran A (2014) Applied corporate finance, 4th edn. Wiley, Hoboken, NJ

de Jager P (2008) Panel data techniques and accounting research. Meditari Account Res 16(2):53–68

Degryse H, Kim M, Ongena S (2009) Microeconometrics of banking: methods, applications and results. Oxford University Press, Oxford
Eckbo BE (ed) (2007) Handbook of corporate finance: empirical corporate finance, North-Holland handbook of finance series, vol 1. Elsevier, Amsterdam
Eckbo BE (ed) (2008) Handbook of corporate finance: empirical corporate finance, North-Holland handbook of finance series, vol 2. Elsevier, Amsterdam
Eckbo BE (ed) (2010) Corporate takeovers: modern empirical developments, vol. 1: takeover activity, valuation estimates and sources of merger gains, vol. 2: bidding strategies, financing and corporate control. Elsevier/Academic Press, Amsterdam
Eckbo BE, Masulis RW, Norli Ø (2007) Security offerings. In: Eckbo BE (ed) Handbook of corporate finance: empirical corporate finance, North-Holland handbook of finance series, vol 1. Elsevier, Amsterdam, pp 233–374
Fan J, Yao Q (2017) Elements of financial econometrics. Cambridge University Press, Cambridge
Fernandez P (2017) The equity premium in 150 textbooks. Available at SSRN: https://ssrn.com/abstract=1473225, 10 October 2017
Flannery MJ, Hankins KW (2013) Estimating dynamic panel models in corporate finance. J Corp Finan 19:1–19
Grinblatt M, Titman S, Hillier D (2011) Financial markets and corporate strategy. 2nd European edn. McGraw-Hill Higher Education, London
Gruszczyński M (2006) Mikroekonometria finansowa. Zarys problematyki. In: Ronka-Chmielowiec W and Jajuga K (ed) Inwestycje finansowe i ubezpieczenia – tendencje światowe a polski rynek. Prace Naukowe Akademii Ekonomicznej we Wrocławiu, no 1133, Wrocław 2006, pp 111–118
Gruszczyński M (2008) Financial microeconometrics in corporate governance studies. FindEcon. Forecasting financial markets and economic decision-making, no. 6: 11–17. Also available as Gruszczyński M (2010) Financial microeconometrics in corporate governance studies. Working Paper No. 07-10, Department of Applied Econometrics, SGH Warsaw School of Economics
Gruszczyński M (2009), Book review. In: Espen Eckbo B (ed) Handbook of corporate finance. Empirical corporate finance (2007–2008), Bank i Kredyt, vol. 40. Elsevier North-Holland, Amsterdam, pp 109–117
Gruszczyński M (ed) (2010) Mikroekonometria. Wolters Kluwer, Warszawa
Gruszczyński M (2012a) Empiryczne finanse przedsiębiorstw. Mikroekonometria finansowa [Empirical corporate finance. Financial microeconometrics]. Difin, Warszawa
Gruszczyński M (ed) (2012b) Mikroekonometria, 2nd edn. Wolters Kluwer, Warszawa
Gruszczyński M (2018) Financial microeconometrics as research methodology in corporate finance and accounting. In: Dudycz T, Osbert-Pociecha G, Brycz B (eds) Efficiency in business and economics. Springer proceedings in business and economics. Springer, New York, pp 71–80
Heckman JJ (2000) Micro data, heterogeneity and the evaluation of public policy: Nobel lecture. J Polit Econ 8(12):2000
Heckman JJ (2001) Micro data, heterogeneity and the evaluation of public policy: part 1. J Polit Econ 109(4):673–748
Heckman JJ (2004) Micro data, heterogeneity and the evaluation of public policy: part 2. Am Econ 49(1):16–44
Khotari SP, Warner JB (2007) Econometrics of event studies. In: Eckbo BE (ed) Handbook of corporate finance: empirical corporate finance, North-Holland handbook of finance series, vol 1. Elsevier, Amsterdam, pp 3–36
Lee CF, Lee JC (eds) (2015) Handbook of financial econometrics and statistics. Springer, New York
Ross SA, Westerfield RW, Jaffe J, Jordan B (2015) Corporate finance, 11th edn. McGraw Hill Irwin, New York
Stock JH, Watson MW (2019) Introduction to econometrics, 4th edn. Pearson, Harlow
Tirole J (2006) The theory of corporate finance. Princeton University Press, Princeton, NJ

Wang P (2009) Financial econometrics, 2nd edn. Routledge, London
Winkelmann R, Boes S (2006) Analysis of microdata. Springer, Heidelberg
Wooldridge JM (2010) Econometric analysis of cross-section and panel data, 2nd edn. MIT Press, Cambridge, MA
Wooldridge JM (2019) Introductory econometrics: a modern approach, 7th edn. Cengage, Boston, MA

Chapter 2
Models of Financial Microeconometrics

The topics presented in this chapter focus on the examination of both the practical and the theoretical issues relating to the application of econometric techniques in corporate finance and accounting research based on microdata. We introduce a range of microeconometric models and techniques with detailed examples of relevant applications. Emphasis is also given to methodology that may be useful in studying causal effects in corporate finance and accounting. The final section takes a fresh look at good practices in financial microeconometrics—in hope of avoiding unnecessary efforts that may lead to inaccurate results.

2.1 The Types of Models Used in Empirical Corporate Finance and Accounting Research

Types of Microdata

This section introduces several microeconometric models, presenting them with direct reference to corporate finance and accounting. This means emphasizing the applied side rather than the formal-mathematical one. Models in microeconometrics use microdata. The type of microdata, as well as the research question, determine how to find the solution (i.e., what kind of model is best applied). According to a classification by Winkelmann and Boes (2006), microdata may be divided into quantitative or qualitative in the following way:

1. *Quantitative microdata*: Discrete or continuous
   (a) Unrestricted range
   (b) Restricted range: (b1) Limited dependent variables, (b2) durations, (b3) counts

M. Gruszczyński, *Financial Microeconometrics*,
https://doi.org/10.1007/978-3-030-34219-7_2

2. *Qualitative microdata*: Discrete
   (a) Binomial (binary)
   (b) Multinomial (unordered)
   (c) Ordered multinomial

In empirical corporate finance, quantitative microdata is, for example, a company's sales value in the last year while qualitative microdata is, for example, information about a company's CEO change last year ("yes" or "no"). Obviously, this makes sense only when there is a sufficiently large set of companies.

In this book, we primarily present examples based on cross-sectional microdata. However, all data types mentioned above can also be in the form of panel data.

Types of Models

A microeconometric model is typically one equation in which the left-hand side variable is to be explained by several right-hand side variables, with the use of microdata. If the model is intended to explain sales value, then "sales value" is the explained variable (endogenous variable, dependent variable) in the model. When "CEO change" is on the left-hand side, then the variable explained in the model is represented by the dummy variable with "ones" and "zeroes" only, where "1" represents YES (CEO has changed) and "0" represents NO (no change of CEO).

In the examples above, "sales value" represents the *quantitative* explained variable while "CEO change" is the *qualitative* explained variable. For quantitative dependent variables, microeconometrics offers classic regression models, usually linear. Major problems remain the same as in basic econometrics: questions of selecting appropriate explanatory variables, tackling the problems of heteroscedasticity, endogeneity, etc.

If the quantitative dependent variable is *limited*, it is partly qualitative and partly quantitative. An example might be the variable representing a company's dividend payment. This variable takes values "0" or "more than 0" and in a sample of companies the number of "zeroes" might be substantial.

Qualitative dependent variables are the core of microeconometrics. Table 2.1 presents a classification of microeconometric models with respect to the type of dependent variable. Most models represent solutions to explaining qualitative dependent variables.

Modeling Equation

The key concept of the microeconometric model is as follows. Assuming we are considering a one-equation model, the problem lies in finding the appropriate determinants—i.e., the explanatory/exogenous variables (X), which are the most suitable for explaining the endogenous variable (y). This relationship between the X variables and the variable y is present in each microeconometric model, although typically it is not direct. Table 2.2 shows schematically the one-equation model.

Table 2.1 Types of microeconometric models and techniques

(a) Multiple regression	**(c) Binomial models** Linear probability model (LPM) Logit model/probit model Complementary log-log model
(b1) Limited dependent variable models Truncated regression Tobit model Two-limit Tobit Sample selection model (Heckman)	**(d) Multinomial unordered variables models** Multinomial logit and probit models Conditional logit (McFadden) Nested logit Mixed logit
(b2) Count data models Poisson regression Negative binomial regression	**(e) Multinomial ordered variables models** Ordered logit and probit models Generalized ordered models Sequential models
(b3) Duration models	**(f) Treatment effects models and other quasi-experimental techniques**

Source: Gruszczyński (2012a) revised

Table 2.2. The single-equation microeconometric model: schematic view

Left-hand side of the model	=	Right-hand side of the model
explained variable: y	is explained by	explanatory variables X
might also be: function of y, probability		accompanied by: random error
		X variables may form a vector $\boldsymbol{x}$

Source: Author

Thus, on the right-hand side of the model, we have the variables X. They are represented in the form of the linear combination $\boldsymbol{x}_i'\boldsymbol{\beta}$. For k explanatory variables X and the constant term, this linear combination is as follows[1]:

$$\begin{aligned}\boldsymbol{x}_i'\boldsymbol{\beta} &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \ldots + \beta_k X_{ki} \\ &= (1\; X_{1i}\; X_{2i} \ldots X_{ki}) \cdot (\beta_0\; \beta_1\; \beta_2 \ldots \beta_k)'\end{aligned} \tag{2.1}$$

[1]In all the formulae throughout this book, small letters in bold denote column vectors. Vectors in Eq. (2.1) are $\boldsymbol{x}_i = \begin{bmatrix} 1 \\ X_{1i} \\ X_{2i} \\ \vdots \\ X_{ki} \end{bmatrix}$ and $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$. Each vector here has the dimension $(k + 1) \times 1$. This means it has $k + 1$ rows and 1 column. The product $\boldsymbol{x}_i'\boldsymbol{\beta}$ has dimension 1×1 and is called *scalar product* since its result is a single number (*scalar*).

where $i = 1, 2, \ldots, n$. Here and further down, the index i denotes the observation number and n is the number of observations (sample size). In Eq. (2.1), there are k explanatory variables $X_1, X_2, \ldots, X_k$ and $k + 1$ parameters (coefficients) $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$.

On the left-hand side of the model, we may have a variable y directly or a transformation of variable y (e.g., probability that $y = 1$).

The variables X on the right-hand side of the equation represent determinants of y (or a function of y). It is obvious that we cannot name (or find or measure) all possible determinants. Therefore, we also include on the right-hand side a random variable called the *error term* (disturbance, error in the equation) customarily denoted by ε.

To sum up

- The explained (endogenous) variable is y or a function of y.
- The explanatory (exogenous) variables are $X_1, X_2, \ldots, X_k$ and enter the model on the right-hand side in the form of a linear combination (Eq. 2.1), together with the error term.

Modeling Strategy

The modeling strategy amounts to selecting the explanatory variables $X_1, X_2, \ldots, X_k$ in such a way that they appropriately explain the y variable (or a function of y). One of the principles to follow is the minimization of the multicollinearity effect. Multicollinearity means high mutual correlation between explanatory variables. This effect may occur when the information from the various X variables is practically the same. This may mean that there is no need to include in the model three or more variables representing the same subject (e.g., the profitability indices). In the sample they are commonly highly correlated with each other, therefore, just one of them may suffice as the explanatory variable X in the model.

Other good practices in finding suitable explanatory variables for the model are shown later in this chapter.

Tables 2.3 and 2.4 show examples of microeconometric models applied in the domain of corporate finance and accounting. The left-hand column in each table presents the object of modeling—the endogenous variable y or the specific probability of y. The right-hand column displays the list of variables $X_1, X_2, \ldots, X_k$ that are explanatory in the model.

The topics identified in the examples presented in Tables 2.3 and 2.4 are very diverse. There is a model of voting "yes" or "not-yes" by an institutional investor in the case of a firm's acquisition in the USA, a dividend model for Jordanian companies, and an interesting model of firm duration in Spain, among others. The common denominator of these models is large sets of microdata. Methodologically, the models differ but, in terms of the field of application, they all belong to financial microeconometrics.

Table 2.3 Examples of financial microeconometric models: Linear probability, binomial logit, multinomial ordered probit, and multinomial unordered logit

Linear probability model	Matvos and Ostrovsky (2008)
The probability of voting for the merger from family cross ownership: $\boldsymbol{y} = 1$ if the vote in the acquirer is "for", $\boldsymbol{y} = 0$ otherwise ($n = 6369$ mutual fund votes in acquirer's shareholders meetings in 114 completed mergers and acquisitions of US public companies, 2003–2006).	$\boldsymbol{X} =$ holdings in the target ($= 1$ if the fund holds shares in the target, $= 0$ otherwise); family and cross ($= 1$ if the fund holds shares in the target and some other fund in the same family holds shares in the target as well, $= 0$ otherwise); family and no cross ($= 1$ if the fund does not hold shares but some other fund in the same family does, $= 0$ otherwise)
Binomial logit model	Moeller et al. (2004)
Probability of bidder success: $\boldsymbol{y} = 1$ if the bid is classified by SDC as successful, $\boldsymbol{y} = 0$ otherwise ($n = 1761$ bids, all offers by US firms to acquire publicly traded US firms 1980–2001 as listed by SDC whose transaction value is at least 1 million USD and 1% of acquirer's market value)	$\boldsymbol{X} =$ premium (aggregate value of cash, stock and other securities offered by the bidder to the target shareholders divided by the market value of equity of the target 50 days prior to the takeover announcement); log of market value of equity of the target and of the bidder, toehold ($= 1$ if the acquirer holds at least 5% of the target shares, $= 0$ otherwise; all cash ($= 1$ if only cash is used, $= 0$ otherwise); Tobin's q for the bidder and for the target; hostile ($= 1$ if acquisition is hostile according to SDC, $= 0$ otherwise); and other variables
Multinomial ordered probit model	Kamstra et al. (2001)
Probability of firm's bond rating (Moody's): $\boldsymbol{y} = 0$ for B, $\boldsymbol{y} = 1$ for Ba, $\boldsymbol{y} = 2$ for Baa, $\boldsymbol{y} = 3$ for A, $\boldsymbol{y} = 4$ for Aa, $\boldsymbol{y} = 5$ for Aaa ($n = 265$ new US industrial bonds issued in 1993).	$\boldsymbol{X} =$ interest coverage (net income plus interest expense, divided by interest expense); debt ratio (total debts divided by total assets); return on assets (ROA); total assets; subordination status ($= 1$ if debt issue has seniority, $= 0$ otherwise)
Multinomial unordered logit model	Hensher and Jones (2008)
Probability of the firm being in one of four states of financial distress: $\boldsymbol{y} = 0$ non-failed firm, $\boldsymbol{y} = 1$ insolvent firm, $\boldsymbol{y} = 2$ financially distressed firm that was subject to a merger or takeover arrangement, $\boldsymbol{y} = 3$ firm that filed for bankruptcy followed by the appointment of liquidators ($n = 2259$ firm-years, Australia 1992–2004)	$\boldsymbol{X} =$ excess market return (above market return); (cash + deposits + marketable securities)/total assets; 4 consecutive annual periods of negative operating cash flow ($1 =$ yes, $= 0$ otherwise); EBIT/total assets; working capital/total assets; log of total assets; age of firm ($= 1$ if the firm was established in the previous 6 years,$= 0$ otherwise); total debt/gross operating cash flow and other variables

Source: Gruszczyński (2012a); Matvos and Ostrovsky (2008); Moeller et al. (2004); Kamstra et al. (2001); Hensher and Jones (2008)

In the following sections of this chapter, we present several of these models in greater detail, together with more examples from corporate finance and accounting. As indicated in Chap. 1, our aim is not the exposure of microeconometrics that can be found in textbooks like Cameron and Trivedi (2005), Winkelmann and Boes (2006), and Gruszczyński (2012b). In particular, specific questions of statistical testing and related issues are not discussed.

Table 2.4 Examples of financial microeconometric models: Tobit, duration model, and multiple regression

Tobit model	Al-Malkawi (2007)
Dividend yield: ***y***—dividend-to-price ratio; for public companies in Amman ($n = 1511$ observations, where $\boldsymbol{y} = 0$ for 853 observations; balanced panel, Jordan 1989–2000)	***X*** = percentage held by insiders, family dummy (= 1 if firm is family owned, = 0 otherwise); state dummy (= 1 if firm is owned by government agencies, = 0 otherwise); age of the firm; age squared; debt-to-equity ratio; natural log of market capitalization; dummy for industry effects (= 1 if firm belongs to nonfinancial sector, = 0 otherwise); after-tax earnings per share
Duration model	Esteve-Pérez and Mañez-Castillejo (2008)
Model of firm duration: The probability that a company exists in year *t* assuming it existed in year *t*-1 (hazard function model) ($n = 14{,}193$ observations, 2028 industrial firms, Spain 1990–2000)	***X*** = firm size (= 1 for firms with more than 200 employees, = 0 otherwise); advertising expenditures (1 = yes, 0 = no); R&D strategy (three categories); industry technological intensity (three categories); export intensity (three categories); productivity (three categories); limited liability company (1 = yes, 0 = no); and other variables
Multiple regression	Beekes et al. (2007)
Model of financial disclosure by a firm: ***y***—natural log of the number of documents released by the company over the 250 trading days ending ten trading days after the company's fourth quarter earnings ($n = 216$ companies, Canada 2004).	***X*** = Corporate governance rating (BSCI rating) for the company; firm size (natural log of firm's market value of equity); good news (= 1 when the company's share price outperforms the market over the 250-trading-day period, = 0 otherwise)

Source: Gruszczyński (2012a); Al-Malkawi (2007); Esteve-Pérez and Mañez-Castillejo (2008); Beekes et al. (2007)

2.2 The Binomial Model: An Auditor Change After a Going-Concern-Modified Audit Opinion in Australia

Outline

In the binomial model, the endogenous variable *y* takes two possible values (two states, two categories, two answers). The variable *y* represents choice, decision, "state of affairs." Here are examples:

- Last year, a company went public having decided to carry out an IPO (yes–no)
- An auditor gives a going concern opinion for a company (yes–no)
- A company has directors (i.e., members of the supervisory board) from academia (yes–no)
- A firm experienced CEO turnover last year (yes–no)
- The first bidder for the acquisition of a firm is successful (yes–no)

It is convenient to express "yes–no" situations by two numbers "1" and "0." One can ask, what is the probability that $y = 1$ or $y = 0$ ("yes" or "no")? This question is typical for qualitative dependent variables and is not very feasible for quantitative ones.

The binomial model explains the probabilities that $y = 1$ or $y = 0$ by relating them to the X variables. Let us discuss briefly the probabilities $P(y = 1)$ and $P(y = 0)$. This consideration introduces variable y as a random variable because, in the population and in the sample, it may take either of two values with given probabilities. Assume that probability $P(y = 1) = p$ and that probability $P(y = 0) = 1 - p$.

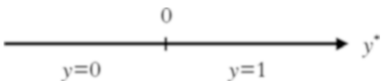
Thus, variable y has two possible values with the probabilities just specified. This means that y has a rather simple probability distribution: two distinct values (0 and 1) with associated probabilities (p and $1 - p$). Such distribution is called Bernoulli's distribution.

Knowing this, it is not difficult to show that the expected value of this random variable y is equal to $E(y) = p$ and the variance of y is equal to $Var(y) = p(1 - p)$.

Table 2.5 presents the description of the binomial model with the use of the latent variable y^* representing the tendency (propensity) to the decision (choice, state) $y = 1$. The variable y^* is continuous and unobservable. When the value of y^* reaches some point (called the *cutpoint*), then what we observe is $y = 1$. If not, then $y = 0$. The propensity y^* can represent, for example, the willingness of the company to decide to go with an IPO this year, the inclination of a company board to pay a dividend last year, etc.

Thus, we think of the binomial model as representing the unknown latent variable y_i^*, which is continuous and is related to variables X. Such relationship is typically linear, as follows:

Table 2.5 The binomial model

Characteristics	What is modeled?
Variable y takes two values $y = 1$ and $y = 0$. Values represent states, categories, etc. *Proposition*: There is an unobserved continuous variable y^* representing the propensity for y to surface as $y = 1$ (e.g., the inclination of company governing bodies to pay a dividend for last year). Assume that if $y^* \geq 0$, then the dividend is paid, and if $y^* < 0$, then it is not paid. But what we observe are only values of y, the binary variable. 	Probability $p = P(y^* \geq 0)$, i.e., $p = P(y = 1)$ and $1 - p = P(y^* < 0)$, i.e., $1 - p = P(y = 0)$. In the binomial model, the probability p is modeled as the function (usually nonlinear) of explanatory variables X.

Source: Gruszczyński (2012a)

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \tag{2.2}$$

In Eq. (2.2) the index i indicates the observation number ($i = 1, 2, \ldots, n$) and n is the number of observations. The expression $\boldsymbol{x}_i'\boldsymbol{\beta}$ is the linear combination of the X variables introduced in Eq. (2.1). The new element is ε_i, the stochastic disturbance term, typically present in regression equations and representing "error in equation." It is usually assumed that its expected value is 0 and its variance is constant over "i."

As previously mentioned, since we do not know y_i^*, we use y_i and the binomial model instead. The binomial model for y_i is the relationship between the probability that $y_i = 1$ and the explanatory variables X. This equation may be written as

$$p_i = F\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right) \tag{2.3}$$

for $i = 1, 2, \ldots, n$. Function F in Eq. (2.3) is the specific function that distinguishes between the types of binomial models. Most popular are *logit*, *probit,* and the *linear probability model* (LPM). Logit model is also known as *logistic model*.

The error term ε_i from Eq. (2.2) is also present in Eq. (2.3) but is omitted here. Its stochastic characteristics depend on a type of F function. Function F is the cumulative distribution function: of logistic distribution for the logit model, and of normal distribution for the probit model. For the linear probability model, we have

$$F\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right) = \boldsymbol{x}_i'\boldsymbol{\beta} \tag{2.4}$$

Obviously, probability p_i is the number from the interval <0,1>, therefore, we should expect also that $0 \leq F\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right) \leq 1$. This holds for logit and probit but not always for the LPM. For large samples, however, the LPM performs equally as well as logit and probit.

The Logit Model

In the logit model, the relationship between p_i and $\boldsymbol{x}_i'\boldsymbol{\beta}$ can be conveniently expressed as

$$p_i = F\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right) = \frac{\exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)}{1 + \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)} = \frac{1}{1 + \exp\left(-\boldsymbol{x}_i'\boldsymbol{\beta}\right)} \tag{2.5}$$

Therefore,

$$\frac{p_i}{1 - p_i} = \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right) \text{ and } \ln \frac{p_i}{1 - p_i} = \boldsymbol{x}_i'\boldsymbol{\beta}$$

The expression $\ln \frac{p_i}{1-p_i}$ is called *logit*.

Therefore, the logit model is the relationship between the logit (explained) and the X variables (explanatory) or their linear combination $\boldsymbol{x}_i'\boldsymbol{\beta}$:

$$\text{logit}\, p_i = \boldsymbol{x}_i'\boldsymbol{\beta} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \ldots + \beta_k X_{ki} \tag{2.6}$$

Let us recall that the ratio of p_i and $1–p_i$ is called *odds*

$$\frac{p_i}{1 - p_i} = odds$$

Logit is the logarithm of odds. Therefore,

$$p_i = \frac{odds}{1 + odds}$$

Table 2.6 sums up the definitions of odds, logit, and the logit model.

Estimation

The parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ of the logit model (Eq. 2.3) are to be estimated from the sample of n observations on the variables y_i,X_{1i}, X_{2i}, . . ., X_{ki} ($i = 1, 2, \ldots, n$). The logit model as well as the probit model are commonly estimated using the maximum likelihood (ML) method.

The estimation results indicate the statistical quality and the validity of the model in verifying the underlying research hypothesis. There are a few specific items to check and, possibly, an iterative process of improving the model (e.g., by choosing the "best" subset of X variables). Some of these questions are discussed later in this chapter.

With the satisfactory statistical checks (described below), one important task is to check if the signs of the parameter estimates are in accord with the theoretical (research) expectations. Why? The sign of the estimate of parameter β_j for variable X_j is the same as the sign of the association between X_j and the probability p_i that $y_i = 1$:

- For positive β_j, the increase of X_j is associated with the increase of probability p_i and the decrease of X_j is associated with a decrease of p_i.
- For negative β_j, the increase of X_j is associated with the decrease of probability p_i and the decrease of X_j is associated with the increase of p_i.

Table 2.6 Probability, odds, logit, and the logit model

Probability	p_i	$P(y_i = 1)$
Odds	$odds = \frac{p_i}{1-p_i}$	$p_i = \frac{odds}{1+odds}$
Logit	$\text{logit}\, p_i = \ln \frac{p_i}{1-p_i}$	$\text{logit}\, p_i = \boldsymbol{x}_i'\boldsymbol{\beta}$

Source: Author

Now, assume that one of our research hypotheses is as follows: having directors (members of the supervisory board) from academia (variable y) is positively associated with the governance level (variable X_j). Assume that the variable X_j is placed among other explanatory variables in the logit model explaining variable y. Then this hypothesis translates into the following statement: the sign of the estimate of parameter β_j (standing by the X_j variable) is positive. What remains is to ensure that the X_j variable appears among the explanatory variables in your final model and to check whether the estimate is significantly positive.

The Marginal Effect and the Odds Ratio

A similar refined question is how large would the effect on p_i be of increasing X_j by one unit, ceteris paribus (i.e., without changing other X variables). The answer is called the *marginal effect* (ME) of variable X_j. For the logit model, marginal effects vary across observations. One possibility is to calculate the ME for some representative values of X variables, usually for the mean values. The more general result is the ratio of the MEs for two variables X_j and X_l. It turns out that the ratio of the MEs is constant over the entire sample and also over various models (logit, probit, LPM). This ratio is equal to β_j/β_l.

One more interpretation can be derived from the estimated logit model. It is called the *odds ratio* (OR) for the variable X_j. It is the ratio of odds with the X_j variable changed by one unit ($X_j + 1$) to the odds without this change. The OR is equal to $\exp(\beta_j)$. For $\beta_j > 0$, the OR is larger than one and we could say the odds are $\exp(\beta_j)$ times larger for $X_j + 1$ than for X_j. Similarly, for $\beta_j < 0$, the odds are $\exp(\beta_j)$ times smaller.

So, the ME tells us what is happening with the probability $P(y_i = 1)$ for the unit change in X_j (holding all other variables constant) and the OR tells the same thing about the odds of $y_i = 1$.

Statistical Validity

Judging the statistical validity of the estimated logit model requires several assessments:

- The statistical significance of the explanatory variables (z test) and of the entire model (LR = likelihood ratio test).
- An acceptable *goodness-of-fit measure* like the *pseudo-R-squared* (or *pseudo-R2*), in relation to the other competitive specifications of the model.
- An acceptable level of prediction accuracy measured, for example, by using a *classification table* such as the following:

Classified (predicted)/true	True $y_i = 1$	True $y_i = 0$
Predicted $y_i = 1$	n_1	n_2
Predicted $y_i = 0$	n_3	n_4

where n_1, n_2, n_3, n_4 are the numbers of cases (observations) representing one of four possible situations. Obviously, $n_1 + n_2 + n_3 + n_4 = n$ and, since we are interested in predictions of “ones” and “zeroes,” the measure *count-R2* $= (n_1 + n_4)/n$ represents the model’s prediction accuracy.

The question remains how to classify the model’s forecast: if p_i predicted from the model is equal to or is higher than the share of ones in the sample, then we predict $y_i = 1$—this is Cramer’s rule (Cramer 1999). Another popular *cut-off value* is 0.5, which may not be appropriate when the sample is highly imbalanced (numbers of “zeroes” and “ones” for the y variable in the sample differ significantly). In Chap. 3, we discuss this question in more detail.

A high value of AUC (i.e., the *Area Under Curve*) is an indication of the model’s good prediction accuracy. The curve is called ROC (i.e., the *Receiver Operating Characteristic*). It connects classifications resulting from all cut-off values between 0 and 1 and is obtained by plotting A against B, where A is the probability of predicting $y_i = 1$ for an actual $y_i = 1$, and B is the probability of predicting $y_i = 1$ for an actual $y_i = 0$ [Using the notation above, we have $A = n_1/(n_1 + n_3)$ and $B = n_2/(n_2 + n_4)$]. In other words, the ROC graphs the true positive rate against the false positive rate; the AUC value lies in the interval $<0,1>$; the higher the value of AUC, the better the fit of the model.

More details on estimating and verifying binomial models (logit, probit, LPM) are to be found in textbooks on microeconometrics (e.g., Cameron and Trivedi 2005; Long 1997; Gruszczyński 2012b).

In Example 2.1, the binomial logit model is applied to the sample of Australian companies considered from the standpoint of the decision regarding the auditors examining the companies’ books.

Example 2.1 A Change of Auditor After a Going-Concern-Modified Audit Opinion: Australian Audit Market

This example considers an important aspect of corporate governance i.e., a change of auditor after obtaining a going-concern-modified audit opinion. Obviously, companies are not fond of such opinions and, in response, sometimes replace the auditor. One hypothesis studied in the paper by Carey et al. (2008) is that “companies receiving first-time going-concern-modified audit opinions are more likely to switch auditors than similarly stressed companies not receiving a going-concern-modified audit opinion.”

Table 2.7 presents the results of the estimation of the logit model, which show that an auditor switch is positively associated with receiving a first-time going-concern-modified audit opinion. The probability of company failure—calculated from the model by Zmijewski (1984)—is also positively related to an auditor switch. The Big 5 auditor firms—it was the nineties, today we have the Big 4 or the “Big N”—have less chance for auditor switch than other audit firms.

Table 2.7 Auditor change after going-concern-modified opinion: Australian companies 1994–1998: binomial logit estimation results

Explained variable: logit (the probability of AUDCHG); AUDCHG = 1 if (after going-concern-modified opinion) the client subsequently switches auditor, = 0 otherwise	
Explanatory variables:	Estimates (p)
PRB = probability of failure (Zmijewski 1984)	1.42 (0.00)
LNTA = log of total assets	−0.40 (0.00)
BIG5 = 1 if audited by a BIG 5 firm, = 0 otherwise	−0.46 (0.17)
GC = 1 if audit report was first-time going-concern-modified, = 0 otherwise	1.00 (0.01)
Constant	0.68 (0.44)

$n = 112$, Pseudo $R^2 = 0.203$
Source: Carey et al. (2008)

In another part of their paper, Carey et al. (2008) show that "lost audit fee revenue from client switching after the issuance of a first-time going-concern-modified audit opinion is greater than lost audit fee revenue from client switching for similarly stressed companies that had not received a going-concern-modified audit opinion." ■

Comment

Table 2.7 presents estimates of the parameters of the binomial logit model. Following the line of reasoning presented before this example, it can be seen that the positive (negative) estimate of the parameter β_j means a positive (negative) association of the variable X_{ji} and the probability p_i which is the probability of an auditor change. This is measured by the marginal effect (ME) which is the derivative of p_i with respect to X_{ji}. The unit increase of X_{ji} is associated with the increase of p_i by the ME.

The goodness-of-fit measure in Table 2.7 is *pseudo-R2*. This is of practical value when comparing this specification (the set of X variables) with alternative ones. Other measures of classification and fit are not included here.

To conclude this exposition about the binomial model, it should be noted that the popular model of *linear multivariate discriminant analysis* (LDA) may be considered as a special case of the linear probability model (LPM). This has been shown by, for example, Maddala (1983).

2.3 Practical Use of the Binomial Logit: Prior Correction

Applying the binomial model to real data typically results in the number of observations with $y_i = 1$ being significantly different than the number of observations with $y_i = 0$. In a large population of companies for a single year, only a few go bankrupt, a few are the object of acquisition or merger, a few change the CEO, etc. So, the population is “unbalanced” in terms of the endogenous variable y_i.

It is customary for modeling to use a balanced sample where the “ones” and “zeroes” are equally frequent. There are several reasons for this. The primary rationale is the accessibility of data for an entire population. What we usually have are the observations for an infrequent category—e.g., for bankrupt companies: in Poland, the number of companies declaring bankruptcy annually amounts to 0.1% of the number of all commercial companies. A random sample of that population is not feasible. Therefore, the infrequent observations—bankrupt companies—(e.g., with $y_i = 1$) are complemented by frequent ones—non-bankrupt—with $y_i = 0$. Various forms of sampling the “zeroes” or matching “zeroes” to “ones” are possible. In effect, we obtain a balanced sample (“50–50 sample”).

It may be argued that balanced samples give a better estimation outcome since the infrequent category of the y_i variable is represented on equal terms with the frequent one: “we select the sample on the dependent variable to learn more about rarer cases than a random sample would be able to tell us” (unknown author in a discussion on stats.stackexchange.com). This is also because the features of the rarer category, represented by the X variables, are more distinctly present when the sample is balanced.

However, there is a price for using a balanced sample. A model estimated on a balanced sample does not correspond directly to the population. Luckily, this problem may be solved when the binomial model is logit. The binomial logit estimated on a sample can be converted into one for the population. It is possible when we know the proportions of the observations (e.g., companies) selected for the sample (Anderson, 1972; Maddala, 1983; Gruszczyński, 2012b, 2017, 2019).

Let us assume that $y_i = 1$ means a bankrupt company, and $y_i = 0$ a non-bankrupt one and that we know the proportions of the companies selected for the sample from both groups: pr_1 for companies with $y_i = 1$ and pr_2 for companies with $y_i = 1$. Then, after estimation of the logit model, the intercept should be diminished thereby: $\delta = \ln(pr_1) - \ln(pr_2)$. This correction is equal to zero, when proportions pr_1 and pr_2 are identical (e.g., when a random sample from both groups is used). The correction is called *prior correction* (King and Zeng 2001) or the *Anderson-Maddala correction*.

The Anderson-Maddala correction coincides with the formula by Skogsvik and Skogsvik (2013) as indicated in Gruszczyński (2019). The Skogsviks’ equation helps to determine the relationship between the biased bankruptcy probability of a given company (from the model—*sample based*), and the unbiased probability resulting from the proportion of bankrupts in the population. More discussion on this topic is presented in Chap. 3 on bankruptcy and financial distress.

2.4 Multinomial Ordered Variables Model: The Security Choice by US Companies

When an endogenous qualitative variable has more than two categories (states), we say it is multinomial. Multinomial variables can be ordered or unordered. This section is dedicated to ordered multinomial variables. The model itself may be termed multinomial ordered variables model, ordered response model, ordered multinomial model, or multinomial ordered model.

The qualitative variable with ordered or ordinal categories (outcomes) is, in some sense, similar to a quantitative variable: categories can be ranked from lowest to highest, like numbers. However, the distances between two neighboring categories are not known, which is not true in the case of numbers. The modeling outcomes are similar: the model for an ordinal variable is just one equation and its parameters are conveniently interpretable (Gruszczyński 2001, 2006, 2007).

The following are examples of ordered variables which may be modeled in corporate finance and accounting:

- Corporate bond rating: *Junk bond* ($y = 1$), *low-grade bond* ($y = 2$), *investment-grade bond* ($y = 3$)
- Credit risk of a client: Very low ($y = 1$), low ($y = 2$), medium ($y = 3$), high ($y = 4$), very high ($y = 5$)
- Company's financial distress: Low ($y = 1$), medium ($y = 2$), high ($y = 3$)
- Corporate governance rating class for a company[2]: C+, B–, B, B+, A–.

Other examples may be found in Gruszczyński (2001, 2008) and in later chapters of this book.

The multinomial ordered model, like the binomial, may be interpreted with the use of an unobserved latent continuous variable y_i^*, as in this example with three categories:

$$y_i = 1 \ \textit{if} \ -\infty \leq y_i^* < \tau_1$$

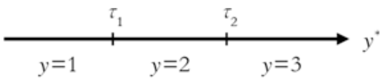
$$y_i = 2 \ \textit{if} \ \tau_1 \leq y_i^* < \tau_2$$

$$y_i = 3 \ \textit{if} \ \tau_2 \leq y_i^* < \infty$$

The τ_1 and τ_2 values of variable y_i^* are called *cutpoints*. We believe that variable y_i^*, if known, represents the inclination (tendency) to attain higher values of variable y_i. For example, variable y_i^* could be a bank's propensity to classify the i-th client into a risk category—the higher the propensity, the higher the category. Table 2.8 presents the characteristic of the ordered multinomial model using the latent variable assumption.

[2]Rating of Polish Corporate Governance Forum, 2004. For more information on corporate governance rankings and ratings, see Sect. 5.3.

Table 2.8 The ordered multinomial model

Characteristics	What is modeled?
Variable y has several categories (states) which can be ordered—e.g., $y = 1$ (company is in good financial condition), $y = 2$ (company is financially distressed) $y = 3$ (company is bankrupt). *Proposition*: There is an unobserved continuous variable y^*= "propensity to go bankrupt." On the horizontal y^*axis there are two points τ_1 and τ_2, such that $\tau_1 < \tau_2$. We presume that: $y = 1$ for $y^* < \tau_1$ $y = 2$ for $\tau_1 \leq y^* < \tau_2$ $y = 3$ for $y^* \geq \tau_2$ 	Probabilities: $p_1 = P(y^* < \tau_1)$ i. e. , $p_1 = P(y = 1)$ $p_2 = P(\tau_1 \leq y^* < \tau_2)$ i. e. , $p_2 = P(y = 2)$ $p_3 = P(y^* \geq \tau_2)$ i. e. , $p_3 = P(y = 3)$ and $p_1 + p_2 + p_3 = 1$ In the ordered multinomial model, the probabilities p_m (here $m = 1, 2, 3$) are modeled as nonlinear functions of the explanatory variables X. For each category m, the parameters of X variables are the same (this is the parallel regression assumption).

Source: Gruszczyński (2012a)

The linear model for variable y_i^* is the same as in Eq. (2.2).

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$$

but now we have more than one cutpoint on the y_i^* axis (for the binomial model, the customary cutpoint is $y_i^* = 0$). In the trinomial ordered model, there are two cutpoints (τ_1 and τ_2), as in the example above.

For the i-th observation (i-th company, i-th transaction, etc., $i = 1, 2, \ldots, n$), the probability that the y_i variable is in the m-th category is equal to

$$p_{im} = P\left(y_i = m | \boldsymbol{x}_i'\boldsymbol{\beta}, \boldsymbol{\beta}, \boldsymbol{\tau}\right) = F\left(\tau_m - \boldsymbol{x}_i'\boldsymbol{\beta}\right) - F\left(\tau_{m-1} - \boldsymbol{x}_i'\boldsymbol{\beta}\right) \tag{2.7}$$

where $m = 1, \ldots, J$ (J is the number of categories for the y_i variable) and F is the cumulative distribution function (CDF). If F is the CDF of logistic distribution, the model is the multinomial ordered logit where

$$F\left(\tau_m - \boldsymbol{x}_i'\boldsymbol{\beta}\right) = \frac{\exp\left(\tau_m - \boldsymbol{x}_i'\boldsymbol{\beta}\right)}{1 + \exp\left(\tau_m - \boldsymbol{x}_i'\boldsymbol{\beta}\right)} \tag{2.8}$$

and $F(\tau_0) = 0$ and $F(\tau_J) = 1$. For the trinomial example, we have $J = 3$ and

$$P(y_i = 1 | x_i) = \frac{\exp\left(\tau_1 - \boldsymbol{x}_i'\boldsymbol{\beta}\right)}{1 + \exp\left(\tau_1 - \boldsymbol{x}_i'\boldsymbol{\beta}\right)} = A \tag{2.9}$$

$$P(y_i = 2|x_i) = \frac{\exp\left(\tau_2 - \boldsymbol{x}_i'\boldsymbol{\beta}\right)}{1 + \exp\left(\tau_2 - \boldsymbol{x}_i'\boldsymbol{\beta}\right)} - \frac{\exp\left(\tau_1 - \boldsymbol{x}_i'\boldsymbol{\beta}\right)}{1 + \exp\left(\tau_1 - \boldsymbol{x}_i'\boldsymbol{\beta}\right)} = B - A \tag{2.10}$$

$$P(y_i = 3|x_i) = 1 - \frac{\exp\left(\tau_2 - \boldsymbol{x}_i'\boldsymbol{\beta}\right)}{1 + \exp\left(\tau_2 - \boldsymbol{x}_i'\boldsymbol{\beta}\right)} = 1 - B \tag{2.11}$$

Parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ and $\tau_1, \tau_2, \ldots, \tau_{m-1}$ of the multinomial ordered logit are estimated using the method of maximum likelihood. It is customarily assumed that one of the coefficients (e.g., β_0) is equal to 0, to assure the model's identifiability.

The multinomial ordered model answers the same questions as the binomial, with the capacity to encompass a broader set of categories. As above, we try to find a reasonable set of X variables for explaining variable y_i, and also to obtain a fair classification result, acceptable goodness-of-fit measures, etc.

On the other hand, there is a specific issue to check regarding the estimated ordered model. It is the assumption of "parallel regressions" which is embedded in the model's setup. This means that for each category of the y_i variable, the model has the same parameter values. Perhaps, the parameters for the categories differ. This should be verified statistically after estimation. If this assumption is not supported by the data, then the model must be modified—typically in the direction of solutions for unordered models where different parameter values for different categories are admissible.

Another modification arises when we use sequential data. The ordered multinomial sequential model describes the y_i^m variable as belonging to one of J categories ($m = 1, 2, \ldots, J$), which are sequentially ordered. What counts is the transition to the next category ($m + 1$). If the process continues (i.e., there is transition), then $y_i^m = 0$, if not then $y_i^m = 1$. These and other topics are described in specialized textbooks on microeconometrics (e.g., Winkelmann and Boes 2006; Gruszczyński 2012b).

Example 2.2 The Security Choice by US Companies

Erel et al. (2009) applied the multinomial ordered logit to the relationship between the type of securities issued by companies and a set of variables representing market and firm-specific characteristics. The paper appeared at a time of the financial crisis, which dramatically affected the financing of US companies.

The ordered endogenous variable y is selected because it represents the gradual level of sensitivity of the debt issued by the companies: "we expect that during poor financial conditions, firms will, at the margin, be more likely to issue less information-sensitive securities than during good financial conditions." The authors claim that, during a recession, or periods of low economic growth and/or shaky capital markets, firms are more willing to issue debt which is less sensitive to external information. In hard times firms prefer convertible bonds to issuing equity and rely on internal financing rather than private debt, etc.

The explained variable y is defined as follows:

$y = 0$ when a firm is financed by internally generated funds (no external debt)

$y = 1$ private debt (bank loan)
$y = 2$ public bond
$y = 3$ convertible bond
$y = 4$ equity offering

It is a multinomial ordered variable with five categories. The model includes four cutpoints τ_1, τ_2, τ_3, τ_4. The underlying y_i^* latent variable may be understood as the firm's inclination to issue debt as measured by information sensitivity: lower in poor market financial conditions and higher in good conditions.

The explanatory variables in the model are the financial characteristics of companies as well as the ratios representing dynamics like sales growth and stock return. The authors' prediction is that the explained variable is "positively related with market-wide conditions so that a recession, a period of low growth, or tight capital markets should be negatively related to this variable." This model—one of many presented in the paper—appears to confirm the hypothesis that the sign (of the estimate) by the "low growth" variable is negative, meaning that "if there is a low growth, it becomes less likely that there is any security issue at all, and if there is an issue, it is likely to be a less information-sensitive one."

The regression equation presented in Table 2.9 includes industry fixed effects (dummies). The table does not show estimates of the parameters τ_1, τ_2, τ_3, τ_4 (the

Table 2.9 The security choice by US companies 1988–2007: multinomial ordered logit estimation results

Explained variable: The multinomial ordered variable $\boldsymbol{y} = 0$ for no external debt, $\boldsymbol{y} = 1$ for bank loan, $\boldsymbol{y} = 2$ for public bond, $\boldsymbol{y} = 3$ for convertible debt, $\boldsymbol{y} = 4$ for seasoned equity offering.	
Explanatory variables	Coeff. Estimates (significance)
Firm age	−0.005
Ln (total assets)	0.224
Leverage ratio	0.476∗∗∗
Market-to-book ratio	0.033∗∗∗
Fixed-assets ratio	−0.088
Cash flow	−0.022
Cash	−0.862∗∗∗
Inverse interest coverage = ln (1 + interest/EBIT)	−0.041∗∗∗
Debt rating dummy	0.562∗∗∗
Sales growth	0.360∗∗∗
Stock return (over previous 12 months)	0.162∗∗∗
Term spread (difference between yields on 10-year treasuries and 1-year treasuries)	4.603∗∗∗
Low growth dummy (yes-no)	0.116∗∗∗
Dummies for industries	Yes

$n = 737{,}433$: Including 7170 equity issues; 2546 convertible bonds; 10,400 public bonds; 20,322 bank loans; cross sections of time series (firms months)
pseudo-R2 = 0.05; ∗∗∗ means statistical significance at 0.01
Source: Erel et al. (2009)

constant term is assumed to be 0). Also, there is no information about testing the parallel regression assumption.
■

More on Ordered Models

The example demonstrates one of the possible uses of multinomial ordered models in finance. As mentioned at the beginning of this section, other applications include corporate credit ratings (e.g., Hirk et al. 2017), bond default ratings (e.g., Mizen and Tsoukas 2012), and many others.

The advantages of using ordered models can be summarized as follows:

- The number of categories is larger than in binomial models.
- They are explainable by the quantitative latent variable; therefore, they are similar to quantitative variable models.
- They produce a single set of parameter estimates.
- They are the standard tool for modeling ratings.

This last characteristic is elaborated in the work on ordered models by Greene and Hensher (2010). Their book concentrates on technical issues, presenting many applications from the social sciences as well as from finance. Their reference list includes papers such as Hausman et al. (1992) on the ordered probit analysis of stock prices and Mora (2006) on sovereign credit ratings.

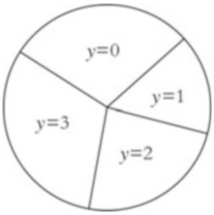
Other applications of ordered models are shown elsewhere in this book.

2.5 The Multinomial Unordered Variables Model: The Choice of Auditor by Chinese Companies

Unordered multinomial variables are modeled in microeconometrics with more than one equation, not as in the case of binomial or ordered multinomial variables. Each equation represents one category related to a base category chosen from all possible states of the multinomial variable y being modeled. So, each equation describes the variable y not "fully" but in "one installment." Equivalent names of such model are multinomial unordered variables model, multinomial response model, unordered multinomial model, and multinomial unordered model.

An analogy to the latent variable y^* for the multinomial unordered model might be the utility function π_{im} of the i-th unit (individual, household, firm, etc.) when choosing category m ($m = 1, 2, \ldots, J$). Utility is a function of the X variables, specific to the choice of category. The unit chooses a given category if its utility is the highest.

Table 2.10 The unordered multinomial model

Characteristics	What is modeled?
Variable y has several categories (states) which cannot be ordered—e.g., $y = 0$ (client chooses bank A), $y = 1$ (chooses bank B) $y = 2$ (bank C), $y = 3$ (bank D)[a]. *Proposition*: Client chooses the bank which maximizes her/his utility. The probability of choosing a particular bank—p_0, p_1, p_2 or p_3—can be determined assuming the type of probability distribution.  	Probability ratios: p_1/p_o p_2/p_o p_3/p_o (category $y = 0$ is treated as base category) and: $p_0 + p_1 + p_2 + p_3 = 1$ In the unordered multinomial model, the probability ratios p_m/p_0 (here $m = 1, 2, 3$) are modeled as nonlinear functions of the explanatory variables X. For different categories m, the parameters of X variables differ.

Source: Gruszczyński (2012a)
[a]Inspired by Ardic and Yuzereroglu (2006)

Table 2.10 presents one possible narrative for the unordered multinomial variable. The y variable here represents the client's choice of bank. We are interested in the probabilities p_0, p_1, p_2, or p_3 of choosing, respectively, bank A, bank B, bank C, or bank D. The client chooses the bank, which maximizes her/his utility. It is assumed that utility is the sum of the nonrandom element associated with the choice (function of the X variables) and a random error with the distribution representing the type of model being used.

In the unordered multinomial logit, the objects of modeling are logarithms of the following probability ratios—for simplicity we assume that $m = 0, 1, 2, 3$ as in the example:

$$\frac{p_{i1}}{p_{i0}} = \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_1\right) \tag{2.12}$$

$$\frac{p_{i2}}{p_{i0}} = \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_2\right) \tag{2.13}$$

$$\frac{p_{i3}}{p_{i0}} = \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_3\right) \tag{2.14}$$

where we assume

$$p_{i0} = \frac{1}{1 + \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_1\right) + \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_2\right) + \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_3\right)} \tag{2.15}$$

From the above, we also have

$$p_{im} = \frac{\exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_m\right)}{1 + \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_1\right) + \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_2\right) + \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_3\right)} \qquad \text{for } m = 1, 2, 3 \tag{2.16}$$

From Eqs. (2.12, 2.13), and (2.14), it follows that the three models

$$\ln\left(\frac{p_{im}}{p_{i0}}\right) = \boldsymbol{x}_i'\boldsymbol{\beta}_m \qquad \text{for } m = 1, 2, 3 \tag{2.17}$$

have the same set of X variables. The estimation of Eq. (2.17) using the maximum likelihood method yields estimates of three sets of parameters: $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\beta}_3$. The estimates indicate the direction of association between a variable X and the probability that $y = m$ related to the probability that $y = 0$.

The unordered multinomial logit is one of many possible approaches to modeling unordered categories, one example of which is the mixed logit model described in Chap. 3. There are also available several references providing broader and more detailed expositions of multinomial models—e.g., Long (1997); Maddala (1983); Cameron and Trivedi (2005); Winkelmann and Boes (2006); Gruszczyński (2012b).

Example 2.3 Choosing Big 4 Companies as IPO Auditors in China
Gao et al. (2013) studied the effect of the so-called "Kelon affair" on Big 4 firms as IPO auditors. Kelon—one of the largest manufacturers of household goods in China—was a client of Deloitte and Touche. In May 2005, Kelon was investigated by the China Securities Regulatory Commission (CSRC) for potential violations of securities laws and accused of recognizing false revenues, underestimating expenses (the allowance for bad accounts), and self-dealing in related party transactions. The authors studied the consequences on Deloitte from this auditing failure, and also the contagion effects on other Big 4 audit firms in China—the others then being PWC, Ernst & Young, and KPMG.

One of the estimated models is the multinomial logit regression model employed to analyze the likelihood of choosing Big 4 firms as IPO auditors after the Kelon failure. There were 374 IPOs in China in the period of 2003–2007. As the IPO process usually takes at least a year and the CSRC suspended new IPOs between June 2005 and June 2006, the effect of the Kelon affair could be first experienced in 2007. In Table 2.11, this effect is represented by dummy variable *year2007*. In the second model, the estimate of the parameter by this variable is negative and significantly different than zero. It means that the odds of choosing a Big 4 firm as IPO auditor was lower than the odds of choosing a small local audit firm.

The results indicate that smaller companies (with lower total assets) and companies with higher leverage have lower preferences for employing Big 4 firms as IPO auditors. In addition, companies controlled by the central or local government have less incentive to employ a Big 4 firm as auditor. Obviously, the study presents only one specific period from the past and has no direct relevance to the current auditing landscape in China.

■

Table 2.11 The choice of audit firm as IPO auditor in China in 2003–2007: unordered multinomial logit estimation results

Explained variable: IPO auditor: $\boldsymbol{y} = 0$ if the company chooses a small local audit firm as IPO auditor, $\boldsymbol{y} = 1$ if the company chooses a big local audit firm as IPO auditor, $\boldsymbol{y} = 2$ if the company chooses a BIG 4 audit firm as IPO auditor.		
Explanatory variables	Model for $y = 1$ [ln (p_1/p_0)]	Model for $y = 2$ [ln (p_2/p_0)]
size = log of total assets	0.366**	1.540***
leverage = total liabilities/total assets	−0.630	−4.091**
ROA (net income/total assets)	−0.582	−9.192
centralgov (= 1 if the firm controlling party is the central government)	−0.964**	−1.792**
localgov (= 1 if the firm controlling party is local government)	−0.726**	−1.447**
year2007 (= 1 for year 2007)	0.211	−1.422**
Constant	−8.106**	−31.834***
$n = 374$ IPOs	$-2\ln L = 513{,}425$	$-2\ln L = 513{,}425$

*** means statistical significance at 0.01, ** at 0.05
Source: Gao et al. (2013)

This example also shows a methodological solution that can be applied even if we know that the categories might be somehow ordered—like the auditor's category. The unordered model has the advantage that probabilities p_1, p_2 are explained separately: for categories $y = 1$ and $y = 2$, we obtain distinct parameter estimates with the same explanatory variables. This feature may be important when we concentrate on a specific category—such as Big 4 firms in the example.

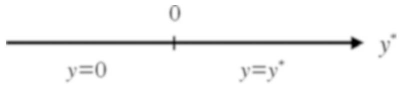
2.6 The Tobit Model: Why Foreign Outside Investors Provide Capital to a Country's Firms?

In financial microeconometrics limited dependent variable models like the Tobit model are used in cases in which multiple regression needs refinement due to the nature of the explained variable. The Tobit model is the regression of the quantitative y variable against the explanatory variables X when the y is censored. This happens when several observations on y are equal to a specific value (e.g., zero), while the remaining observations take values other than zero.

In the example of the Tobit model shown in Table 2.4, the dependent variable y is the dividend ratio. Within a sample of companies, there are usually many zeroes for y (e.g., for companies choosing not to pay dividends), while other values of y are positive (e.g., for companies paying dividends). We say that y is censored from below. If the values of the X variables for the entire sample are known, then the Tobit regression can be estimated.

The Tobit model can be defined in many ways. One explanation assumes that there is a latent variable y_i^* representing the "propensity" of y_i to attain values higher

Table 2.12 The Tobit model

Characteristics	What is modeled?
Variable y is censored. It has two states: $y = 0$ or $y > 0$. Example: "What was the price of the car your family bought last year?" Variable y represents car price. Answers to these questions are many zeroes ($y = 0$) for the families not buying a car, and positive values ($y > 0$) for the families that actually bought a car. *Proposition*: There is an unobserved continuous variable y^*= "propensity to buy a car." We presume that if $y^* > 0$, then the car is bought for the price $y = y^*$. If $y^* \leq 0$, then there is no car purchase and $y = 0$. In economics, we say that the family has chosen a *corner solution* ($y = 0$) or an *interior solution* ($y > 0$) to the problem of maximizing the household's utility function.	Variable y as the linear function of the explanatory variables X. Specifically: The modeled variable y has a conditional discrete–continuous distribution, which is the mixture of the continuous distribution (values $y > 0$) and the one-point distribution ($y = 0$). 

Source: Gruszczyński (2012a)

than zero (e.g., the propensity of the i-th company to pay a dividend for the year in question). Variable y_i^* is the linear function of the explanatory variables X, the same as in Eq. (2.2)

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$$

In fact, what we observe is y_i which is equal to

$$y_i = y_i^* \text{ } if \text{ } y_i^* > 0 \tag{2.18}$$

$$y_i = 0 \text{ } if \text{ } y_i^* = 0 \tag{2.19}$$

If the i-th company pays a dividend, then its propensity y_i^* is positive. If the i-th company does not pay a dividend, then its propensity y_i^* is equal to zero.

The Tobit model is estimated using the maximum likelihood method with the assumption that the error terms ε_i have identical independent normal distributions. Compared to the classical regression of y against the X variables, the estimation of the Tobit model is more complicated. This is because variable y is not continuous but has a discrete–continuous distribution: one point at $y = 0$ and a continuous distribution for $y > 0$.

Table 2.12 presents characteristics of the Tobit model

To summarize, if the endogenous variable y is continuous but has many zeroes then it should be modeled with the use of the Tobit model, not with classical multiple regression. However, there exists a very approximate relation between estimates of the Tobit model obtained using the maximum likelihood method (ML) and the least squares (LS) estimates of regression model.

If we estimate the model

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \tag{2.20}$$

where $y_i \geq 0$ and where the values $y_i = 0$ are frequent, then the relation between the estimated parameters is as follows:

$$\boldsymbol{b}_{LS} \approx \frac{n_1}{n}\boldsymbol{b}_{ML} \tag{2.21}$$

where $\boldsymbol{b}_{LS}$ and $\boldsymbol{b}_{ML}$ are, respectively, estimates of the parameters $\boldsymbol{\beta}$ for the regression model (LS) and for the Tobit model (ML), n is the sample size (the number of all observations), and n_1 is the number of observations with $y_i > 0$. So, the Tobit estimates are equal to the LS estimates divided by the share of the uncensored observations. If there is no censoring, then the two estimates are equal.

Example 2.4 Why Foreign Outside Investors Provide Capital to a Country's Firms?

Leuz et al. (2010) studied the determinants of US investments in companies around the world. The sample includes 4411 firms in 29 countries (1997 data), of which 2469 are European companies (including 1077 from the UK), 61 are from South America, 800 are from developing Asian economies, and 1081 are other (including 979 from Japan). Around 25% of the companies selected for the sample have zero US investments. Therefore, the Tobit model was chosen for examining the relationship between US investors' share in the firm's equity and its determinants. Table 2.13 shows the estimation results.

Table 2.13 US investment in 4411 companies in 29 countries (1997): Tobit model estimates

Dependent variable: Percent of the firm's float[a] that was held by US portfolio investors as of year-end 1997, $y = 0$ or $y > 0$.	
Explanatory variables	Estimate (significance)
Family/management = percentage of control rights held by family groups and the top management group	−0.018∗∗∗
XLIST = 1 if the firm's equity is listed on a US exchange, = 0 otherwise	9.952∗∗∗
MSCI = 1 if the firm's equity is listed on the MSCI Word Index, = 0 otherwise	3.549∗∗∗
Ln(size) = log of total assets in millions USD	3.115∗∗∗
Leverage = ratio of total liabilities to total assets	−0.103∗∗∗
Book-to-market = book equity value/market equity value	−1.021∗∗∗
Dividend yield = dividends paid/price	0.105
Country dummies	Yes
Industry dummies	Yes

$n = 4411$ firms in 29 countries; data for 1997

Pseudo $R^2 = 0.07$; ∗∗∗ means significance at 0.01

Source: Leuz et al. (2010)

[a]Float is the percentage of shares not held by large blockholders (as provided by Worldscope's Closely Held variable) multiplied by the market value of equity in billions of US dollars

It was found that a lower level of corporate governance is associated with lower foreign (US) investment. Foreigners invest less in firms that are poorly governed i.e., have an ownership structure more "conducive to outside investor expropriation" (e.g., more family and management ownership).

In the same paper, the authors employ other Tobit models and show that the share of foreign investors in a country's firms is related to the country's information rules and legal institutions. In addition, it is lower in firms that appear to engage in more earnings management.

■

2.7 Multiple Regression: CEO Cash Compensation, Accounting Performance, and Compensation Committee Quality

Qualitative and limited dependent variable models are the core of microeconometrics. In previous sections, we presented examples of binomial, multinomial, and Tobit models. However, microeconometrics is all about using a regression-type approach to relate the explained variable y (qualitative or quantitative) to a set of explanatory variables X when the sample contains numerous microdata. Therefore, classical multiple regression, with the quantitative y variable, is also part of microeconometrics, including financial microeconometrics.

Linear multiple regression is, as in Eq. (2.20) without restrictions on y_i, namely

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \tag{2.22}$$

where y_i is the i-th observation of the explained variable y and ε_i is the error (disturbance) term introduced in Eq. (2.2). Let us also recall that $\boldsymbol{x}_i'\boldsymbol{\beta}$ is the linear combination of the explanatory variables X presented earlier as (Eq. 2.1)

$$\boldsymbol{x}_i'\boldsymbol{\beta} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \ldots + \beta_k X_{ki} = \left(1\; X_{1i}\; X_{2i} \ldots X_{ki}\right) \cdot \left(\beta_0\; \beta_1\; \beta_2 \ldots \beta_k\right)'$$

where $i = 1, 2, \ldots, n$.

Assumptions about the error terms ε_i typically depend on the sample type. For cross-sectional data (observations on multiple units at one point in time), it is assumed that the error terms ε_i have (conditional on the X variables) expected value of zero and have the variance that depends on i (also conditional on X) (see Cameron and Trivedi 2005).

Example 2.5 shows such a "traditional" microeconometric model explaining a quantitative variable.

Example 2.5 CEO Cash Compensation, Accounting Performance, and Compensation Committee Quality
This is an example from the domain of corporate governance. Sun and Cahan (2009) examined the effects of compensation committee quality on the association between the cash compensation of chief executive officers (CEOs) and accounting earnings and other variables. The sample comprises data from 2001[3] on 812 US companies with various compositions of compensation committees whose quality has been examined relative to six characteristics:

1. The proportion of CEO appointed directors[4] on the compensation committee.
2. The proportion of directors on the compensation committee with 20 or more years of board service time for the current company.
3. The proportion of CEOs of other firms on the compensation committee.
4. The percentage of shares held by the directors on the compensation committee.
5. The proportion of directors with three or more additional board seats on the compensation committee.
6. The number of directors on the compensation committee.

These six variables were used to calculate a composite variable *CCQUAL* representing the level of compensation committee quality.[5] *CCQUAL* is used as important explanatory variables in the regression model that explains CEO cash compensation.

One model estimated by Sun and Cahan (2009) is shown in Table 2.14. It presents the impact of overall compensation committee quality on the association between CEO cash compensation and accounting earnings and the moderating effects of growth opportunities and earnings status.

The model may verify the hypothesis (from agency theory) that the positive relationship between CEO compensation and the company's financial results is higher when the compensation committee is of higher quality. This hypothesis is verified by checking the significance and the sign of the parameter estimate by the interaction variable $CCQUAL * \Delta ROE$. The authors state that a positive and significant parameter means, in this case, that CEO compensation is tied to a company's results when the quality of its compensation committee is high. This important result is additionally verified with more regression models, including various control variables and examining the endogeneity question.

■

[3]Data from the year before the Sarbanes-Oxley act of 2002.

[4]US directors are (roughly) the same as members of supervisory boards in European companies.

[5]*CCQUAL* is constructed in such a way that its higher level is associated with higher compensation committee quality. Quality is measured individually for each characteristic. Initially, the sign of the regression parameter by that characteristic in the model explaining the remuneration of CEOs was found. The value of *CCQUAL* for each company is the sum of five numbers (0 or 1) for each of five characteristics (the variable representing the percentage of shares held by directors is ignored due to its insignificance). A value of 1 is given for the company if a specific characteristic is higher than the sample median for positive characteristics, or lower for negative characteristics. The average value of *CCQUAL* in this sample is 2.08.

Table 2.14 The relationship between CEO compensation and accounting earnings and compensation committee quality (on boards of directors/supervisory boards) in the US in 2001: OLS estimation

Explained variable: $\Delta\ln(COMP)$ = annual change in log value of CEO salary plus bonus	
Explanatory variables	Parameter estimates (t statistics)
$CCQUAL$ = compensation committee quality	−0.01 (−1.09)
ΔROE = change in ROE (return on assets)	0.05 (0.39)
$CCQUAL*\Delta ROE$	0.23 (3.84)***
RET = stock return (buy-hold returns monthly for fiscal year)	0.18 (6.37)***
Constant	0.07 (1.49)
Dummy variables for sectors	Included

$n = 812$, Adjusted $R^2 = 0.1654$, $F = 6.54$***
*** $p < 0.001$ (one-tailed)
Source: Sun and Cahan (2009)

More on Multiple Regression and Statistical Signification

Example 2.5 presents a typical use of multiple regression in the fields of finance, accounting, and corporate governance. The sample is not random but is meticulously constructed taking into consideration the diversity embedded in the variable (i.e., compensation committee quality).

The estimated model includes X variables that are statistically insignificant but are not removed due to their merit in the research project. While this is a habit which gives statisticians and econometricians a headache, it should not. The so-called *insignificance of explanatory variable* is the result of a statistical test and may tell something about the estimation accuracy but definitely does not refer to a causal relationship. We should remember that an “insignificant” variable deleted from the model “moves” to the error term where it can potentially affect the correlation of this error with the explained variable. This might cause more of a serious problem than leaving the variable in the model with a less accurate estimate.

In addition, the variable which is important for the theory and hypotheses underlying the model is usually present in various model specifications (various sets of X variables) where its parameter β is estimated more or less precisely depending on the adjacent X variables. Another reason for leaving insignificant variables in the model is their possible usefulness in generating better forecasts from the model—it may be of value if the model is operationally used with new data.

The American Statistical Association (ASA) published a formal statement with principles on statistical significance and p-values. The six principles are as follows (Wasserstein and Lazar 2016):

1. *A* p-value can indicate how incompatible the data are with a specified statistical model.
2. *A* p-value does not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.

3. Scientific conclusions and business or policy decisions should not be based only on whether a *p*-value passes a specific threshold.
4. Proper inference requires full reporting and transparency.
5. A *p*-value, or statistical significance, does not measure the size of an effect or the importance of a result.
6. By itself, a *p*-value does not provide a good measure of evidence regarding a model or hypothesis.

Those views strengthen our hesitation toward putting too much weight on "statistical significance" in financial microeconometrics. More views on statistical significance can be found in the influential book by Ziliak and McCloskey (2008).

2.8 How to Prove Causality in Regression[6]: The Advent of "Metrics"

The most intriguing question for researchers is how to prove causality between modeled categories—with the use of microdata. As we know, multiple regression as a powerful tool may give evidence of association/correlation between the explanatory variable *X* and the explained variable *y*. Such association, if properly shown, might be of great value in discussing the research question but does not represent the effect of causality. Using regression, we commonly seek to show that some activity represented by the exogenous variable *X* is (or not) the determinant of the behavior of the *Y* variable. This is acceptable as long as by "determinant" we understand *X* being associated or correlated with *Y*. To prove that *X* is the "cause" of *Y,* we need more than a regression model and, even so, the proof is not very general.

In their survey paper on causality in empirical corporate finance research, Atanasov and Black (2016) begin with the following paragraph:

> Much corporate finance research is concerned with causation – does a change in some input cause a change in some output? Does corporate governance affect firm performance? Does capital structure affect firm investments? How do corporate acquisitions affect the value of the acquirer, or the acquirer and target together? Without a causal link, we lack a strong basis for recommending that firms change their behavior or that governments adopt specific reforms.

In order to prove the validity of these types of questions for data representing a particular time and space setting, it is necessary to use techniques other than regression. They are based on the "treatment effects" approach (i.e., on understanding that there is a treatment variable like "new governance order in company" and an effect variable like "company's financial performance." For examining causal relationships, the most appropriate methods are counterfactual analysis, such as

[6]This section uses several paragraphs from Gruszczyński (2018b).

matching techniques (including propensity score matching), difference-in-differences (Diff-in-Diff) methods, the regression discontinuity approach, among others.

These methodological approaches are appealing, although not easy to apply correctly. Atanasov and Black (2016) suggest a comprehensive list of conditions to be met in order to accept analyses based on the counterfactual approach. The authors propose consideration of the following simple regression:

$$q_i = b * gov_i + \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i$$

where q_i is the outcome variable (e.g., Tobin's q), gov is the governance variable, b is the parameter for gov, and $\boldsymbol{x}_i' \boldsymbol{\beta}$ is the linear combination of the other explanatory variables (covariates), including the constant term.

Assume that the estimate of b is significant and positive. This means that, conditional on the covariates, on average higher gov predicts higher q. However, this does not mean that a change in gov will *cause* a change in q. By *cause* we mean that, if one increases gov, changing nothing else, q will change. The following problems enumerated by the authors arise:

1. *Reverse causation*: Meaning the possibility that q causes gov
2. *Omitted variable bias*: Perhaps an unobserved variable causes both q and gov or intervenes in the relation between q and gov
3. *Specification error*: Possibly the relation between q and gov is different; this problem includes also *functional form misspecification*
4. *Signalling*: Firms may change the level of gov "to signal to investors something about management attitudes, or other factors which investors can't readily observe"
5. *Simultaneity*: It is possible that q and gov and the X variables are determined simultaneously; it might be that simultaneously gov determines q and q determines gov
6. *Heterogenous effect*: The causal effect of gov on q may depend on the observed (X) and the unobserved (U) firm characteristics
7. *Construct validity*: Both gov and q variables represent multifaceted constructs which are not well defined; the gov variable is the level of governance—usually in the form of an index—encompassing many governance issues; similarly, the construct q imperfectly measures the firm's result.
8. *Measurement error*: All variables, q, gov, and the Xs are subject to measurement error; if the error is not random, then its consequences are similar to those in the case of specification error
9. *Observation bias*: "observed subjects behave differently because they are observed;" firms might change gov level because managers think this change matters, while it has no direct effect
10. *Interdependent effects*: These are effects on firms that adopt "a reform"; "for example, a governance reform that will not affect share price for a single firm might be effective if adopted widely, because investors will then appreciate the reform's impact."

This list does not include concepts from classical econometrics like *identification* or *endogeneity*. The authors believe that they mean "different things to different people." Instead, in the foregoing catalogue, the authors try to itemize more specifically problems that have a lot in common with those classic concepts.

The authors studied some 13,000 articles published in 2001–2011 in 22 leading journals of economics, finance, law, and management. Among them, 863 papers focused on the relationship between corporate governance and the companies' performance and, of those, 74 papers used the *shock-based research* approach (with 40 different shocks/treatments). The authors concentrate on examining three exceptional papers and on showing that—even in such cases—there are doubts as to all the stages of the research process, especially as to techniques aimed at proving causality effects.

The "checklist" for the accurate proof of causality makes it look like it is a task impossible to attain—with so many problems and challenges to face. However, the techniques employed in attempting to show causality effects, sometimes incorrectly applied, are in constant use in finance and accounting research.

Techniques that help to make inferences regarding causality and endogeneity in econometric modeling embrace increasingly popular approaches that all belong to the so-called *'metrics*, a term introduced by Angrist and Pischke (2015) in their book *Mastering 'Metrics: The Path from Cause to Effect*. The text is used in many classes on econometrics and data analysis worldwide (details are provided on Joshua Angrist's webpage).

However, these remedies for the inference about causality or the endogeneity problem are not common knowledge among econometricians. Angrist and Pischke (2017) comment on that in a survey of undergraduate econometrics syllabi at leading US universities. From 18 detailed schedules of econometrics classes, the percentage of time devoted to three (out of 14) major topics is:

- Causal effects: 2.5%
- Differences-in-differences: 2%
- Regression discontinuity methods: 1.4%.

Similar proportions are seen in the content of contemporary econometrics textbooks: 3%, 0.5%, and 0.1%, respectively.

The sections that follow present an eclectic selection of several techniques and problems associated with the modern use of microdata in financial and accounting research and also devoted to researching causality effects. The next section is a short review of treatment effects in corporate finance and accounting research.

2.9 Treatment Effects in Empirical Corporate Finance: Effects of French IPOs

In some cases, microdata enable the analysis of *treatment effects* (TE)—i.e., the results of treatment applied to some elements of the sample. It is argued that the methodology of TE enables the demonstration of the influence (causality) of certain events, decisions, etc. on the explained variables.

The estimation of TE is gaining popularity in corporate finance and accounting research. The treatment effect is simply the positive or negative difference between the result (the variable in question) when the company "is treated" and the result when the company "is not treated." *Treatment* is not necessarily applied from outside, like legal change, but may be a decision of the company itself. Research on *treatments* examines issues such as the effect of a bond IPO on the interest spread of bank loans, the effect on executive pay of management's direct hiring of compensation consultants, whether firms select highly paid peers to justify their own CEO pay, etc. (Tucker 2010). Other examples of treatments are mergers, acquisitions, IPOs, stock splits, redemptions of shares, earnings announcements, auditor's going-concern opinions, etc.

The theory of treatment effects uses the concept of *counterfactual* result, which cannot actually be observed (see Angrist and Pischke 2009). Suppose a company is contemplating whether to undertake an IPO (*initial public offering*). The following explanation also uses an exposition in Gruszczyński (2018a). In the language of "treatments," there are two treatment levels: deciding and not deciding on an IPO. If a company decides on the IPO, then its financial result (e.g., ROE) attains the level y_1 after a while. If the company does not decide on the IPO, the ROE level reaches y_0. The question is whether the IPO decision influences the level of y (i.e., the level of ROE).

In order to estimate the effect of this treatment (IPO) for the sample of companies, we should know two values of ROE for each company: the ROE with the IPO and the ROE without the IPO—i.e., the result with and without the treatment. Then, the (causal) effect of the treatment would simply be the difference between the two results. Unfortunately, a single company (i-th company in the sample) may be observed only after it takes the IPO decision—i.e., either the decision is "yes" or "no." Therefore, we can observe only one result of the possible two: y_0 or y_1. This second result is counterfactual (hypothetical or potential).

ATE and ATT

Let us assume that the IPO decision of the i-th company is represented by the random variable D_i with two possible values: 0 and 1 i.e., $D_i = \{0, 1\}$ where 0 means "no" and 1 means "yes." The potential result for the i-th company is

$$\text{Potential result} = \begin{cases} y_{1i} \text{ if } D_i = 1 \\ y_{0i} \text{ if } D_i = 0 \end{cases} \tag{2.23}$$

where y_{0i} denotes ROE for the i-th company—assuming that there is no IPO—regardless of what really happens; similarly, y_{1i} denotes ROE for the i-th company—assuming that there is an IPO—regardless of what really occurs. The treatment effect for i-th company would be simply $y_{1i} - y_{0i}$. Unfortunately, we observe only one of the values y_{0i}, y_{1i} (the second one is counterfactual).

What we observe in reality for the i-th company is

$$\mathrm{y}_i = \begin{cases} y_{1i} \text{ if } D_i = 1 \\ y_{0i} \text{ if } D_i = 0 \end{cases} \tag{2.24}$$

and this can be written as

$$\mathrm{y}_i = D_i y_{1i} + (1 - D_i) y_{0i} \tag{2.25}$$

where $D_i = 1$ if IPO and $D_i = 0$ if non-IPO.

The average observed difference in ROE values (between companies with and without an IPO) is called the *average treatment effect* (*ATE*) and is equal to

$$ATE = E(y_i|D_i = 1) - E(y_i|D_i = 0) = E(y_{1i}|D_i = 1) - E(y_{0i}|D_i = 0) \tag{2.26}$$

Thus, the value of *ATE* may be quickly calculated. But what we would like to know is the treatment effect "on the treated" (*ATT*)—i.e., the change in ROE for companies deciding on an IPO as compared to the same companies not deciding on an IPO.

The right-hand side of Eq. (2.26) can be written as

$$\begin{aligned} ATE &= E(y_{1i}|D_i = 1) - E(y_{0i}|D_i = 0) \\ &= \{E(y_{1i}|D_i = 1) - E(y_{0i}|D_i = 1)\} + \{E(y_{0i}|D_i = 1) - E(y_{0i}|D_i = 0)\} \end{aligned}$$

So, *ATE* is the sum of two differences. The first is *ATT* (*average treatment effect on the treated*)

$$ATT = E(y_{1i}|D_i = 1) - E(y_{0i}|D_i = 1) \tag{2.27}$$

and shows the causal effect of an IPO in companies that actually decided on an IPO. This is the difference between the ROE for those companies (i.e., $E(y_{1i}|D_i = 1)$) and the ROE for the same companies assuming (hypothetically) that they did not decide on an IPO (i.e., $E(y_{0i}|D_i = 1)$). We cannot calculate *ATT,* but we can calculate *ATE*, which differs from *ATT* by

$$E(y_{0i}|D_i = 1) - E(y_{0i}|D_i = 0)$$

which is called *selection bias*. This is the difference between the average y_{0i} for companies that did and did not decide on an IPO. Remember that y_{0i} is ROE for "non-IPO" companies. Thus, the selection bias is also, as with *ATT*, a hypothetical difference.

Summing up, what we know to this point is

$$ATE = ATT + selection\ bias \tag{2.28}$$

and what can be calculated is *ATE*. We do not know the magnitude and the direction of the selection bias. The question is when the selection bias may equal zero. Firstly, it happens when the variables D_i and y_i are independent. This is possible only for randomized experiments or a simple random sample. In our case this would mean that companies are randomly selected and floated an IPO, or not. Obviously, such scenario is not valid here.

Propensity Score Matching

Another possibility for significantly lowering the selection bias occurs when we can adopt the *assumption of conditional independence* (CIA)—i.e., the independence of variables D_i and y_i, conditionally on the explanatory variables ("covariates") X_i. This may be written as:

$$\{y_{0i}, y_{1i}\} \perp D_i|X_i \tag{2.29}$$

where the sign $\perp$ denotes the relation of independence. The X_i's are like "control variables" in the relationship between D_i and y_i. The CIA is also called the assumption of "selecting on observables." This means that if, conditionally on the X_i's, both groups of companies (the IPO and the non-IPO) choose an IPO, then their results (ROE) have the same distribution. The CIA assumption means that if in the X variables there is enough information about the decision on the IPO then the correlation between y_{0i}, y_{1i}, and D_i can be minimized by conditioning on the X variables (X. Li 2004).

Rosenbaum and Rubin (1983) showed that in such case the treatment effect is equal to *ATE* i.e., the difference between the results of companies "with the treatment" and companies "without the treatment" provided that they have the same probability of treatment. This probability is called *propensity score*.

The *propensity score matching* (PSM) method enables the creation of a comparison group by matching the "IPO" observations (companies) to the "non-IPO" observations for similar values of propensity score. It may be said that the score represents the "propensity of a company to be treated." The general idea is to match "treated" to "non-treated" companies that are as similar as possible. Instead of matching against all the X_i variables, the match is performed with a single measure called the propensity score. "Because the score incorporates the effects of X_i on the probability that $D_i = 1$, PSM ideally creates a sample of treated and untreated observations that are similar across X_i, minimizing correlation between D_i and X_i and reducing concerns about FFM (functional form misspecification)" (Shipman et al. 2017).

The propensity score can be estimated with the use of the binomial logit model with D_i as the explained variable and the covariates X_i as the explanatory variables. Sometimes, it is recommended that interactions, squares, and transformations of the X_i variables be included (Rosenbaum 2001). The estimated probability that $D_i = 1$ for each observation is the propensity score.

After calculating the propensity score for each observation in the sample, the "treated" observations (with $D_i = 1$) are matched with the "untreated" ($D_i = 0$) that have the same or a very close propensity score. Finally, the *ATE* is calculated as the average difference between y_i for pairs of matched observations (with $D_i = 1$ and with $D_i = 0$).

Example 2.6 The Post-Issue Operating Performance of French IPOs: Use of PSM

Chaouani (2010) studied the effects of initial public offerings (IPOs) for 3950 French companies. The operating performances of IPO and non-IPO companies were compared using the propensity score method (PSM).

There are 139 observations for IPO companies and 14,385 observations for non-IPOs for the years 1998–2002. As the first step, the logit model for the endogenous choice variable D_i (IPO of non-IPO) is estimated. The explanatory variables X selected for the model are company and industry characteristics, such as annual rate of sales growth, leverage (total debt over total assets), ROA, log of total assets, log of firm age (years), median market-to-book ratio for firms in the same industry, the number of IPOs in a given year, and dummy variables for industries.

Next, the propensity score (probability calculated from the estimated logit) is obtained for each firm. In the second step, each observation of the IPO group (treated) is matched with observations of the non-IPO firms (control) based on their propensity score within the same year (1998–2002). The matching method used here is "kernel matching."

In the third step, the *t*-tests are performed to check if the means of each X variable differ between the treated and the control firms—before matching and after matching. This shows whether, after matching, the X variables are balanced between the two groups (i.e., do not significantly differ). It turns out that the balancing hypothesis is satisfied for the matched sample (i.e., in this sample the IPO and the non-IPO firms do not differ in terms of X variables).

Next, in the fourth step, the operational effects of the decisions to go public (IPO) are calculated (as average treatment effects) for 1, 2, 3, 4, and 5 years after the IPO. The following financial indices are considered:

Profitability

- *OROA*, operating return on assets: EBITDA[7] divided by total assets
- *NROA*, net return on assets: net income divided by total assets
- *OROS*, operating return on sales: EBITDA divided by total assets

Efficiency

- *AST*, asset turnover: total sales divided by total assets

Output

- *SALEG*, sales growth rate: Total sales in year t minus total sales in year t–1 divided by total sales in year t–1

Leverage

- *DEBT*, debt-to-asset ratio: Total debt divided by total assets.

Table 2.15 presents the estimates of the average treatment effects on the treated *ATT* for various performance measures. *ATT* is the average difference between the "treatment group" (IPO firms) and the "non-treatment group" (non-IPO). It turns out that such differences are negative (i.e., practically all performance measures for the IPO firms are lower than for the non-IPO firms in the years after the IPO). This means that the "IPO effect" is negative: the operational results (the first five indices) are worse in the companies that went public in comparison with the companies that remained private. On the other hand, the debt ratio (sixth measure) is lower for the IPO firms as compared with the non-IPO firms. The author refers to many other studies that confirm such results.

In addition, Chaouani (2010) shows that the operational results of the IPO companies in the years directly before the IPO were much better than in the years after the IPO. Other studies also demonstrate that new listing firms time their IPO to coincide with high performance that is not maintained after the IPO. Also, it is common that after going public the companies reduce their leverage (debt ratio).

■

More on Treatment Effects Methodology

Although matched–comparison evaluation, like PSM, belongs among the quasi-experimental design techniques, it is becoming increasingly standard, and in corporate finance and applied accounting as well. For example, Shipman et al. (2017) examined papers on the use of PSM in leading accounting journals in 2008–2014.

[7]Earnings before interest, taxes, depreciation, and amortization.

Table 2.15 The treatment effects estimated for various performance measures for French IPO firms

Average treatment effect (*ATT*) estimates for post-IPO performance in years 1–5					
Years after IPO	*ATT*	*t*-statistic	Years after IPO	*ATT*	*t*-statistic
OROA			*AST*		
1	−0.044	−3.08	1	−0.599	−5.24
2	−0.029	−2.05	2	−0.377	−3.30
3	−0.044	−3.69	3	−0.448	−3.85
4	−0.059	−6.06	4	−0.713	−6.13
5	−0.055	−5.77	5	−0.716	−6.07
NROA			*SALEG*		
1	−0.024	-1.47^{N}	1	0.116	2.31
2	−0.050	−2.05	2	0.121	1.99
3	−0.048	−2.31	3	−0.099	−2.03
4	−0.043	−3.20	4	−0.463	-0.69^{N}
5	−0.039	−2.44	5	−0.107	−2.03
OROS			*DEBT*		
1	−0.041	−2.31	1	−0.111	−4.68
2	−0.015	-0.43^{N}	2	−0.125	−2.52
3	−0.036	−2.65	3	−0.086	−3.24
4	−0.049	−3.90	4	−0.116	−5.95
5	−0.056	−2.63	5	−0.075	−2.96

N denotes *ATT* insignificantly different than 0
Source: Chaouani (2010)

They revealed that the popularity of this method in accounting research—86 articles in 5 journals, including 70 articles in 2012–2014 alone—is not necessarily associated with its proper use: "studies often oversell the capabilities of PSM, fail to disclose important design choices, and/or implement PSM in a theoretically inconsistent manner." In their extensive review, the authors state that PSM is not an alternative to selection models. Furthermore, they highlight the inaccuracy of statements that "PSM alleviates broad concerns relating to endogeneity, omitted variables, or self-selection, without qualification." PSM also does not emulate experimental conditions and has limited external validity.

The authors produce their own comprehensive example of accounting research for examining possible angles of application of PSM (also versus multiple regression). For their sample of 30,000 observations (firm-years) from the US market, they use three treatments: auditor size (*BIG4*), internal control weakness (*WEAK*), and being followed by at least one analyst (*ANALYST*). Treatment is represented by the dummy variable D_{it} equal to 1 for treatment and 0 for no-treatment. In those settings the authors "demonstrate the impact of common PSM design choices on sample composition and *ATE* estimates." The major suggestion from the authors' undertaking is that "PSM estimates can be fickle and difficult to replicate, indicating the need for stress testing matched sample results and supplementing PSM with alternative research designs." Shipman et al. (2017) propose a set of good practices for situations in which PSM can actually be useful, especially as the remedy for functional form misspecification (FFM) as the identification concern in multiple regression.

Treatment effects modeling should be distinguished from self-selection models (see next section). Self-selection modeling begins with estimation of the selection equation and its results are used for estimating the treatment effects. In the matching model (this section), the treatment effects are estimated directly (see K. Li and Prabhala 2007).

One complementary survey-type paper on selection bias in accounting and finance research was published by Tucker (2010). The author correctly states that PSM is not the remedy for self-selection bias and presents the division of selection bias into two types: (A) due to observables and (B) due to unobservables. The remedies described are PSM for case A and the Heckman inverse Mills ratio for case B (see next section). The latter case (i.e., selection bias due to unobservables) results from a "failure to control for the differences researchers cannot observe," usually due to the smaller information sets accessible to researchers than to managers and market participants.

Tucker (2010) interestingly exposes different uses of the popular notion of "selection bias." Selection bias has the meaning in both cases (A) and (B), as in Eq. (2.27) for case A, but originally and more frequently this term has been used for problems of selection on unobservables (i.e., case B). The author's advice for researchers using PSM in finance and accounting is as follows:

- PSM does not address selection bias due to unobservables.
- PSM does not guarantee that treated and non-treated companies are well matched by the X variables (company characteristics); thus, it is advisable to test the differences in distributions of the explanatory variables between the treated and the non-treated companies matched by their propensity scores and to restrict the inference to companies whose characteristics can be found in both groups of companies.

As in the Shipman et al. (2017) paper described above, Tucker (2010) in her survey also highlights the growing number of articles addressing selection bias in two leading accounting journals. Unfortunately, most articles using this methodology have various flaws. This creates the risk of drawing improper conclusions from otherwise interesting research.

Treatment effects and PSM technology are experiencing increasing popularity in corporate finance and accounting research, along with other quasi-experimental techniques such as regression discontinuity design, and differences-in-differences estimation. They all attempt to investigate cause–effect relationships. It is crucial to study both the details and the impediments of those methods before attempting their use in research. There are many publications to follow as reference (e.g., Caliendo and Kopeinig 2008; Shipman et al. 2017; Atanasov and Black 2016; Angrist and Pischke 2009, 2015).

2.10 Self-Selection Modeling in Empirical Corporate Finance and Accounting Research: Dividend Decisions, Dividend Payments, and Corporate Social Responsibility

Like the treatment effects models, *self-selection* (*sample selection*) modeling is proving to be popular and useful in financial and accounting research. And like the "treatments" described above, self-selection is also about the various decisions made by companies. Firms make decisions about financing and investments, about issuing corporate bonds, about underwriters, etc. At the very beginning of their survey, K. Li and Prabhala (2007) write that "corporate finance decisions are not made at random but are usually deliberate decisions by firms or their managers to *self-select* into their preferred choices."

The nonrandom decisions of firms as the results of self-selection can be grouped in samples of similar decisions made by other firms. These might include decisions about a stock split, the redemption of shares, hiring a Big 4 auditor, a going-concern audit opinion, etc. Such samples are not random and are typical in corporate finance and accounting research.

Outline

The following is a short exposition of a self-selection model (after K. Li and Prabhala 2007; Lennox et al. 2012; Gruszczyński 2012a; Winkelmann and Boes 2006). Let us assume that, for the sample of companies, we aim to estimate parameters $\boldsymbol{\beta}$ in the linear regression model (the formula is the same as 2.22)

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \tag{2.30}$$

The explained variable y_i typically denotes an outcome, effect, or cost (e.g., profitability, return, cost of equity, auditor's fee, etc.) The explanatory variables X (represented by the vector $\boldsymbol{x}_i'$) are supposed to be associated with y_i. If the error term ε_i follows classical assumptions, then the ordinary least squares (OLS) or generalized least squares (GLS) give a consistent estimator of parameters $\boldsymbol{\beta}$.

Now, let us consider a subsample including only companies, which make self-selection D (examples of D are given in the previous paragraphs). We assume that $D_i = 1$ represents companies included in the subsample while companies with $D_i = 0$ are not included. For self-selected companies, the equation is as follows:

$$y_i|(D_i = 1) = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i|(D_i = 1) \tag{2.31}$$

This equation is called an *outcome equation*. The difference between Eqs. (2.30) and (2.31) is central to self-selection: Eq. (2.30) is dedicated to the entire sample (population), while Eq. (2.31) applies only to the subsample of companies choosing decision $D_i = 1$. If the latter are not randomly chosen from the population, then the OLS/GLS estimators for (Eq. 2.31) are no longer consistent. This is a common situation.

Now, suppose that $D_i = 1$ denotes companies that choose an auditor from a Big N firm. The first step in self-selection modeling is to estimate the *selection equation* (i.e., the choice of $D_i = 1$). This should be done only for the entire sample (for companies with $D_i = 1$ and companies with $D_i = 0$ together).

The selection equation is as follows:

$$D_i^* = z_i'\gamma + u_i \tag{2.32}$$

where latent variable D_i^*, as in Eq. (2.2), represents the propensity to choose an auditor from a Big N firm. If $D_i^* \geq 0$, then $D_i = 1$ ("Big N"), and if $D_i^* < 0$, then $D_i = 0$ ("no Big N"). Equation (2.32) with D_i as the explained variable is estimated with the use of binomial probit on the full sample.

The variables X explain variable y and the variables Z explain the choice D. The variables X and Z may overlap. Those Z variables that do not overlap with X in Eq. (2.31) are called *exclusion restrictions* because the researcher presumes they have no direct impact on y.

The Heckit Method

We are interested in estimating $\boldsymbol{\beta}$ in outcome equation (2.31). The error terms ε_i and u_i are assumed to have bivariate normal distribution, independent of X and Z, with means equal to zero, variances equal to, respectively, σ^2 and 1, and correlation equal to ρ. It can be shown that, with these assumptions, the expectation of y_i in Eq. (2.31) is as follows:

$$E(y_i|D_i = 1) = \boldsymbol{x}_i'\boldsymbol{\beta} + \sigma\rho\lambda\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right) \tag{2.33}$$

Expression $\lambda\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)$ is the inverse Mills ratio (*IMR*) and is equal to

$$\lambda\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right) = \frac{\varphi\left(z_i'\boldsymbol{\gamma}\right)}{\Psi\left(z_i'\boldsymbol{\gamma}\right)} \tag{2.34}$$

The *IMR* defined in Eq. (2.34) is the ratio of density and the cumulative density of the standard normal distribution calculated at $z_i'\boldsymbol{\gamma}$. Note that Eq. (2.34) is valid only for observations with $D_i = 1$.

We estimate the outcome equation in the form Eq. (2.33) using the selected sample (only for companies with $D_i = 1$) and the form Eq. (2.33)—i.e., we regress y_i on $\boldsymbol{x}_i$ and $\widehat{\lambda}_i$, where $\widehat{\lambda}_i$ is calculated from Eq. (2.34) with the use of estimates $\widehat{\boldsymbol{\gamma}}$ of the selection equation (2.32). Now estimates of $\boldsymbol{\beta}$ are consistent.

It is worth noting that this two-stage procedure is called the *Heckit method* after Heckman (1976). It consists of:

1. Estimating the probit model (2.32) for the full sample (all i) and
2. Estimating the regression equation (2.33) for the subsample ($D_i = 1$) with the use of $\boldsymbol{\gamma}$ estimated in stage 1.

Self-Selection, the Tobit, and Treatment Effects Modeling

Self-selection models are similar to limited dependent variable models such as the Tobit model described in Sect. 2.6. The difference is that the inclusion of a unit (firm, observation) in the sample is modeled by the additional regression equation (selection equation). Also, the outcome equation here is estimated on the selected sample while the Tobit model uses the full sample.

Lennox et al. (2012) indicate that the self-selection model is "conceptually the same as treatment effect model." The latter might be estimated on the full sample as the regression

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + D_i + \varepsilon_i \tag{2.35}$$

with the selection Eq. (2.33) where the inverse Mills ratio (*IMR*) formula for $D_i = 0$ observations should be added. No matching is assumed here. Hence, this approach is much less restrictive than those described in the previous section (Strawiński 2007, 2014).

Example 2.7 Dividend Decisions, Dividend Payments, and Corporate Social Responsibility

Cheung et al. (2016) examine how corporate social responsibility (CSR) issues are connected with firms' decisions to pay dividends. The authors find support for the view predicting that firms with a stronger involvement in CSR activities may be associated with higher dividend payouts. The sample comprises data from Compustat for which the CSR measures are available for the years 1991–2010 (excluding financials and utilities). There are 1965 firms in the sample with a total of 15,561 firm-years.

The authors employ the Heckman sample selection model describing two company's decisions about the dividend y_i

1. Pay or not pay ($y_i = 0$ or $y_i > 0$).
2. The amount of dividend to pay (for $y_i > 0$).

Table 2.16 Dividend decisions and CSR: Parameter estimates for the Heckman sample selection model

Explained variable: Selection equation: probability of paying dividends Outcome equation: log of (dividend payout ratio = dividends per share/diluted earnings per share)		
Variable	Outcome equation	Selection equation
Constant	−2.607***	−1.055***
csr = CSR score (average of seven CSR dimensions[a])	0.237**	0.194
nysep = percentile of NYSE[b]	−0.001***	0.005***
mtb = market-to-book ratio	0.012	0.011
lev = leverage (long-term debt/total assets)	0.436***	−0.294***
eta = earnings/total assets	0.323	0.578
r_com = cost of equity capital (composite of five measures)	0.486***	−3.691***
beta = beta from market model (3 years daily returns)	−0.383***	−0.126***
idio = idiosyncratic factor[c]	−0.213***	−0.341***
sbuyback = stock buyback[d]	−0.625***	0.167
payer = 1 if dividends paid in previous year, = 0 otherwise	1.899***	2.926***
cash = cash or equivalent to total assets	0.261***	−0.508***
reta = retained earnings/total assets	−0.074**	0.845***
ipo90 = 1 if IPO date after 1990, = 0 otherwise		0.084*
*reta*ipo90*		−0.762***
taxregime = 1 if dividend tax cut in a year, = 0 otherwise		0.409***
mills = inverse Mills ratio (*IMR*)	0.740***	

*** means statistical significance at 0.01, ** at 0.05, * at 0.10
Source: Cheung et al. (2016)
[a]From KLD database
[b]Percentile of New York Stock Exchange to which a firm's market capitalization belongs
[c]Residual variance from the Fama–French three-factor model over the previous 3 years
[d]Increase in treasury stock or total expenditure on the purchase of common and preferred stock minus any reduction in the redemption value of the net number of preferred stock outstanding (the greater of the difference between purchases and sales of common and preferred stock or zero)

Estimation results from Heckman's two-stage procedure show that CSR is not associated with a propensity to pay dividends (selection equation) but has significant relation to dividend payout (outcome equation). The authors state that "CSR is a driver of the dividend payout ratio but not of the propensity to pay because firms with high CSR scores are as likely to pay dividends as those with low CSR scores; among those firms that pay dividends, firms with high CSR scores tend to pay more dividends out of their earnings." It should also be noted that the estimate of the parameter by the inverse Mills ratio (*IMR*) is statistically significant. It is evidence of selection bias (Table 2.16).

■

2.11 Endogeneity

Endogeneity means the situation in which the explanatory (exogenous, regressor) variable X in a regression model is correlated with its error term (i.e., correlated with another regressor that has not been included in the model). This means that X does not explain the dependent variable y "from outside" but explains variable y in connection (correlation) with the error term. In such a case, the X variable is in fact "endogenous" (i.e., explained by factors present in the error term, similar to those explaining the y variable). Moreover, the error term in this situation is not actually random because it is partially predictable by the X variable.

Endogeneity means that, for example, the ordinary least squares estimator in a linear regression is not consistent. The solution of endogeneity in a specific model depends on various factors and is never simple. F. Li (2016) states that "most corporate financial decisions are determined endogenously in a complex network of relationships."[8]

Endogeneity in Examining Company Performance Versus CEO "Power"

The following narrative is taken from F. Li (2016). Assume that the performance of a company (endogenous variable) is explained by CEO "power" and by other explanatory (exogenous) variables. CEO power is expressed by means of an index showing the "salary distance" between the CEO and the no. 2 person in the company—i.e., (CEO compensation minus no. 2 person compensation)/(CEO compensation). This index is called the *GAP*. The financial performance of the company is represented by Tobin's q. How can the *GAP* be endogenous in the model explaining the company's performance (q)? According to the author there are two possibilities:

> . . . either causality runs from q to the *GAP*, or causality runs both ways. A random shock that enters the regression model through the error term affects q. Because q affects the *GAP*, *GAP* will be correlated with the error term, generating a biased coefficient on the *GAP*. The second situation is that the *GAP* and q have no direct effect on each other, but they are spuriously correlated through some third variable. If we do not explicitly control for the third variable, the error term will absorb the effect of this variable. Thus, the error term will be correlated with the GAP, causing biased and inconsistent estimates. (F. Li 2016)

So, in the regression model we have the left-hand side variable q and the right-hand side variable *GAP*. To address the possible endogeneity of the *GAP*, one can use various techniques. F. Li (2016) uses the instrumental variables approach, the fixed effects model (firm and year effects), lagged dependent variables, and control variables. All this results in a significant change in the regression coefficient by the *GAP* variable: from positive when the endogeneity effect is not addressed to

[8]This section uses several paragraphs from Gruszczyński (2018b).

negative when the techniques of accounting for endogeneity are applied. In another words: more CEO power is associated with poorer company performance.

Reverse Causality and Unobserved Heterogeneity

In a survey paper on corporate governance in financial and accounting research, Brown et al. (2011) identify problems of *reverse causality* and *unobserved heterogeneity*.

In corporate governance studies, the relationship between corporate governance level ($X = CG$) and variables such as a company's result ($Y = result$) often represents simultaneous causality. In fact, *CG* and *result* are potentially correlated with variables representing the error term. In a model where *result* is explained and *CG* is explanatory, the disturbance represents all variables that have relation to *result* and are not included in the model. A company's result has many determinants, not only governance level ($X = CG$). Other determinants include the business environment, innovativeness, the propensity to adapt to changing market conditions, etc. Some of these variables are also associated with $X = CG$ because a company that is more responsive to challenges usually has a better level of governance.

The model may also suffer from "unobserved heterogeneity," where the identified relationships are symptoms of some unobservable factor(s) that drive both X and *result*.

> Because in both of these cases the explanatory variable(s) will be endogenous and correlated with the residuals in the regression model, OLS (ordinary least squares) is biased and inconsistent. It follows that any study that unreasonably ignores the possibility of endogeneity, but makes a causal argument that, say, better *CG* leads to better firm performance, is at the very least incomplete. (Brown et al. 2011)

In order to solve this type of endogeneity, the authors propose the use of the fixed-effect panel approach and the instrumental variables (IV) approach.

Instrumental Variables Approach

Referring to an example with *results* as the dependent variable and *CG* as the explanatory variable in the model, the IV method begins with identification of a set of so-called "instruments" Z for *CG* and estimate the model consistently using two-stage least squares (2SLS).

So, there is a second equation (model) where *CG*, being "suspected" of endogeneity is explained by variables-instruments Z and other exogenous variables. The estimated value of *CG* from this equation is now entered into the main equation explaining *result*. This approach is equivalent to the two-stage least squares (2SLS) known for estimating simultaneous equations. All stages in this procedure should be accompanied by specific testing (such as the Hausman test). There is also the question of identification of the instruments. The method of IV is presented in many textbooks (e.g., Maddala and Lahiri 2009, Wooldridge 2019).

The Surveys of Roberts and Whited (2013), Atanasov and Black (2016), and Gippel et al. (2015)

In their survey on endogeneity in corporate finance, Roberts and Whited (2013) address the challenges of dealing with endogeneity in a specific model. They indicate two groups of techniques as remedies for endogeneity. First are those which consider the source of the variability of the exogenous variable: the instrumental variables approach, the differences-in-differences estimators, and regression discontinuity design. The second group of techniques makes use of endogeneity in the modeling itself (e.g., the use of panel data or matching estimation).

We now know that addressing endogeneity in regression-type models may give inconclusive results, depending on the research setup, the researcher's creativity, etc. Nevertheless, always considering this question is advocated. Atanasov and Black (2016) comment on this:

> We share neither the perspective of some researchers, whose view can be caricatured as "endogeneity is everywhere, one can never solve it, so let's stop worrying about it"; nor the "endogeneity police," whose attitude is that "if causal inference isn't (nearly) perfect, a research design is (nearly) worthless"; nor that of authors who know they have an endogeneity problem, but say little or nothing about it in their paper, hoping the referee won't notice, or else use a weak instrument to address endogeneity and hope the referee won't object. Our anecdotal sense is that paper acceptance and rejection decisions often turn on which position—endogeneity is everywhere, endogeneity police, or our middle ground—best describes the referee and the editor. (Atanasov and Black 2016, 210)

Similar objections are raised in the survey paper on endogeneity in empirical corporate finance and accounting by Gippel et al. (2015). The authors state that endogeneity is often the problem "that researchers either do not address at all, or simply note in passing that the problem exists." They indicate that the solution to endogeneity depends not only on sophisticated econometric technique but also on the underpinning theory. In their survey, the authors discuss typical solutions to endogeneity: 2SLS, instrumental variables, and generalized method of moments (difference GMM and system GMM)—together with relevant tests. Their major proposal lies in advocating *natural experiment* "as the way of mitigating endogeneity and building stronger theory." In the case of research in finance, natural experiment is a naturally occurring event or state that originates from legislation, policy, etc. Such an experiment is often called a *quasi-experiment*. A methodology that proposes dealing with endogeneity by using naturally occurring exogenous events seems to have limited applicability in numerous research studies in corporate finance and accounting.

2.12 Difference-in-Differences Estimators and Regression Discontinuity Designs in Corporate Finance and Accounting Research

The techniques of Diff-in-Diff and RDD that are outlined here belong to the quasi-experimental techniques mentioned earlier in Sect. 2.8 as tools for studying causal effects. A more detailed presentation of 'metrics methodology is contained in the book by Angrist and Pischke (2015).

Difference-in-Differences Estimators

Difference-in-Differences estimators are applicable in a situation in which a "treatment" group of observations and a "non-treatment" (control) group differ but move in parallel. To apply Diff-in-Diff, we need panel data (at least two time periods) and the exogenous event, treatment, and change that occurs in between.

In the classic study by Card and Krueger (1994), the research question is: What are the employment effects of the minimum wage increase in New Jersey (NJ) effective April 01, 1992? The authors collected data from 410 fast food restaurants near the border between the states of NJ and Pennsylvania (PA) two months before and seven months after the minimum wage increase in NJ. The PA restaurants represent the untreated (control) group and the NJ restaurants represent the treatment group.

The outcome variable y_i is employment, the indicator variable D_i has two values: $D_i = 1$ for the treatment group and $D_i = 0$ for the control group, the time variable t_i represents two periods: $t_i = 1$ and $t_i = 0$ [treatment occurs between these periods], i is the index of restaurant. Then the Diff-in-Diff estimator is the "difference in y for NJ" minus the "difference in y for PA"

$$\beta_1 = [E(y_i|D_i = 1, t_i = 1) - E(y_i|D_i = 1, t_i = 0)]$$

minus

$$[E(y_i|D_i = 0, t_i = 1) - E(y_i|D_i = 0, t_i = 0)] \tag{2.36}$$

Another way to obtain β_1 is to estimate the following regression equation:

$$y_i = \beta_0 + \beta_1 D_i * t_i + \beta_3 D_i + \beta_4 t_i + \varepsilon_i \tag{2.37}$$

This is because

$$E(y_i|D_i = 1, t_i = 1) = \beta_0 + \beta_1 + \beta_3 + \beta_4 \text{ and } E(y_i|D_i = 1, t_i = 0) = \beta_0 + \beta_3$$
$$E(y_i|D_i = 0, t_i = 1) = \beta_0 + \beta_4 \text{ and } E(y_i|D_i = 0, t_i = 0) = \beta_0$$

that coincides with Eq. (2.36). So, the Diff-in-Diff estimate eliminates the "state effect" β_3 and the "time effect" β_4.

Roberts and Whited (2013) provide another example of treatment and Diff-in-Diff. The state of Arizona passed anti-takeover legislation in 1987, at which time Connecticut had not passed similar legislation. The year 1986 represents the pre-treatment period, the year 1987 the post-treatment period. Firms registered in Arizona represent the treatment group; those registered in Connecticut the control group. Variable y may represent companies' outcomes.

The key assumption underscoring such exercises is that, in the absence of treatment, the average change in the response variable would have been the same for the treatment and the control groups. The assumption is termed "parallel trends assumption" because it requires that trends in the y variable in the treatment and the control groups before the treatment are the same. Under such condition, the Diff-in-Diff-estimator is consistent.

Roberts and Whited (2013) point out that this key assumption of parallel trends in untestable. Therefore, they present "a laundry list" of sensitivity and robustness tests that should be performed if one employs the Diff-in-Diff approach in research.

Regression Discontinuity Design

Regression discontinuity design (RDD) is another quasi-experimental technique, popular among researchers in all social sciences, including corporate finance and accounting.

We use RDD when for a certain variable X the observation is "treated" if $X > X_0$. Value X_0 represents the threshold (cutoff) that is known. The variable X is called the running variable or the forcing variable. The observations representing the "recipients" of the treatment (firms, managers, clients, investors, etc.) whose X variable is above the cutoff are assigned to the treatment group, those below to the nontreatment group.

Now, the X variable is the regressor (possibly with other regressors) in a regression model describing the outcome variable y. The primary idea is that observations falling just below and just above the cutoff are relatively comparable. Angrist and Pischke (2015) give the example of Americans aged 21 years and older that can drink legally (X variable = age, with the threshold at 21) and the death rate (from all causes) as the y variable. In the field of finance, Roberts and Whited (2013) quote the example of financial covenants that specify thresholds for some accounting variables for a firm (X variables) and the level of financing y (bank loan) (source: Chava and Roberts 2008).

Two types of RDD may be considered:

1. *Sharp RDD*—The assignment depends only on the forcing variable (e.g., the Minimum Legal Drinking Age [MLDA] is the sharp threshold and, in studying the effect of MLDA on mortality, we use sharp RDD).
2. *Fuzzy RDD*—Passing X_0 increases the probability of treatment but also other variables X may determine if the observation is treated or not.

There are variants of both types of RDD. Generally, for estimating the treatment effect in a sharp RDD case, we may estimate a single equation regression model, for fuzzy RDD more equations are needed.

Testing in this methodology might be complex, as in the Diff-in-Diff case. For example, one of the assumptions in sharp RDD is local continuity, which ensures that the expected outcome is similar for observations close to but on different sides of the threshold (i.e., in the absence of treatment, the outcomes would be similar). Roberts and Whited (2013) mention in this regard the question of manipulation: "the ability of subjects to manipulate the forcing variable and, consequently, their assignment to treatment and control groups." These and other considerations should be given a lot of attention when designing research with the use of RDD.

As with other "new" methods, there are questions of proper application of Diff-in-Diff and RDD to relevant research situations. However, it is always worth considering and attempting the use of such techniques in searching for proof of causality in relations between variables in corporate finance and accounting.

2.13 Good Practices

Modeling Strategy in Financial Microeconometrics

This chapter has presented many issues of modeling in financial microeconometrics. The most welcome modeling strategy is the use of foundations from the theory of economics and finance. The discipline of finance has a very quantitative background and several solid theories. However, as with other social science disciplines, theories are often far from current reality, their longevity is low, and they have limited validity across countries and across financial regulations. Similarly, research in corporate finance and accounting frequently yields inconclusive and/or dissimilar results, depending on the market, the sample, the observation period, etc. It should be accepted that the research result is limited to the "here and now" and is not general like a mathematical theorem.

The previous chapter presented comments on corporate finance theories and their empirical counterparts. Most corporate finance and accounting research, quoting relevant theories, relies on stylized approaches from previous research in the relevant stream of literature. In many cases, the lack of a mature or adequate theory forces researchers to adopt a greater reliance on data. Modern (advanced) data analysis techniques with exponentially growing amounts of available data seem to be prevailing in various practical uses of empirical corporate finance and accounting.

On the other hand, the abundance of data may serve to develop fascinating new methodologies as in agent-based economics and other fields.

The Deficiencies of the Regression Model

In Sects. 2.7 and 2.8, we mentioned the inadequacies connected with the use of regression. Researchers in corporate finance and accounting typically believe in the power of regression—i.e., the equation in which there is a left-hand side variable y (explained, endogenous) and a number of right-hand side variables X (explanatory, exogenous). That model can take many forms, but the concept remains the same: there are some determinants (explanatory variables), which explain the behavior (variability) of the explained variable. And, when one shows—with some care—that "statistically" the model is correct, then this constitutes the "proof" of the validity of the examined relationship. Yes, this might be the case, but usually it is not. Why? There are several aspects, controls, and questions that such model should answer. If we do not follow a specific "checklist" of such items, the model may seem to be satisfactory but is, in fact, neither correct nor appropriate.

Knowing that, is it worth using regressions in corporate finance and accounting research? The answer is obviously, "yes." Regression analyses are valuable, especially when there is no alternative. Their outcomes often come close to ascertaining causality, particularly when panel regression techniques are used. The common regression (correlation) technique based only on observational data has no ability to evidence causality. However, it has significant interpretative value, especially when the sample includes companies properly assembled.

The issue of causality and methods for showing causal relationships between variables were presented in Sects. 2.8, 2.9, 2.10, 2.11, and 2.12. The methodology based on the concept of treatment is what we advocate attempting for successful research on causality in finance and other fields.

Good Practices[9]

The questions and discussion in this chapter show that it is of vital importance to have some common understanding about the research methodology of statistical-econometric origin as applied to corporate finance and accounting. The obvious starting point is always a world literature survey where one finds the major stream of the hypotheses in question. At the core are papers in leading A-journals, mostly papers with a quantitative (statistical) focus. Statistical-econometric methodology in contemporary corporate finance and accounting research is dominated by techniques

[9]This subsection first appeared in Gruszczyński (2018b).

based on the use of microdata. In typically proposed regression-type models, the techniques of microeconometrics and advanced data analysis are prevalent. The variety of research questions—and proposals as to how to solve them—might certainly suggest some standards or rules of good practice to follow.

Good practices in microeconometric modeling are proposed on the basis of our experience and on survey articles of similar type. The paper by Renée Adams (2017) provides important inspiration for researchers in corporate governance topics. We have also used the papers of Kennedy (2002) on the 10 commandments of applied econometrics as well as the web page of Rob Hyndman (https://robjhyndman.com). The pitching template by Robert Faff (2017) has also proved very useful.

And now the good practices.

1. Dedicate a reasonable share of your time to this[10]:
   (a) Formulate major features of your research question in one sentence.
   (b) Identify the sources (papers) of the scientific mainstream that are the contemporary foundation of your topic.
   (c) Write in one paragraph the motivation for your intended research, indicating the "puzzles" you are going to solve; avoid questions which are incorrect, remember that "an approximate answer to the right question is worth a great deal more than a precise answer to the wrong question."
   (d) Identify the fundamental idea that "drives the intellectual content" of your research topic, along with the major research hypothesis.
   (e) Determine the key explained variable in your model as well as the principal explanatory variables; consider the possibilities of endogeneity and the viable remedies.
   (f) Carefully explore the availability and the quality of your data.
   (g) Discuss the major steps in your research schedule and the choice of (quantitative) methodology.
   (h) Clarify again the originality of the intended research and how you will argue that it is really novel.
   (i) Answer the questions: So what? Why is your topic sensible? How will major decisions, behavior, activity, etc. be influenced by the outcome of this research?

2. Once you decide to apply an econometric/microeconometric model, remember the following[11]:
   (a) Propose your model employing relevant theory (if possible), make use of the results of other researchers, and apply a lot of common sense.
   (b) If you employ microdata, do not forget that they have a low level of aggregation, due to which classical linear relations are rarely applied; for

[10]Following Faff (2017) and Faff et al. (2017); also, Kennedy (2002) and Hyndman (robjhyndman.com/).

[11]Following Kennedy (2002); Gruszczyński (2012b).

example, this is the reason why goodness-of-fit measures, like R-squared, have low values. Possible substantial heterogeneity of units (companies) should be taken into account.

(c) Samples of microdata in corporate finance and accounting researches are typically not random. Do not avoid questions about sample biases. Most research uses samples of large public companies, commonly the best on the market. Remember that the results of your investigation pertain only to the companies in the sample.

(d) Use the rule of KISS: *Keep It Sensibly Simple* (but not, *Keep It Simple, Stupid*). While this rule of simplicity may mean many things (see Kennedy 2002), we advocate the following: do not include too many explanatory variables in your model (e.g., if you have five profitability ratios, use only one or two). Your explanatory variables should have some merit, so avoid adding new variables unless you are sure that they belong to the model—otherwise, you are quickly pursuing a data mining exercise.

(e) In most models, including nonlinear ones, the explanatory variables enter in the form of a linear combination (i.e., as the sum of products of each variable by its parameter). So, it is advisable that in such linear combination we minimize the effects of multicollinearity (i.e., high mutual correlation of explanatory variables), which might be done, for example, by selecting for the model variables that are not much correlated with each other.

(f) Make sure that the estimation result can be sensibly interpreted: that the signs of parameter estimates are as expected or fit to the theory (are sometimes the same as signs of simple correlations[12]), that the variables which are important are also statistically significant.[13] Do not forget that small values of estimates are not a symptom of their "lesser validity" (all is decided by statistical tests).

(g) Do not forget that in order to show causality, we need to use special techniques (see Sects. 2.8, 2.9, 2.10, 2.11, and 2.12); otherwise, the only interpretation will be using the terms "correlation" and "association."

(h) Use the potential of data mining carefully. Modeling in econometrics begins with theory as the inspiration for specifying model equations. Once we employ data mining in order to find the model that "best fits to the data," then the result may not be correct, especially when appropriate theory exists; however, data mining exercises may reveal regularities that can "be seen in the data," and this may constitute a good hint for further modeling. Sometimes the modeling goal is not about revealing relationships between explanatory and explained variables but about predicting (*ex post*) the values of the explained variable, in which case, data mining techniques are also appropriate.

[12] Gruszczyński (2012a, p.82).

[13] However, bear in mind that the *p*-value "is not the king" (as discussed in Sect. 2.7).

3. Quantitative research in corporate finance and accounting as reported in published papers—especially in top quality journals—is scrutinized by reviewers. The following is based on the set of "good practices" formulated by Prof. R. Adams (2017)[14]:

    (a) Remember your assumptions: Most papers are rejected because of critical reviews concerning the choice of instrumental variables or the treatment effects approach with no discussion regarding the assumptions necessary for model identification. Another key element is the proper description of the institutional setup concerning the investigated issue of corporate finance or accounting (e.g., the legal framework).
    (b) The clarity and transparency of the text are as important as the research result itself. Follow the literature and avoid preparing an article already written by someone else. At the same time, do not believe everything which has been published, question what you read. Do not forget about the visualization of your thoughts—essential when you use Diff-in-Diff and regression discontinuity techniques.
    (c) If you intend to demonstrate causal effects, it is necessary to show a relevant strategy for identifying those effects—which is not simple. Most papers ignore such discussion, while others employ incorrect identification techniques (e.g., the popular technique of Diff-in-Diff is usually not very appropriate—on corporate governance topics, for instance, it is difficult to find a setup corresponding to a proper medical-type experiment). Also, matching techniques are usually not applied correctly. Obviously, neglecting discussion about causality is not advisable; if identification of causal effects is not feasible then the best solution is to apply regression and correlation techniques—with deep analysis of possible biases in the results—for this, Prof. Adams advocates a paper by Miller (2013).
    (d) A good dataset is crucial for good results. For researchers in corporate finance and accounting, a common headache is triggered by various deficiencies like holes in the data, limited availability of data for "soft" variables, not unified data for quantitative variables, data incomparable between companies, etc. On the other hand, the availability of more data is not always positive for modeling (Prof. Adams points out that in some corporate governance research the cross-sectional data fare much better than panel data).
    (e) It is advisable to consider all "classic" econometric topics including correction for heteroscedasticity, use of fixed effects (if possible), OLS as the benchmark for more advanced techniques, etc.; however, the question of the statistical significance of the variables is not very important since for large datasets it is almost always assured—more important is the interpretation and the feasibility of the result.

[14] Adams (2017); Prof. Adams reports her hints and suggestions in the form of "Adams' alphabet" from A to Z, of which we have presented here only a few select items.

(f) The replicability of your outcomes is important, as journals are increasingly requiring that your dataset be supplied along with the paper—sometimes, to be placed online.

Finally, it is worth mentioning Winston Churchill and his confidential memo, dated August 09, 1940, entitled *Brevity* (to be found online). It is devoted to eliminating unnecessary jargon and concentrating on the intended message: "discipline on setting out the real points concisely will prove an aid to clearer thinking."

* * *

The topics presented in this chapter have focused on examining both the practical and the theoretical questions of applying econometric techniques in corporate finance and accounting research based on microdata. We have introduced a range of microeconometric models and techniques, with detailed examples of relevant applications. Emphasis has also been on methodology that may be of help in studying causal effects in corporate finance and accounting.

Research projects in corporate finance and accounting applying the techniques of microeconometrics are exposed to many risks, most of which are connected with uncertainties about the relevant methodological approach. To mitigate such risks, it is advisable to use recommended "good practices" to avoid unnecessary efforts that may lead to inaccurate results.

The remainder of the book consists of four chapters devoted to financial microeconometrics in bankruptcy research, corporate governance, applied accounting, among other topics.

References

Adams RB (2017) The ABCs of empirical corporate (governance) research. Corp Gov Int Rev 25:461–464

Al-Malkawi H-AN (2007) Determinants of corporate dividend policy in Jordan: an application of the tobit model. J Econ Admin Sci 23:44–70

Anderson JA (1972) Separate sample logistic discrimination. Biometrika 59:19–35

Angrist JD, Pischke J-S (2009) Mostly harmless econometrics. Princeton University Press, Princeton, NJ

Angrist JD, Pischke J-S (2015) Mastering 'metrics: the path from cause to effect. Princeton University Press, Princeton, NJ

Angrist JD, Pischke J-S (2017) Undergraduate econometrics instruction: through our classes, darkly. J Econ Perspect 31(2):125–114

Ardic OP, Yuzereroglu U (2006) A multinomial logit model of bank choice: an application to Turkey. Working Paper ISS/EC/2006-02. Department of Economics, Bogazici University, Istanbul

Atanasov V, Black B (2016) Shock-based causal inference in corporate finance and accounting research. Critic Finan Rev 2016(5):207–304

Beekes W, Brown PR, Chin G (2007) Do better-governed firms make more informative disclosure? Canadian Evidence. Available at SSRN: https://ssrn.com/abstract=881062

Brown P, Beekes W, Verhoeven P (2011) Corporate governance, accounting and finance: a review. Account Finance 51:96–172

Caliendo M, Kopeinig S (2008) Some practical guidance for the implementation of propensity score matching. J Econ Surv 22(1):31–72
Cameron AC, Trivedi PK (2005) Microeconometrics: methods and applications. Cambridge University Press, New York
Card D, Krueger AB (1994) Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania. Am Econ Rev 84(4):772–793
Carey PJ, Geiger MA, O'Connell BT (2008) Costs associated with going-concern modified audit opinions: an analysis of the Australian audit market. Abacus 44(1):61–81
Chaouani S (2010) Using propensity score matching and estimating treatment effects: an application to the post-issue operating performance of French IPOs. Int Res J Financ Econ 48:73–93
Chava S, Roberts M (2008) How does financing impact investment? The role of debt covenant violations. J Financ 63:2085–2121
Cheung A, Hu M, Schwiebert J (2016) Corporate social responsibility and dividend policy. Account Finance 58(3):787–816
Cramer JS (1999) Predictive performance of the binary logit model in unbalanced samples. The Statistician 48(Part 1):85–94
Erel I, Brandon J, Kim W, Weisbach M (2009) Market conditions and the structure of securities. NBER Working Paper No. 14952
Esteve-Pérez S, Mañez-Castillejo JA (2008) The resource-based theory of the firm and firm survival. Small Bus Econ 30:231–249
Faff R et al. (2017) Increasing the discoverability of non-English language research papers: a reverse-engineering application of the pitching research template. Available at SSRN: https://ssrn.com/abstract=2948707
Gao Y, Jamal K, Liu Q, Luo L (2013) Does reputation discipline big 4 audit firms? CAAA annual conference 2011, University of Alberta, School of Business Research Paper No. 2013-1006
Gippel J, Smith T, Zhu Y (2015) Endogeneity in accounting and finance research: natural experiments as a state-of-the-art solution. Abacus 51(2):143–168
Greene WH, Hensher DA (2010) Modeling ordered choices: a primer. Cambridge University Press, Leiden
Gruszczyński M (2001) Modele i prognozy zmiennych jakościowych w finansach i bankowości [modeling and forecasting qualitative variables in finance and banking]. Oficyna Wydawnicza SGH, Warszawa
Gruszczyński M (2006) Corporate governance and financial performance of companies in Poland. Int Adv Econ Res 12(2): 251–259. Also available as Gruszczyński M (2005) Corporate governance and financial performance of companies in Poland. Working Paper No. 2-05, Department of Applied Econometrics, SGH Warsaw School of Economics
Gruszczyński M (2007) Corporate governance ratings and the performance of listed companies in Poland. Working Paper No. 4-07, Department of Applied Econometrics, SGH Warsaw School of Economics
Gruszczyński M (2008) Financial microeconometrics in corporate governance studies. FindEcon. Forecasting financial markets and economic decision-making, no. 6: 11–17. Also available as Gruszczyński M (2010) Financial microeconometrics in corporate governance studies. Working Paper No. 07-10, Department of Applied Econometrics, SGH Warsaw School of Economics
Gruszczyński M (2012a) Empiryczne finanse przedsiębiorstw. Mikroekonometria finansowa [Empirical corporate finance. Financial microeconometrics]. Difin, Warszawa
Gruszczyński M (ed) (2012b) Mikroekonometria, 2nd edn. Wolters Kluwer, Warszawa
Gruszczyński M (2017) Błędy doboru próby w badaniach bankructw przedsiębiorstw (sample bias in the research on corporate bankruptcy). Kwartalnik Nauk o Przedsiębiorstwie 44(3):22–29
Gruszczyński M (2018a) Financial microeconometrics as research methodology in corporate finance and accounting. In: Dudycz T, Osbert-Pociecha G, Brycz B (eds) Efficiency in business and economics. Springer proceedings in business and economics. Springer, New York, pp 71–80

Gruszczyński M (2018b) Good practices in empirical corporate finance and accounting research. J Bank Financ Econ 2(10):45–51

Gruszczyński M (2019) On unbalanced sampling in bankruptcy prediction. Int J Financ Stud 7 (2):28

Hausman J, Lo A, MacKinlay C (1992) An ordered probit analysis of transaction stock prices. J Financ Econ 31:319–379

Heckman JJ (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Ann Econ Soc Meas 5:475–492

Hensher DA, Jones S (2008) Mixed logit and error component model of corporate insolvency and bankruptcy risk. In: Jones S, Hensher DA (eds) Advances in credit risk modeling and corporate bankruptcy prediction. Cambridge University Press, Cambridge, pp 44–79

Hirk R, Hornik K, Vana L (2017) Multivariate ordinal regression models: an analysis of corporate credit ratings. Research report series/Department of Statistics and Mathematics, 132. WU Vienna University of Economics and Business, Vienna

Kamstra M, Kennedy P, Suan TK (2001) Combining bond rating forecasts using logit. Financ Rev 37(2):75–96

Kennedy PE (2002) Sinning in the basement: what are the rules? The ten commandments of applied econometrics. J Econ Surv 16:569–589

King G, Zeng L (2001) Logistic regression in rare events data. Polit Anal 9(2):137–163

Lennox CS, Francis JR, Wang Z (2012) Selection models in accounting research. Account Rev 87:589–616

Leuz C, Lins KV, Warnock FE (2010) Do foreigners invest less in poorly governed firms? Rev Financ Stud 23(3):3245–3285

Li X (2004) Three applications of propensity score matching in microeconomics and corporate finance; US internal migration; seasoned equity offerings; attrition in a randomized experiment. Unpublished doctoral dissertation, Ohio State University

Li F (2016) Endogeneity in CEO power: a survey and experiment. Invest Anal J 45(3):149–162

Li K, Prabhala NR (2007) Self-selection models in corporate finance. In: Eckbo BE (ed) Handbook of corporate finance: empirical corporate finance, North-Holland handbook of finance series, vol 1. Elsevier, Amsterdam, pp 37–86

Long JS (1997) Regression models for categorical and limited dependent variables. Sage, Thousand Oaks, CA

Maddala GS (1983) Limited dependent and qualitative variables in econometrics. Cambridge University Press, Cambridge

Maddala GS, Lahiri K (2009) Introduction to econometrics, 4th edn. Wiley, Chichester

Matvos G, Ostrovsky M (2008) Cross-ownership, returns, and voting in mergers. J Financ Econ 89:391–403

Miller MK (2013) The uses and abuses of matching in political science. Working Paper, George Washington University. Available at https://sites.google.com/site/mkmtwo/research

Mizen P, Tsoukas S (2012) Forecasting US bond default ratings allowing for previous and initial state dependence in an ordered probit model. Int J Forecast 28(1):273–287

Moeller SB, Schlingemann FP, Stulz RM (2004) Firm size and the gains from acquisitions. J Financ Econ 73:201–228

Mora N (2006) Sovereign credit ratings: guilty beyond reasonable doubt? J Bank Financ 30:2041–2062

Roberts MR, Whited TM (2013) Chapter 7: Endogeneity in empirical corporate finance. In: Constantinides G, Harris M, Stulz R (eds) Handbook of the economics of finance, vol 2. North Holland, Amsterdam, pp 493–572

Rosenbaum PR (2001) Observational studies: overview. In: Smelser NJ, Baltes PB (eds) International encyclopedia of the social and behavioral sciences. Elsevier, Amsterdam

Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for casual effects. Biometrika 70:41–55

Shipman JE, Swanquist QT, Whited RL (2017) Propensity score matching in accounting research. Account Rev 92(1):213–244
Skogsvik K, Skogsvik S (2013) On the choice-based sample bias in probabilistic bankruptcy prediction. Invest Manag Financ Innovat 10(1):29–37
Strawiński P (2007) Przyczynowość, selekcja i endogeniczne oddziaływanie. Uniwersytet Warszawski, Wydział Nauk Ekonomicznych (unpublished)
Strawiński P (2014) Propensity score matching. Uniwersytet Warszawski, Warsaw
Sun J, Cahan S (2009) The effect of compensation committee quality on the association between CEO cash compensation and accounting performance. Corp Gov 17:193–207
Tucker JW (2010) Selection bias and econometric remedies in accounting and finance research. J Account Lit 29:31–57
Wasserstein RL, Lazar NA (2016) The ASA's statement on p-values: context, process, and purpose. Am Stat 70(2):129–133
Winkelmann R, Boes S (2006) Analysis of microdata. Springer, Heidelberg
Wooldridge JM (2019) Introductory econometrics: a modern approach, 7th edn. Cengage, Boston, MA
Ziliak S, McCloskey D (2008) The cult of statistical significance: how the standard error costs us jobs, justice and lives. University of Michigan Press, Michigan
Zmijewski M (1984) Methodological issues related to the estimation of financial distress prediction models. J Account Res 20:59–82

Chapter 3
Modeling Financial Distress and Bankruptcy

Modeling bankruptcy and financial distress is the most common area of statistical–econometric research in empirical corporate finance. Information regarding financial distress, bankruptcy, and other types of cessation of company activities is of critical importance for owners of equity, management, lenders, investors, and other stakeholders. In this chapter, we try to enumerate the essential issues of microeconometric modeling in this field, including the selection of predictions, unbalanced samples, modeling for financial distress versus bankruptcy, as well as modeling firm exit and firm survival.

3.1 Research on Corporate Financial Distress and Bankruptcy

Fifty Years of the Altman Z-Score

The first attempts to employ statistical approaches to modeling and predicting bankruptcy date back to 1960s. It might be argued that the jump start of methodological innovations in the field was the paper of Edward Altman in the *Journal of Finance* in 1968 (Altman 1968). It was preceded by an equally statistically sound paper by William H. Beaver published in the *Journal of Accounting Research* in 1966 (Beaver 1966). Altman's approach involves expressing financial distress with the multivariate model; Beaver uses a set of single variables for the same purpose.

During the ensuing years, the field has grown immensely—in terms of methodology, accessibility of data, and applicability. The chief reason is the skyrocketing demand for assessments of corporate standing and predictions of corporate failure. As a result, it has become the field where statistics, econometrics, data science, and corporate finance join in striving for the best methodologies. The demand for

M. Gruszczyński, *Financial Microeconometrics*,
https://doi.org/10.1007/978-3-030-34219-7_3

predictions is both internal and external to the firm. Referring to the current fate of his models, Altman (2018) states

> The applications have taken two forms: (1) Those applied by analysts external to the firm, primarily for credit and investment purposes, and; (2) Those applied by managers and board members from within the distressed firm to gauge its strengths and weaknesses and, in some cases, to help guide the firm to a successful financial turnaround.

This statement is also true for other methodological approaches.

Researchers into financial distress rarely explore the underlying concept that stems largely from the theory of the enterprise and agency theory. This topic is expanded by Hotchkiss et al. (2008), for example. Typical research into financial distress concentrates on verifying the hypotheses based on intuition and/or the results of other researchers. This is understandable because the results are usually specific to place, scope, and time period. However, in distress and bankruptcy research, it is common to apply research outcomes to many markets and countries (e.g., Altman Z-score).

The major goals of research into financial distress are as follows:

- Search for the determinants of financial distress and/or bankruptcy.
- Prediction of the state of financial distress.
- Composition of investment portfolios with the use of distressed stocks.

Bankruptcy/insolvency and the preceding phases of financial distress are of utmost interest to the company itself and to parties outside the company, such as lenders, regulators, and analysts. Research-wise, the topics belong to management, finance, accounting, economics, as well as to civil and administrative law. Thus, we have the "real" framework represented by the distressed company, its management, and equity owners. We also have the "research" framework that includes a plethora of methodologies competing for the best predictions of the company's fate. Then there is the "legal" framework within which the fate of a bankruptcy filing is settled. Finally, the "state/public" framework of governments and regulators is concerned with aggregate distress and bankruptcy in a region/country.

The foregoing are the reasons underlying the vast and multidimensional library of literature on the topics of company distress and bankruptcy that has emerged over the years and is constantly expanding worldwide. The most apparent feature of such studies is the application of quantitative statistical methods, yet journals of statistics and econometrics rarely allocate space to findings regarding financial distress. The major outcomes are published in journals of accounting and of corporate finance. Several titles have already been mentioned in Chap. 1, including repositories like the SSRN e-journals. Among them, it is worth noting the "Econometric Modeling: Microeconometric Models of Firm Behavior eJournal."

Survey papers on research into bankruptcy and distress appear regularly in books and journals. Examples are

- Hotchkiss et al. (2008)—A survey on the empirical research into the use of mechanisms for resolving default and reorganizing companies in financial distress.

- Aziz and Dar (2006)—A survey into the methodologies applied in ten countries.
- Senbet and Wang (2010)—A survey made from a legal perspective.
- Linden (2015)—A survey into the methodology of selecting variables for bankruptcy models.
- Prusak (2018)—A survey on bankruptcy prediction research in Central and Eastern European countries.

The books of Altman et al. (2019) and Beaver et al. (2010) should also be mentioned as references.

Some 50 years after the works of Beaver (1966) and Altman (1968), the topics of financial distress and predicting bankruptcy should have become staple components of corporate finance textbooks and courses. In terms of methodology, it seems that the finance profession in academia still does not recognize them as basic content in core corporate finance and accounting courses. The notable exceptions are textbooks by Damodaran (*Applied Corporate Finance,* 4th ed., 2014), Berk and DeMarzo (*Corporate Finance,* 4th ed., 2017), and Ross et al. (*Corporate Finance*, 11th ed., 2015). Instruction in corporate finance and accounting should incorporate the notion that some rules and hypotheses might only be evidenced by research on large groups of companies. Such research and the tools derived from it are much advocated in regard to a company's financial distress/bankruptcy as well as the going-concern audit opinion.

Nevertheless, it is essential to recognize that aspects of distress/bankruptcy are studied in many disciplines: finance, accounting, economics, management, and law. The latter may be evidenced, for example, in the paper by LoPucki and Doherty (2015) as well as the surveys by Hotchkiss et al. (2008) and Senbet and Wang (2010). In this chapter, we concentrate on both the financial-managerial aspects and on the research methodology of distress and bankruptcy.

The Notion of Financial Distress

Financial distress (FD) is not easy to define. Ross et al. (2015) list examples of events that might evidence the financial distress of a company: dividend reduction, plant closings, losses, layoffs, CEO resignation, or a plummeting stock price. The authors state that "financial distress is a situation where a firm's operating cash flows are not sufficient to satisfy current obligations . . . and the firm is forced to take corrective action."

Corporate finance and accounting define financial distress in various ways. While everyone may think they understand what financial distress or corporate insolvency is, when it comes to a precise definition, the result is likely less than satisfactory. Platt and Platt (2006) state that the "definition of financial distress is less precise than the legal actions that define proceedings such as bankruptcy or liquidation; despite this uncertainty, it is clear that the condition of being financially distressed deviates from corporate normality in a manner similar to bankruptcy." The authors identify the following symptoms of FD:

- Layoffs, restructurings, missed dividend payments.
- Low interest coverage ratio.
- Cash flow less than the current maturities of long-term debt.
- Change in the equity price or negative EBIT.
- Negative net income before special items.

Information about a company's financial health is crucial for many stakeholders.

Equity Owners: Signs of distress are expressed, for example, in the going-concern audit opinion, which should be publicly disclosed and is of key importance to company owners. The company's market valuation may be affected.

Creditors: Information entering banks' internal rating systems in regard to the FD of companies in their credit portfolios is crucial to the prediction and evaluation of risk in the banks.

Equity Investors: Investors tracking the financial performance of companies have obvious interest in any news regarding possible distress. Their likely strategy would be to reduce positions in such stocks.

Here are some issues relevant to understanding and explaining financial distress.

- Distress lies in the gap between a company's good financial health and bankruptcy.
- Distress precedes bankruptcy, although it is not clear that the same factors cause both.
- A financially distressed company may have an unclear future with a significant probability of discontinuation. A bankrupt company terminates its activity under a specific legal form but may sometimes continue with good prospects.
- The category of financial distress is both fuzzy and dynamic. Data on a company's financial state are usually delayed and of little use to investors. Prevailing studies concentrate on a cross-sectional view, while a time series analysis of distress might be better suited for practical purposes.

Does Financial Distress Lead to Bankruptcy?

Various studies indicate the necessity of clearly distinguishing between Yes–No bankruptcy modeling and possibly more than two states in the case of financial distress research. As a result, studies into financial distress and bankruptcy consider two or more states. This means that the explained variable in distress models represents two or more categories.

Here is a selection of studies and the binomial or multinomial variables used to express distress:

- Cheng et al. (2006); Lau (1987)—Five states of increasing severity of financial distress

  0 = financial stability
  1 = omitting or reducing dividend payments
  2 = technical default and default on loan payments
  3 = protection under Chap. 10/11 of the US Bankruptcy Act
  4 = bankruptcy and liquidation

- Campbell et al. (2008, 2011)—Two states

  0 = non-failed firm
  1 = filing for bankruptcy (Chap. 7/11), delisting for performance-related reasons, receiving a D rating from a rating agency

- Dahiya et al. (2003)—Two states

  0 = non-failed firm
  1 = a firm is financially distressed if it has insufficient cash flow to meet its debt payments; two types of FD announcement—(1) default on a firm's public debt, and (2) a firm's filing for bankruptcy protection under Chap. 11

- Platt and Platt (2006)—Two states

  0 = non-failed firm
  1 = financially distressed firm—meets all the following criteria in two consecutive years: negative EBITDA interest coverage, negative EBIT, and negative net income before special items

- Hensher and Jones (2008)—Four states

  0 = non-failed firms
  1 = insolvent firms—(1) failure to pay Australian Stock Exchange (ASX) annual listing fees; (2) a capital raising specifically to generate sufficient working capital to finance continuing operations; (3) loan default; and (4) a debt/total equity restructure due to diminished capacity to make loan repayments
  2 = financially distressed firms—delisted from ASF due to being subject to a merger or takeover arrangement
  3 = firms filed for bankruptcy—followed by the appointment of liquidators, insolvency administrators, or receivers

- Gruszczyński (2004)—Three states

  0 = non-failed firm
  1 = firm with undetermined financial condition
  2 = financially distressed firm

The studies on bankruptcy plainly identify two states: "the bankrupt firm" and "the non-bankrupt firm," which differ according to the specific definition of the bankrupt firm entering each sample. Without going into detail, we should emphasize that the prevalent methodologies applied in bankruptcy studies (the binomial model, multivariate discriminant analysis, etc.) also provide the possibility of calculating "many shades of grey"—i.e., predictions/probabilities indicating the degree of financial distress—between financially healthy and bankruptcy.

Going-Concern Opinions

The first indication of financial distress leading to bankruptcy may come from the auditors of a company's financial records in the form of the going-concern opinion. Gerakos et al. (2016) used US data from 2000 to 2011 with 72,580 client-year observations including 794 bankruptcies and 11,696 going-concern opinions. Their dataset had, therefore, 1.1% bankruptcy filings and 13.1% going-concern issuances. The key finding is that "going-concern opinions do not predict more bankruptcies than a statistical model based solely on observable data, which suggests that auditors do not efficiently use information when generating going-concern opinions." This may suggest that going-concern predictions generated from bankruptcy models (in this case, the binomial probit) are more accurate and will assist auditors in preparation of their reports.

In the next sections, we discuss questions of FD and bankruptcy modeling using microeconometric methods. We also elaborate on problems that are common to various methodological approaches—i.e., selection of predictor variables and sample composition.

3.2 Microeconometric Models of Bankruptcy and Financial Distress

Methods for Predicting Bankruptcy/Distress

The methodology employed in financial distress research now covers almost all techniques of data analysis, specifically methods of statistics, econometrics, survival analysis, and data mining. A common feature of most approaches is their probability focused nature, as it is customary to express financial distress in terms of the probability of corporate failure.

Nowadays, methodology might be split into the classic stream and the new methods of advanced data analysis. The classic stream includes

- The Altman Z-score (1968)
- Ohlson's O-score (1980)

- Moody's KMV[1] model based on Merton (1974)
- The Shumway hazard model (2001)

The Altman Z-score uses linear discriminant analysis (LDA), while Ohlson's O-score is based on the logit model. From the above list, the first two are accounting-based models, while the latter two are market based.[2] All four have been followed by numerous researchers worldwide and, despite having been subjected to many modifications by both their original creators and other researchers, they are still regarded as the most popular approaches in distress modeling.

Some interesting new methods have originated from modeling fraud detection, from the methodology of text mining, and from modeling companies' churn for loans. New approaches of advanced data analysis have established a significant alternative to the classic stream in this field.

Data-driven models (e.g., machine learning models) that have recently been used in predicting financial distress and bankruptcy include

- Support vector machines (SVMs) (Shin et al. 2005)
- Bagging, boosting, random forest, neural networks, SVMs (Barboza et al. 2017)
- Generalized boosting, AdaBoost, and random forest (Jones et al. 2017)
- Ensemble boosted trees (Zięba et al. 2016)
- Ensemble learning (Choi et al. 2018)

In machine-learning vocabulary, the above methods are called "classifiers," which can also be used to describe models such as LDA or logit.

A popular idea for research is comparing the results of bankruptcy predictions derived by various methods from both streams—applied to the same dataset on bankruptcy. Results differ between studies and datasets. For example, Jones et al. (2017) state that "simple classifiers such as logit and LDA perform reasonably well in bankruptcy prediction; however, we recommend the use of 'new age' classifiers." The reasons are better prediction in samples used in that study, the ease of estimating and implementing—"with minimal researcher intervention"—and a "relatively good level of interpretability." Barboza et al. (2017) find that machine learning models have roughly 10% more accuracy relative to traditional models (for their sample). Nehrebecka (2018) compared the logit with SVM models and found that SVM performed better for the training sample and logit performed better for the validation sample. Studies by Zięba et al. (2016) and by Choi et al. (2018) confirm that the use of "ensembles" of classifiers for bankruptcy prediction gives better results than the use of single classifiers. This outcome is similar to that obtained when comparing methods of forecasting time series. For example, in the last M4 competition (in 2018), the best results were achieved with the use of combined forecasts and hybrid methodologies (Makridakis et al. 2018).

[1] Moody's (2000).

[2] The distinction by Outecheva (2007).

In this book, we concentrate on the classic stream as representing the methodology of microeconometrics. However, all methods applied for modeling and predicting distress/bankruptcy experience similar problems, at least in two respects:

- Sample composition (the question of unbalanced sampling)
- The choice of predictor variables (explanatory variables, input variables)

These issues are considered in more detail later in this chapter.

The Choice of Predictors and the Drawbacks of Modeling

All attempts to model and predict financial distress and bankruptcy rely on the selection of appropriate predictors (i.e., covariates/explanatory variables). Predictors are usually chosen from financial ratios calculated using companies' financial statements. For example, the classic set of predictors in the Altman (1968) Z-score includes:

X_1 = working capital/total assets (liquidity ratio)
X_2 = retained earnings/total assets (profitability/leverage ratio)
X_3 = EBIT/total assets (productivity ratio)
X_4 = equity market value/book value of total debt (solvency ratio)
X_5 = sales/total assets (activity ratio).

Throughout this chapter, we present many models and many predictor variables. As in all econometric/regression research, there is no golden rule for finding the best set of explanatory variables. In corporate finance and accounting, it is customary to narrow the choice to company finances, the market situation, sometimes also to corporate governance issues. The factors for predicting financial distress are commonly represented by variables calculated from data in the financial reports, market data, and ownership- and industry-specific information.

Lennox (1999) and Kaiser (2001) specify, for example, that the following predictors be considered in explaining financial distress/bankruptcy:

- *Unprofitability*—The more unprofitable the company, the higher the probability of failing.
- *Debt*—The higher the debt, the higher the probability of default.
- *Cash flow difficulties*—A company with healthy cash flow has relatively easy access to external financing.
- *Age of the firm*—During the initial growth period the chance of failure increases, the medium age period sees stable probability of default, and afterwards the chance of failure decreases.
- *Size of the firm*—An inverse U-shaped effect on the probability of moving into/out of financial distress.
- *Legal status*.

- *Corporate shareholder*—The existence of corporate shareholders has a negative effect on the probability of moving into financial distress.
- *Multiple creditors*—Firms with multiple creditors are less likely to run into financial distress.
- *Diversification*—Diversified firms are less likely to move into financial distress.

The choice of an "appropriate" set of predictors is of major importance in distress and bankruptcy modeling. Various techniques of econometrics and advanced data analysis compete for the best classification of observations in-sample and out-of-sample. On the other hand, we believe that economics and finance research should try to avoid a mechanical approach to selecting predictors. When researching distress/bankruptcy with the use of financial ratios, researchers are encouraged to consider the following:

- Each group of financial ratios (e.g., profitability, liquidity, operational efficiency, leverage) includes many ratios that are strongly correlated. The standard model should include at most 1–2 ratios from each group.
- Market ratios should be carefully connected with accounting ratios in regard to the time frame.
- Incremental ratios (e.g., the percentage increase of sales) introduce dynamics that should be methodologically recognized.
- Ratios are useful for comparing firms of various sizes in the numerator and the denominator; however, they are sometimes overused (e.g., by comparing to the industry average).
- Pre-classification of ratios as "good" or "bad" to explain financial distress should be done with great caution. Distress can be defined, although vaguely; therefore, the prior assumption that some variables should be considered as "for" or "against" might not be valid. The result is always sample specific.

These recommendations may be added to a list of drawbacks formulated by Balcaen and Ooghe (2004) in their survey on the application of quantitative approaches to modeling bankruptcy.

- Most distress/bankruptcy studies use an initial set of explanatory variables, often chosen arbitrarily "on the basis of their popularity in the literature and their predictive success in previous research." This is due partly to the lack of a sound theory, which may serve as the theoretical basis for variable selection. The most popular techniques for selecting variables are based on statistical considerations. Consequently, the final set of variables depends on the sample and, therefore, is sample specific and unstable.
- Bankruptcy models use financial ratios calculated from the annual financial statements. In many countries, the requirement to publish these statements is restricted to companies that meet certain criteria (e.g., over a specified size), and some companies are, therefore, excluded from samples. Another problem is that the researcher may assume that annual accounts give fair and true views of the financial state of companies, when there is strong global evidence that this may

not be the case, especially for failing companies. Generally, any deficiencies in the financial statements are transferred to the samples.

- The ill-defined dichotomy of corporate failure. The classification of companies into failing and non-failing populations is arbitrary, because of the various definitions of failure and the way in which the selected one is applied. A researcher may use, for example, the legal definition of bankruptcy, an economic definition of failure, a set of default events (as utilized in credit risk analyses), the concept of financial distress (in lieu of failure), etc. Consequently, the assumption of the dichotomous dependent variable may be violated.
- Nonstationary and unstable data. If the bankruptcy model is to be used for prediction, then the question of its stability over time gains importance. Bankruptcy/failure literature uses the term nonstationarity or the instability of data for those situations in which the relationship between a model's variables is not stable over time and the values of the independent variables differ markedly between the estimation period and the forecast period. The unstable data situation results in the model's poor predictive ability and calls for frequent re-estimation of the model. The assumption of stability over time is also important for "pooled samples" of companies failing in different years, which is common practice in bankruptcy modeling.

Balcaen and Ooghe (2004) also comment on nonrandom sampling and sampling biases, which are discussed later in this chapter. The catalogue of drawbacks may be augmented by other problems of bankruptcy modeling in emerging economies—e.g., in Poland (Gruszczyński 2005):

- The limited availability of sound financial data.
- The lack of nonfinancial "soft" variables in bankruptcy models.
- The heterogeneity of samples in terms of time span, size of companies, and industry sector.

The above recommendations and warnings about distress and bankruptcy modeling may supplement the set of good practices presented in Chap. 2.

Comparing Financial Distress and Bankruptcy Models

As stated in Sect. 3.1, the variable that is to be explained in microeconometric models for financial distress may be binomial or multinomial. For bankruptcy modeling, however, the only choice is a binomial variable. Therefore, the dependent variables are:

1. Financial distress model

$y_i = 1$ The company is financially distressed ("severe problem" company)
$y_i = 0$ The company is financially sound ("no problem" company)

or (for example)

$y_i = 1$ The company is financially distressed
$y_i = 2$ The financial condition of the company is undetermined
$y_i = 3$ The company is financially sound

2. Bankruptcy model

$y_i = 1$ The company is bankrupt
$y_i = 0$ Company is non-bankrupt

Example 3.1 Predicting Financial Distress and Bankruptcy for US Companies
In their study comparing FD and bankruptcy models, Platt and Platt (2006) considered a sample of companies that are financially distressed but not yet bankrupt. The distressed companies are those with negative EBITDA covering interest expense, negative EBIT, and negative net income before special items. There were 276 such US companies in the Compustat 1999–2000 database (all three items were negative in both years). The remaining 1127 companies comprise a subsample of non-financially distressed companies.

Part A. For modeling FD, the binomial logit is applied with the dependent variable equal to 1 for a distressed firm and equal to 0 for a non-distressed firm.

The explanatory variables (predictors of distress) were carefully chosen, initially comprising one variable from eight specified groups of financial ratios. The expectation was that financial distress would be negatively related to profit margin, profitability, liquidity (cash position), growth, and operating efficiency, and would be positively related to operating or financial leverage.

After developing the core group of predictors, the additional variables were added only when they yield the coefficient with the expected sign, statistical significance, and improved classification accuracy. The final set includes five variables representing profit margin, profitability, financial leverage, and liquidity:

1. CF/Sales = Cash flow/sales, a measure of profit margin
2. EBITDA/TA = Earnings before interest, tax, depreciation, and amortization/total assets—a measure of operating profitability
3. Current LTD due/TA = The current portion of long-term debt due/total assets—a measure of leverage
4. TIE = Times interest earned [basically, earnings before tax/interest expense]—a measure of leverage
5. QR = Quick ratio [(current assets − inventories)/current liabilities]—a measure of liquidity.

All company ratios were transformed into industry-relative ratios with the formula: company ratio divided by the mean ratio in the industry. Predictor variables from 1999 were used to explain financial distress in 2000. In other words, the estimated model is of the form

$$probability\ \left(y_{i,2000} = 1\right) = logit\left(predictor\ variables_{i,1999}\right)$$

where *logit* means that we use a binomial logit model of the form introduced in Chap. 2 (Eq. 2.5) to explain the probability.

The signs of the estimated parameters of the logit models are as expected: negative for variables 1, 2, and 4, and positive for variables 3 and 5. Higher cash flows (CF/Sales and EBITDA/TA) and greater times interest earned (TIE) are associated with a lower probability of financial distress. Higher leverage (Current LTD Due/TA) and greater liquidity (QR) are related to a higher probability of distress.[3]

Part B. The second part of the study by Platt and Platt (2006) is devoted to verifying the hypothesis that FD and bankruptcy belong to the same "single on-going corporate decay" process. Therefore, it was expected that they are explained by identical variables. The competing hypothesis is that FD and bankruptcy are different to such an extent that separate variables should explain each.

In order to verify this notion, the set of five predictors selected for the financial distress model was expanded by the following seven predictors of bankruptcy that were selected in another study by Platt and Platt (1991).

1. CF/Sales = Cash flow/sales
2. Short-term debt/total debt
3. Net fixed assets/total assets
4. Total debt/total assets
5. Percent change in sales/percent change in industry output
6. (CF/Sales) × percent change in industry output
7. (Total debt/total assets) × percent change in industry output

The first variable (CF/Sales) is the same as in the financial distress model.

The extended model (with 11 predictors) was estimated for the same sample that was used in the FD exercise. Afterwards, tests for included/omitted variables were applied to test the main hypothesis that the predictors from the bankruptcy model do not "add incrementally" to the predictors from the distress model. With the use of the Davidson-MacKinnon J test, the authors arrive at the conclusion that such hypothesis should be rejected. Apart from some doubts concerning this methodology, the result that the set of financial distress predictors is different from the set of bankruptcy predictors seems appropriate. The authors offer a detailed explanation of this result. We quote only this:

> Bankruptcy is the decision that firms make when they need, for example, to protect their assets from creditors. Financial distress arises when the firm's operating decisions yield less satisfactory results. . . . Taken together, these results suggest that the bankruptcy process is not just a continuation of a downward spiralling cycle toward ultimate corporate failure.

■

The study by Platt and Platt (2006) stresses that the major difference between financial distress and bankruptcy models lies in the separate sets of predictors for

[3]The estimates of logit model parameters in the paper by Platt and Platt (2006) are given as "uniformly scaled" without their actual values; in addition, their model (3) is presented as explaining probability which may be incorrect.

each model. The concept of this research was not followed in other analyses. Tinoco and Wilson (2013) in their paper on financial distress and bankruptcy modeling use the binomial variable[4] in which the firm is classified as financially distressed when it files for bankruptcy or "whenever it meets both of the following conditions: (1) its earnings before interest and taxes, depreciation, and amortization (EBITDA) are lower than its financial expenses for two consecutive years and, (2) there is a negative growth of its market value for two consecutive periods." In this setup, it is not possible to distinguish between predictors of distress and predictors of bankruptcy.

Thus, we may not have a definitive answer to the dilemma on how to distinguish modeling for FD and bankruptcy. As we indicated in the previous section, the result is always sample specific. However, diverse approaches to modeling with innovative sets of predictors are always worth studying and applying.

Multinomial Models

From this point on, we do not distinguish between models dedicated separately to financial distress and to bankruptcy. We begin with multinomial models. In Sect. 3.1, we presented several examples of the explained variable representing various levels of distress. Lau (1987) used the ordered multinomial variable with five states, the last of which represents bankruptcy. Hensher and Jones (2008) applied the unordered multinomial variable with four states.

Example 3.2 Mixed Multinomial Logit: FD of Australian Companies
The four states of the Hensher and Jones (2008) study shown in Sect. 3.1 are (0) non-failed firm, (1) insolvent firm, (2) financially distressed firm (i.e., was subject to a merger or takeover arrangement), and (3) bankrupt firm (i.e., filed for bankruptcy). The states of distress are ordered from healthy to bankrupt company (with "insolvent" in between). This research was presented in Chap. 2 as an example of an ordered multinomial model. The explanatory variables for this model shown in Table 2.3 are excess market return (above market return); (cash + deposits + marketable securities)/ total assets; four consecutive annual periods of negative operating cash flow (yes = 1, = 0 otherwise); EBIT/total assets; working capital/total assets; the log of total assets; the age of the firm (= 1 if the firm was established in the previous 6 years, = 0 otherwise); total debt/gross operating cash flow; and other variables.

Hensher and Jones (2008) propose the use of the standard unordered multinomial logit and the mixed logit (specifically, the error component logit in this case). The authors state that for the mixed logit "the probability of failure of a specific firm in a sample is determined by the mean influence of each explanatory variable with a fixed parameter estimated within the sampled population, *plus*, for any random parameters, a parameter weight drawn from the distribution of individual firm parameters estimated across the sample."

[4] A more recent paper by Tinoco et al. (2018) uses multinomial variable with three states.

The multinomial mixed logit may be derived from the random utility model. It is assumed that firm q in situation (occasion, period, moment) t faces the "choice" of a certain state of financial distress and focuses on strategy leading to the outcome ("choice") with the highest utility (i.e., non-failure). The utility associated with the i-th outcome evaluated by firm q in occasion t is

$$U_{itq} = \boldsymbol{x}'_{itq}\boldsymbol{\beta}_{itq} + \varepsilon_{itq} \tag{3.1}$$

where $\boldsymbol{x}_{itq}$ is the vector of values of the explanatory variables that represent the firm's characteristics, the attributes of the other outcomes, and "descriptors of the decision content on occasion t." Vector $\boldsymbol{\beta}_{itq}$ and the disturbance ε_{itq} are unobserved and treated as stochastic. The error terms ε_{itq} are IID—independent (between outcome alternatives) and identically distributed. Their distribution is the extreme value type 1 (Gumbel distribution: EV1[5]). The entire unobserved heterogeneity is to be represented by parameters $\boldsymbol{\beta}_{itq}$. This restriction is alleviated by the assumption that the error term in the utility equation is the sum of two uncorrelated parts η_{itq} + ε_{itq}

$$U_{itq} = \boldsymbol{x}'_{itq}\boldsymbol{\beta}_{itq} + \left(\eta_{itq} + \varepsilon_{itq}\right) \tag{3.2}$$

where η_{itq} is the part which is corelated over alternative outcomes and heteroscedastic, and another part ε_{itq}, which is IID over alternative outcomes and firms. The density f $(\eta|\,\Omega)$ of η is defined by the fixed parameters Ω (mean, covariances, etc.).

The heterogeneity of firms is now represented by $\boldsymbol{\beta}_{itq}$ and by η_{itq}. In effect, the probability of the i-th outcome conditional on η is given as (ignoring the t subscript)

$$L_i(\eta) = \frac{\exp\left(\boldsymbol{x}'_{iq}\boldsymbol{\beta}_{iq} + \eta_{iq}\right)}{\sum_j \exp\left(\boldsymbol{x}'_{jq}\boldsymbol{\beta}_{jq} + \eta_{jq}\right)} \tag{3.3}$$

This is the formula known as the standard multinomial model but here we have additional information embedded in η_{iq} for each firm in the sample. This information influences the i-th outcome. Since η is not observed, the unconditional probability of the i-th outcome is obtained by integrating overall values of η (weighted by the density of η)

$$p_i = \int L_i(\eta) f(\eta|\Omega)\mathrm{d}\eta \tag{3.4}$$

[5]Compared with a normal distribution, EV1 has "fat tails" (since the extreme values are in tails). For the binomial model the difference of choice probabilities between the normal distribution and EV1 is minor. In the model with many choices, the choice probabilities are relatively low (e.g., 0.01 or 0.02). Then the difference between the normal distribution and EV1 is significant, which is especially visible when aggregating individual probabilities.

In statistics the weighted mean of many functions is called a mixed function and the density of weights is called a mixing distribution. This model is called the mixed logit since it is a mixture of logits with *f* as the mixing distribution (Train 2009).

The mixed logit and other versions of multinomial modeling were attempted by Hensher and Jones (2008) on a sample of non-failed and distressed firms listed on the Australian Stock Exchange (ASX) collected in the years 1992–2004. The full sample was randomly allocated to an estimation (training) sample (A) and a holdout sample (B). The structure of the sample is as follows (number of firm-years):

State 0: A = 1871, B = 2192
State 1: A = 280, B = 242
State 2: A = 41, B = 37
State 3: A = 67, B = 123

The results of application of the authors' models to this sample encouraged the authors to indicate the advantages of using a more general discrete choice model such as the mixed logit. One result is the recognition of the role that various sources of heterogeneity play in influencing choice outcomes. On the other hand, the study confirms that the "greatest overall statistical influence on the failure outcome" is from the predictors that are consistent with the literature, such as firm size, firm age, retained earnings to total assets, cash resources to total assets, total debt to operating cash flow, and excess market return.

■

Another example of the mixed logit application in corporate bankruptcy research is shown in the paper by Kukuk and Rönnberg (2013). The authors use a set of 5000 observations of German firms for the years 2007–2008 and apply the classical binomial logit versus the mixed logit that allows for stochastic parameters. The study is devoted to the analysis of the quality of the estimates and predictions.

Example 3.3 The Ordered Multinomial Logit: FD of Polish Companies
A study by Gruszczyński (2004) applies the ordered trinomial logit for modeling a distress variable that has three states: (1) $y_i = 1$—a financially distressed firm, (2) $y_i = 2$—undetermined financial condition, and (3) $y_i = 3$—a financially sound firm. The database comprised the financial statements of 200 unlisted companies in Poland (1995–1997). The 1995 statements were examined by accounting and legal experts. The final sample includes 23 companies in a poor financial situation (financially distressed), 23 financially sound companies, and an additional 25 companies representing firms in a "medium" financial shape—an inconclusive state between "no problem" and "severe problem."

The key specifications of the attempted models are as follows:

$$probability\ (y_{it} = k) = logit\big(predictor\ variables_{i,t-1}\big) \tag{3.5}$$

$$probability\ (y_{it} = k) = logit\big(predictor\ variables_{i,t-2}\big) \tag{3.6}$$

where $k = 1, 2, 3$ and *logit* means that we use an ordered logit model of the form introduced in Sect. 2.4 (Eq. 2.8) to explain the probability. The term *predictor* *variables* denotes the list of explanatory variables for the models and $t = 1997$. The specification assumes that the financial state of a company in year $t = 1997$ may be determined by its characteristics for $t - 1 = 1996$ and $t - 2 = 1995$.

The predictors were chosen from four groups of financial ratios: liquidity ratios, profitability ratios, activity ratios, and debt management ratios—17 indicators altogether. The models assume a 1 or 2-year lag to explain the 1997 state of the company.[6] The predictors were selected in the following sequence of steps:

1. The financial ratio X explaining the y variable is significantly correlated with y. For the binomial y, the ordinary correlation coefficient with X suffices. For the ordered trinomial y, the correlation is replaced by the chi-square test of independence: the model may only accept the ratios for which the hypothesis of independence (with y) is rejected. The direction (sign) of this association is then determined by a simple yX-correlation coefficient where the y variable is treated as dichotomous (with $y_i = 2$ rejected).
2. Ratios are accepted to the model as explanatory variables only if they are weakly correlated between themselves.
3. The model is accepted only if the sign of the yX-correlation is the same as the sign of the relevant X parameter estimate in the logit model. In such application, this rule is practical and intuitive. In Polish econometric literature, we call it the *principle of coincidence*. It means that once we are sure that increasing values of X are associated with increasing values of y (from 1 to 2 to 3 in a trinomial model), the models we may reasonably accept shall have the positive sign of the parameter's estimate for the X variable. Decreasing values of X associated with increasing values of y shall result in accepting the model with the negative sign of the parameter's estimate for the X variable. For trinomial models, the principle of coincidence is verified by using the yX-correlation coefficient where the y variable is treated as dichotomous (with $k = 2$ rejected).
4. From each group of financial ratios, the model includes just one or two variables. Because the ratios are highly correlated within the group, each selected predictor conveys most of the information from the entire group.
5. The explanatory variables (predictors) included in the model are significant, although this condition is not applied rigorously. An incorrect indication of significance test is possible here due to the multicollinearity of the explanatory variables as well as to the small sample (see the discussion about significance in Sect. 2.7).
6. The model has reasonable ex post predictive capacity. Forecast accuracy is calculated as the share of correct forecasts of Y in the sample. The forecast of y for the firm is the state ("1" or "2" or "3") with the largest probability predicted from the estimated model. In the case of the trinomial ordered model, the probabilities of $y_i = 1$, $y_i = 2$, and $y_i = 3$ are given by formulae (2.9, 2.10, and 2.11) in Chap. 2.

There are more financial ratios significantly correlated/associated with y_{t-1} than with y_{t-2}. This means that the symptoms of financial distress increase in number

[6]The lag length is due to the format of the financial data. The legal format of financial statements in Poland before 1995 was significantly different from that in 1995 and thereafter, because of major changes in the law on accounting introduced in 1995.

approaching the year $t = 1997$. All liquidity ratios are positively correlated with y (i.e., companies with higher liquidity have better chances of being financially sound after 1–2 years than companies with liquidity problems). The profitability ratio most frequently chosen for the models is "operating ROA" (return on assets defined with operating profit in the numerator). Profitability is also positively correlated with y: the higher the profits of the company, the higher the probability of staying in good financial shape. The asset management ratios selected for the models are liabilities turnover and inventory cycle. The debt ratios significantly correlated with y are the debt ratio and the ratio of liabilities (adjusted for most liquid assets) to sales. Both correlations are negative: increasing debt is associated with a company's decreasing ability to survive.

The procedure of selecting the best set of predictors resulted in a number of competing models. It was shown that, as in other economies, the financial distress of companies in Poland is determined mainly by the degree of liquidity, profitability, and by the size of debt. The best predictors revealed in this study are:

- The loss of liquidity (liquidity ratio)
- Diminishing profitability (return on assets)
- Increasing debt (debt ratio)
- Decreasing turnover of liabilities

An example of an estimated trinomial model with the 1995 predictors is presented in Table 3.1.

This study indicates that models containing a reasonable collection of 2–3 financial ratios can predict the state of a company's financial distress in Poland after 1 or 2 years. The precision of such forecast lies in the range of 85–90%. According to the rules described above, the selection of variables resulted in obtaining a good number of prediction models with acceptable statistical and economic properties. Models with predictors from year $t - 2$ perform worse than

Table 3.1 Financial distress in companies in Poland

Predictor	Estimate	*SE*	*t*-statistic	Prob.
Quick ratio	1.2654	0.4804	2.6340	0.0084
ROA	1.4402	1.6272	0.8851	0.3761
Debt ratio	−2.6851	1.4980	−1.7925	0.0731
Cut points				
τ_1	−0.6002	0.9092	−0.6602	0.5091
τ_2	1.5527	0.9198	1.6880	0.0914
Akaike criterion	1.8490	Hannan–Quinn criterion		1.9124
Schwarz criterion	2.0084	McFadden R-squared		0.2220
Prediction accuracy (y_i)	**Number of companies**	**Predicted number**	**Sum of all probabilities**	**Error**
1	23	25	23.4263	−0.4263
2	25	25	24.5721	0.4279
3	23	21	23.0015	−0.0015

The trinomial logit with predictors from year $t - 2$
Source: Gruszczyński (2004)

models with predictors from $t-1$. The average prediction accuracy is 86.8% for the model with the 1995 predictors and 87.3% for the model with the 1996 predictors. ■

The models of FD presented in this example have recently been compared to a number of other FD and bankruptcy models for Poland by Boratyńska and Grzegorzewska (2018), who applied them to a small number of enterprises in the agricultural sector. The accuracy of the models was again confirmed in this case. Incidentally, exercises of applying estimated models to other samples—sometimes from other geographical areas or legal regimes—are very common in bankruptcy literature. Since each model is "sample specific," such attempts are useful when the core idea (i.e., the set of variables) is maintained and the model is re-estimated on a new sample.

Binomial Models

For completeness and as introduction to the next section, we present two examples of binomial models of bankruptcy.

Example 3.4 Binomial Logit: Bankruptcy of Companies in Poland
The study by Ciesielski (2005) on bankrupt companies in Poland uses data on 96 companies with bankruptcy resolutions announced in the "Monitor Sądowy i Gospodarczy" in 2002. This group was tailored to reflect the structure of the economy in 2002. As a result, 40 bankrupt companies were selected for the training (basic) sample[7] and 20 companies for the holdout sample. A group of 60 "healthy" (financially viable) companies was selected in a similar way. This group was also divided into a training sample (40 firms) and a holdout sample (20 firms).

The author uses the following specification:
The Dependent Variable

$y_i = 1$ company is bankrupt
$y_i = 0$ company is non-bankrupt

The Model

$$probability\ (y_{it} = 1) = logit\left(predictor\ variables_{i,t-2}\right)$$

where *logit* means that we use a binomial logit model of the form introduced in Chap. 2 (Eq. 2.5) to explain probability. The term *predictor variables* indicates the list of predictors (financial ratios) with the year denoted by t. The model assumes a lag of 2 years (i.e., financials from 2000 (X variables) explain variable y in 2002.

[7]The sample comprised 12 companies representing manufacturing, 16 companies from commerce, 5 construction companies, 4 service companies (real estate, management, science), and 3 other companies (food processing, transportation, financial services).

Table 3.2 Bankruptcy of companies in Poland 2000–2002: binomial logit, basic sample

Dependent variable: Binomial variable $\boldsymbol{y} = 1$ if the company was announced bankrupt in 2002, $\boldsymbol{y} = 0$ otherwise			
Predictors (from 2000)	Model 1	Model 2	Model 3
PMO	0.51114	0.25829	−0.04507
NKA	−1.15192	−4.48416∗	
KA		0.47588	
BP	−0.32734		−0.02004
RZ	−0.11627	−0.19866∗	−0.09405
OZ	5.57730∗		
PO	−0.03178∗	−0.00866∗	−0.02330∗
WO	−5.26458∗	−2.14254	−5.26458∗
ROA		−7.59912∗	−9.24862∗
KWA			−7.13808∗
Constant	−3.18543∗	0.79459∗	2.71010∗
Pseudo R^2	0.55160	0.51330	0.59230

$n = 80$ firms that published financial statements in "Monitor Polski B" and "Monitor Spółdzielczy B" in 2000–2002
∗ denotes statistical significance at 0.05
Source: Ciesielski (2005)

The following predictors are considered in the author's models:

PMO—Reserves and short-term liabilities/current assets
NKA—Surplus (deficit) in working capital/total assets (working capital demand is defined as working capital minus net cash)
KA—Working capital/total assets
BP—Current ratio
RZ—Liabilities turnover (net sales/short-term liabilities)
OZ—Debt margin
PO—Gross profit (loss) plus interest/interest
WO—Net cash flow from operating activities/total assets
ROA—Return on assets
KWA—Shareholder equity plus long-term liabilities/total assets.

The sets of predictors were carefully selected in accordance with the model's specifications considered by the author. Table 3.2 shows estimates of the three binomial logit models presented in the chapter.

The three models each have seven predictors, which might be regarded as a relatively large number of predictors. The models have the following accuracy within and out of the samples:

	Model 1	Model 2	Model 3
Basic sample (%)	86	86	88
Holdout sample (%)	75	80	83

■

Example 3.5 "In Search of Distress Risk": Investing in US Distress Stocks[8]
A new angle of classic research on bankruptcy based on binomial modeling was presented by Campbell et al. (2008, 2011). Their attempt is based on previous studies by Shumway (2001) and Chava and Jarrow (2004).

The model is aimed at predicting "failure events" defined as filing for bankruptcy (Chaps. 7 or 11), delisting for performance-related reasons, or receiving a D rating from a rating agency. The variable y_{it} representing a failure event equals 1 in the case of failure, otherwise it equals 0. Thus, we have a simple binomial model. In this case, it is dynamic as failure events are observed monthly. These monthly US data cover the period 1963–2008 comprising some 1.7 million firm-months, among which there are about 1600 failure events.

The dynamic logit model employed here explains the probability of a failure event in month t by means of lagged (by 1 month) explanatory variables. The original expression from the paper is as follows:

$$P_{t-1}(y_{it} = 1) = \frac{1}{1 + exp\left(-\alpha - \boldsymbol{\beta}' \boldsymbol{x}_{i,t-1}\right)} \tag{3.7}$$

where y_{it} equals 1 if the firm goes bankrupt or fails in month t, and $\boldsymbol{x}_{i,\ t\ -\ 1}$ represents a vector of explanatory variables in month t–1 (i.e., at the end of month t–1).

The set of explanatory variables comprises the following accounting and market-based predictors:

- Net income to market valued total assets (NIMTA)
- Net income to total assets (NITA)
- Total liabilities divided by the sum of market equity and book liabilities (TLMTA)
- Total liabilities relative to total assets (TLTA)
- Book to market equity ratio
- Ratio of cash and short-term assets to the market value of assets (CASHMTA)
- Monthly log of excess return on equity relative to the S&P 500 index (EXRET)
- Standard deviation of daily stock return over the preceding 3 months (SIGMA)
- Relative size measured as the log ratio of its market capitalization to that of the S&P 500 index (RSIZE)
- Log of price per share, truncated from above at $15 (PRICE)

The novelty of this approach lies mainly in its fine tuning of the predictors, which are generally the same as in the previous studies by Shumway (2001) and Chava and Jarrow (2004). However, the profitability (NIMTA) and excess stock returns (EXRET) variables were also introduced in a distributed lag form—which helped to increase model performance.

The model outperforms other approaches in prediction: as the forecast horizon increases, the market-based variables are more important than the accounting

[8]This example is taken from Gruszczyński (2015).

variables. The model outperforms the O-score and the Z-score and doubles the accuracy of Moody's KMV distance-to-default. The safest 5% of stocks have an average failure probability of 0.01, while the probability is 0.34 for the riskiest 5%.

This study shows that one may construct financial distress models with increasingly better predictive performance—using only the "classic" market and accounting variables.

■

3.3 Unbalanced Samples in Bankruptcy Prediction[9]

Bankruptcy: A Rare Event

Bankruptcy or insolvency cases are rare within the population of active companies. In Poland 1299 companies filed for bankruptcy[10] in 2017 out of a population of 4,370,412—the total number of companies in the country in that year—yielding a percentage of failed companies of 0.1%. In Germany, those numbers were, respectively: 20,684; 4,923,202; and 0.4%. In Denmark: 6497; 570,496; and 1.2%. In Spain: 4059; 3,561,348; and 0.1%. As is common throughout Europe, the percentage of failed companies within the number of active companies is relatively low and rarely exceeds 1%.

Although instances of insolvency and bankruptcy[11] are rare, the probability of bankruptcy (insolvency, default, etc.) is much talked about in contemporary accounting, corporate finance, and financial markets analysis—for obvious reasons. Bankruptcy is the state many parties would like to be forewarned of. Such parties include company management, equity owners, lenders, potential investors, and insurers. The rarity of actual events of insolvency creates questions of how best to model such predictions for the entire population of companies.

Unbalanced Samples

The infrequent instances of company insolvency in large datasets of companies serve as the benchmark for assessing the probability of bankruptcy for all companies, including those outside the sample. Insolvency, bankruptcy, and default probabilities (PD = probability of default) are expected to be calculated for every firm. The modern methodology for assessing the PD began with the seminal paper by Edward

[9]This section is based on the paper by Gruszczyński (2019).

[10]According to Dun and Bradstreet, companies that file for bankruptcy are classified as failure companies. All numbers in this paragraph are taken from their Global Bankruptcy Report (DNB 2017).

[11]For discussion of insolvency/bankruptcy data see Staszkiewicz and Witkowski (2018).

Altman on Z-score in 1968 (Altman 1968). Altman's model uses linear multivariate discriminant analysis (LDA) to estimate the "score" that apprises stakeholders of a potential state of bankruptcy for the company in question.

Although a Z-score as the outcome of discriminant analysis (LDA) is not expressed in terms of probabilities, Z-scores may be inverted into probabilities with the use of logistic transformation. This conversion is "not strictly correct" (Hillegeist et al. 2004, 16) but may serve as a technique for comparison with other methods. A major complication is that independent variables in LDA must have a normal distribution, otherwise the LDA estimator is not consistent (Maddala 1983, 27). Nonetheless, if the Z-score is taken as the argument in the logistic function, the result may be interpreted as the "default likelihood" and used as the "equity-implied probability of default" (Altman et al. 2011).

The Altman models have been challenged by approaches directly producing probabilities of bankruptcy, such as the logit model, as well as by more advanced machine learning methods.

Bankruptcy prediction models aim at explaining the binomial outcome variable y representing bankruptcy (insolvency) with two possible values: $y_i = 1$ for bankrupt companies and $y_i = 0$ for non-bankrupt companies. Modeling involves the explanation of the y variable with the set of independent variables X (*covariates*).

A correct bankruptcy prediction model has potentially good classification accuracy and supplies reliable predictions of bankruptcy probabilities. The emphasis of the discussion here is on the rarity of bankruptcy cases within the population of all companies. This means that the number of cases with $y_i = 1$ is considerably smaller than cases with $y_i = 0$ and that this population's proportion is usually not represented in research samples.

There are two types of unbalanced samples (some researchers prefer "imbalanced") in bankruptcy prediction modeling. Let us consider the n-element sample for bankruptcy modeling that includes n_1 bankrupt companies and n_2 non-bankrupt companies.

(a) If the proportion of n_1 and n_2 in the sample is different than 50:50, then the sample is considered unbalanced.
(b) If the proportions $p_1 = n_1/n$ and $p_2 = n_2/n$ are different from the fractions of bankrupt and non-bankrupt companies in the population, then the sample is considered unbalanced in terms of p_1 and p_2.

The sample that is balanced in terms of definition (a) requires undersampling of healthy (survival) firms. This occurs because populations of companies are large—e.g., there are more than 4 million companies in Poland. It is, therefore, not feasible to sample healthy firms in the same manner as bankrupt ones. The "50:50 samples" appear in studies that use matching techniques[12]—each bankrupt company is

[12]This is called matched pairs sample design. Skogsvik and Skogsvik (2013) indicate that 70% of the early studies on bankruptcy use this design (Zmijewski 1984). An example of more recent bankruptcy research with matched pairs is the study by Bodle et al. (2016).

matched to a healthy company that is "similar" in terms of size, industry, etc. Such a sample is considered unbalanced from the point of view of definition (b).

Type (b) imbalance, as the more general situation, is discussed in the subsections that follow. In both cases, however, the question is whether the observation enters the sample randomly or not. If not, we have the problem of sampling bias, which is common in bankruptcy prediction models.

Sampling Bias, Weighting, WoE, and Resampling

Since the paper by Zmijewski (1984), the question of unbalanced samples has been examined from many angles but is still far from being resolved (see also Platt and Platt 2002; Chen et al. 2006). Nonrandom samples in bankruptcy models are the source of two types of biases:

- *Choice-based sample bias*: Results when the probability of a company entering a sample depends on the dependent variable attributes (e.g., firstly, data on bankrupt companies is collected and healthy companies are selected using some matching scheme).
- *Sample selection bias*: Results when only observations with complete data enter the sample.

Zmijewski (1984) has shown in several simulations with the probit model that choice-based sample bias declines if the ratio of bankrupt and non-bankrupt companies in the sample approaches that in the population. Also, neither bias appears to affect statistical inferences nor overall classification rates. However, they were shown to have an impact on the estimates for single observations (e.g., on the estimates of the probability of the bankruptcy of a particular company).

Unbalanced samples are also sometimes handled with appropriate weights for observations in bankruptcy modeling. In their exercise of comparing models internationally, Altman et al. (2017) used weights of single units both for unequal numbers of bankrupt and non-bankrupt data and for unequal numbers of observations across countries. In a footnote, the authors state:

> Although the score (logit) in principle has a probability interpretation, the "probabilities" estimated using this weighting scheme in this study do not, however, represent empirical PDs. It would still require calibration procedures for the models to obtain PDs that correspond to associated empirical PDs in the population. But this is not attempted in the study, as our focus is more general (the classification accuracies of the models across countries). It is also worth noting that the original Z-score does not have a PD interpretation either.

Weighting is a technique not often utilized in "classical" bankruptcy studies worldwide, despite the known results generated, for example, for the binomial logit (Manski and Lerman 1977). In the case of logit models, the use of appropriate weights may be as effective as the application of simple correction that is discussed in the next section (see also Maalouf et al. 2018). On the other hand, as stated by Long and Freese (2014), "The use of weights is a complex topic, and it is easy to apply weights incorrectly." For the logit model, the choice of weights is not straightforward.

Another procedure in practical applications is the use of the so-called Weights of Evidence (WoE), popular in scoring modeling. WoE is the univariate relationship between a predictor X_j and a binomial target variable y. The idea is to transform X_j into WoE variables and then fit the model (e.g., the logit model). The relationship that connects the WoE routine to unbalanced sampling is not direct. Attempts to use this technique in bankruptcy prediction are shown, for example, in Nehrebecka (2018).

The new generation of bankruptcy studies that has emerged with the use of machine learning techniques also propose new solutions for handling unbalanced samples. Zhou (2013) describes the use of oversampling and undersampling algorithms applied to 1981–2009 data on US bankruptcies and to 1989–2009 data on Japanese bankruptcies. Oversampling means sampling "the minority class over and over to achieve the balanced distribution of the two classes." Undersampling means "to select a portion of the majority class to achieve the distribution balance of the two classes" (Zhou 2013). The sampling techniques are for oversampling, ROWR (random oversampling with replication) and SMOTE (synthetic minority oversampling technique); and for undersampling, RU (random undersampling), UBOCFNN (undersampling based on clustering from the nearest neighbor), and UBOCFMGD (undersampling based on clustering from a Gaussian mixture distribution). These techniques may generate samples with a 50:50 composition of bankrupt and healthy companies and then may be used to verify various bankruptcy prediction models. What is important is that the major goal in studies using such techniques lies in finding the model that performs best in terms of classification accuracy. Other examples of such an approach are Choi et al. (2018) and Wagenmans (2017).

How does unbalanced sampling in bankruptcy models interfere with bankruptcy probabilities? To answer this question, we will concentrate on the binomial logit model.

Prior Correction in the Logit Model

Let us consider the binomial logit model of bankruptcy and the consequences of unbalanced samples for the prediction of bankruptcy probability. One method for overcoming the effects of unbalancing is weighting, as explained in the previous section. We advocate the use of a simple correction, sometimes called "prior correction" (King and Zeng 2001) or the "Anderson-Maddala correction" (Gruszczyński 2017). King and Zeng (2001) state that although econometricians attribute the correction to Manski and Lerman (1977), in fact, the correction has been well known since 1975 (Bishop et al. [1975] 2007). We challenge this finding by noting that the paper by Anderson (1972) first introduced this result, which was later restated by Maddala (1983, 1991).

Prior correction allows the analyst to convert the binomial logit model estimated on the basis of on an unbalanced sample to the model for the population. The condition is that the "fraction of ones" (i.e., bankrupt companies) in the population is known. As before, $y_i = 1$ means a bankrupt company and $y_i = 0$, a non-bankrupt company. The subject of the modeling is the probability $P(y_i = 1)$. Let us assume that the fraction of ones in the population is equal to π . King and Zeng (2001) state that knowledge of π "can come from census data, a random sample from the population measuring y only." In the case of bankruptcy modeling, the fraction may be established from official data on bankruptcies and companies for a particular country, region, time period, etc. Assume that $\pi = N_1/N$, where N_1 is the number of bankrupt companies in the population and $N = N_1 + N_2$ is the population size (with the number of non-bankrupt companies equal to N_2).

Now, consider the n-size sample for bankruptcy modeling that includes n_1 bankrupt companies and n_2 non-bankrupt companies. The fraction of ones in the sample is $\bar{y} = n_1/n$. The proportion of bankrupt companies selected for the sample is $p_1 = n_1/N_1$ and the analogous proportion of non-bankrupt companies is $p_2 = n_2/N_2$.

Consider also the following binomial logit model with k covariates X and $k + 1$ parameters $\beta_0, \beta_1, \ldots, \beta_k$

$$P(y_i = 1) = \frac{1}{1 + \exp\left(-\boldsymbol{x}_i'\boldsymbol{\beta}\right)} \tag{3.8}$$

where $\boldsymbol{x}_i'\boldsymbol{\beta} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \ldots + \beta_k X_{ki}$. The maximum likelihood estimation of Eq. (3.8) in the n-element sample gives the estimate of intercept β_0 that needs to be corrected—if estimated Eq. (3.8) is to represent the population (not only the sample).

The correction, known as "prior correction" or the "Anderson-Maddala correction" amounts to subtracting the estimate of β_0 by

$$prior\ correction = \ln\left[\left(\frac{1-\pi}{\pi}\right)\left(\frac{\bar{y}}{1-\bar{y}}\right)\right] \text{[King and Zeng 2001]} \tag{3.9}$$

or

$$\delta = ln\left(\frac{p_1}{p_2}\right) \text{[Maddala 1983]} \tag{3.10}$$

Corrections (3.9) and (3.10) are equal, which can be shown using definitions of p_1, p_2, π, and $\bar{y}$. For the randomly selected sample, we have $p_1 = p_2$ and $\pi = \bar{y}$ and the prior correction is equal to zero. Thus, the nonrandom samples inherently imply the need for correcting the model in order to have it represent the population. However, if the population is not precisely known, the fractions p_1, p_2, and π can only be estimates or calibrations. Then the correction (3.9) or (3.10) should be applied carefully and with a relevant comment. In any case, we apply the correction when we make inferences in the context of the entire population and not the sample itself. Later we use the correction δ from Eq. (3.10).

To sum up, a model estimated on a sample not representing the population's proportion of bankrupt companies gives estimates of bankruptcy probability, which are biased regarding the entire population. Unbiased probabilities of bankruptcy can be obtained after adjusting the model. Note that this discussion leaves aside the question of classification and classification accuracy. We concentrate here on estimating the PD—the probability of bankruptcy/insolvency—especially when the model is used for companies outside the sample.

Prior correction in the form of Eq. (3.9) or (3.10) coincides with the findings by Skovsik and Skogsvik (2013). They also emphasize that the bankruptcy probabilities obtained from the bankruptcy prediction models depend on the share of bankrupt companies in the sample, and they are, therefore, biased—if the share of bankrupt companies in the sample is not the same as the share in the population. According to the authors' findings, there is an algebraic relationship between the biased bankruptcy probability of a given company (from the model, *sample-based*), and the unbiased probability, which results from the proportion of bankrupts in the population. This share of bankrupts in the population (denoted by π) is treated as the a priori probability of bankruptcy. The probability of bankruptcy of a single company calculated from the model (*sample-based*) denoted by the Skogsviks as p_{fail}^{prop} is, therefore, biased. It is the function of:

- The unbiased probability p_{fail}^{π}
- The proportion (*prop*) of bankrupt companies in the sample
- The proportion π of bankrupt companies in the population.

The formula derived from the Bayes theorem (Skogsvik and Skogsvik 2013) is as follows:

$$p_{fail}^{prop} = \left[1 + \left(\frac{\pi}{1-\pi}\right)\left(\frac{1-prop}{prop}\right)\left(\frac{1-p_{fail}^{\pi}}{p_{fail}^{\pi}}\right)\right]^{-1} \tag{3.11}$$

It follows that if $prop > \pi$, then $p_{fail}^{prop} > p_{fail}^{\pi}$ and vice versa. This means that, in the typical situation of bankruptcy modeling (i.e., when $prop > \pi$), the PD for a given company calculated from the model is higher than the "population-adjusted" PD for the same company. How much higher? The authors give the example with $prop = 0.5$ as in matched pairs modeling and $\pi = 0.02$. If the unbiased probabilities are $p_{fail}^{\pi} = 0.01, 0.02, 0.10$, they correspond to the (biased) predictions from the model $p_{fail}^{prop} = 0.33, 0.55, 0.84$, respectively. Thus, the model predicted probabilities are considerably higher than the unbiased probabilities. It should be noted that Eq. (3.11) has been derived assuming random sampling from the population of bankrupt companies as well as (separately) from the population of non-bankrupt companies.

Formula (3.11) can be written in terms of p_1, p_2 from Eq. (3.10) as

$$p_{fail}^{prop} = \left[1 + \frac{p_2}{p_1}\left(\frac{1-p_{fail}^{\pi}}{p_{fail}^{\pi}}\right)\right]^{-1} \tag{3.11$'$}$$

The Skogsviks' equation seems to be important for calculating a specific unbiased probability p_{fail}^{π} for a company that is needed in financial risk management or in the valuation of a company's equity or bonds. From Eq. (3.11) it follows that the adjustment of the unbiased PD is equal to

$$p_{fail}^{\pi} = \left[1 + \frac{p_1}{p_2}\left(\frac{1 - p_{fail}^{prop}}{p_{fail}^{prop}}\right)\right]^{-1} \tag{3.12}$$

This result allows the analyst to translate the model's predicted PD for a single company into the PD "calibrated for the fraction of failure companies in the population."

The adjustment of the probability of bankruptcy (default) is only possible when the number representing π, (i.e., the fraction of bankruptcies in a recognized population of companies for a given year) is known or may be feasibly approximated. If we assume that $\pi = 0.02$ and $prop = 0.5$, then the bankruptcy probability estimated from the model as 0.7 corresponds to the "population adjusted" probability of 0.045. This exaggeration of unbiased PDs that is inherent in bankruptcy models should be considered in practical uses.

The transformation of biased into unbiased probabilities can be further specified for various binomial models. We return to prior corrections (3.9) and (3.10) for the logit model (3.8). The correction δ of the constant term β_0 estimate defined in Eq. (3.10) can be shown to coincide with the Skogsviks' equation. From Eq. (3.11) we have

$$-\,ln\left(\frac{p_1}{p_2}\right) - logit\; p_{fail}^{\pi} = -logit\; p_{fail}^{prop}$$

and

$$\delta + logit\; p_{fail}^{\pi} = logit\; p_{fail}^{prop}$$

here δ is defined in Eq. (3.10). Therefore, the logit for the biased bankruptcy probability p_{fail}^{prop} (i.e., what we receive from the estimated logit model) must be reduced by the value of δ in order to obtain the logit of unbiased bankruptcy probability p_{fail}^{π}. The reduction is contained in the constant term.

For example, an estimated logit model with seven explanatory variables is as follows[13]:

$$\widehat{logit}\, P = 0.79 + 0.26X_1 - 4.48X_2 + 0.48X_3 - 0.20X_4 - 0.01X_5 - 2.14X_6 - 7.60X_7$$

[13]Model estimated by Ciesielski (2005). Details in Table 3.2 in Sect. 3.2.

This equation has been estimated for 40 bankrupt and 40 non-bankrupt companies. It means that the proportion of bankrupt companies in the sample is $prop = 0.5$. For the values of the explanatory variables corresponding to one specified case (firm), the probability of bankruptcy resulting from this model is $p_{fail}^{prop} = 0.6$.

Now, let us assume that the proportion of bankrupt companies in the population is $\pi = 0.02$. Therefore, the ratio equation (3.10) is $\frac{p_1}{p_2} = 49$. In order to calculate the population adjusted (unbiased) probability from the estimated equation, the intercept should be reduced by

$$\delta = ln\left(\frac{p_1}{p_2}\right) = 3.89182$$

The new intercept is now equal to -3.10. The probability of bankruptcy obtained from the estimated equation with the new intercept is exactly equal to the probability p_{fail}^{π} calculated from Eq. (3.11). In this case $p_{fail}^{\pi} = 0.0297$.

The Skogsviks' equation (3.11) applies to the outcomes of all binomial models. For example, in the linear probability model (LPM), the probability p_{fail}^{prop} of bankruptcy is equal to the estimate of the dependent variable for a specific company. From Eq. (3.12), the unbiased p_{fail}^{π} is calculated, provided that the proportions of p_1 and p_2 are known. In the probit model, the estimate of bankruptcy probability can also be calculated and inserted into Eq. (3.12) as p_{fail}^{prop}. It should be noted that only in the binomial logit model does there exist a simple correction for the estimated model that coincides with the Skogsviks' formula. King and Zeng (2001) point out that in other binomial models like probit the only possibility is the use of a Skogsviks-like equation.

None of the foregoing considerations refer to the question of classification accuracy (classification of companies by the model). Since the rankings of the companies in terms of probabilities p_{fail}^{prop} and p_{fail}^{π} are identical (Skogsvik and Skogsvik 2013), the classifications of companies within the sample (into two groups: bankrupt and non-bankrupt) based on biased and unbiased predictions are the same, assuming the appropriate choice of the cut-off point.

The cut-off point α is the limit of probability for classification: if the estimated probability is less than α, the company is classified as non-bankrupt; if not, it is classified as bankrupt. The default cut-off point in programs for estimating binomial models such as Stata is $\alpha = 0.5$. We advocate the use of Cramer's rule (Cramer 1999; Śmigielski et al. 2010), according to which the cut-off points α are:

- For the biased predictions from the bankruptcy model, $\alpha = prop = \bar{y} = \frac{n_1}{n},$
- For the unbiased predictions from Eq. (3.12), $\alpha = \pi = \frac{N_1}{N}.$

Cramer's rule is based on the notion that the typical cut-off point of 0.5 applied for unbalanced samples does not allow one to reasonably predict less frequent cases. Cramer (1999) states:

> [The] choice of 0.5 is usually defended by the argument that it is optimal if the predicted y_i determine a course of action and if moreover the cost of misclassification is the same for either form that this may take. But if the cut-off point is optimal for the use of predictions in actual decisions it need not also be optimal for assessing the within-sample performance of the fitted model.

Cramer (1999) proposes using a cut-off point α equal to the proportion of ones in the sample because it yields predictions that are optimal in the sense that they maximize the "index of performance" for each observation.[14] In effect, the success rate for the unbalanced samples is better spread over the two alternatives, $y_i = 1$ and $y_i = 0$.

As noted in Sect. 3.3, models of multivariate discriminant analysis are not directly used to estimate bankruptcy probabilities. However, the LDA estimation results can be corrected by considering the population's proportion of bankrupt and non-bankrupt companies (Zmijewski 1984; Altman and Eisenbeis 1978).

3.4 Models of Firm Exit

Two States of Exit: Binomial Models and LDA

In the previous section, we concentrated on financial distress and bankruptcy, with the emphasis on modeling diverse states of distress. This section is devoted to examining ways and determinants of firm exit—i.e., termination of a company's activity/existence. The chief type of exit is bankruptcy. Typically, this means the involuntary termination of activity, involuntary exit. Other exits are typically voluntary, like restructuring or liquidation due to a merger or acquisition.

Typical models of firm exit are binomial models: usually companies that terminate their activities (e.g., due to bankruptcy) are compared with "healthy" ones. Most of the binomial models discussed in previous sections refer to two states: bankruptcy and non-bankruptcy. The same two states are used in models of linear discriminant analysis that became the staple methodology in bankruptcy prediction after the seminal work of Altman (1968) and his Z-score.

Z-score methodology and binomial modeling (especially the binomial logit) have had numerous extensions made by Altman himself, his collaborators, and many other authors (Altman 2018; Altman et al. 2017). Also, new methodological approaches have emerged, such as the data-driven models (machine learning models) mentioned in Sects. 3.2 and 3.3.

In Sects. 3.1, 3.2, and 3.3, we discussed many aspects of modeling a firm exit in the form of bankruptcy using a two-state dependent variable. We mention it again here for the sake of completeness.

[14] The "index of performance" is defined as the probability of the observed outcome estimated from the model—related to the "null value" of this probability (i.e., estimated from the model containing only the constant term).

Models for Many States of Exit

Many studies consider more than one type of company exit, including perhaps restructuring, mergers, and acquisitions. In the study by Cefis and Marsili (2012) in the Netherlands, the following states are considered:

1. Exits

Failure: A firm is dropped from the register because of the termination of activities (voluntary or by bankruptcy).
Mergers and acquisitions: A firm is merged with other firm(s) into a firm with new identity or a firm is acquired by another firm which maintains its identity.
Radical restructuring: A firm is decomposed into several units or a firm undergoes a major transformation (which results in a change to its legal or economic identity).

2. Continuation (survival)

Cefis and Marsili (2012) apply the multinomial logit model and the complementary log-log model with the base category being continuing state.[15]

Example 3.6 Exit of Firms in Belgium
Balcaen et al. (2012) in their study of determinants of exit for Belgian firms consider three types of exit after distress.

Court-driven exit

Bankruptcy, Compulsory liquidation, Moratorium on payments

Voluntary liquidation

Not ordered by court, a self-imposed windup and dissolution of the firm approved by shareholders; effective when the liquidation value of a firm's assets is higher than the liquidation value of its liabilities: the firm's assets are sold, creditors are paid off, and any remaining funds are distributed to shareholders.

M&A exit

Acquisition, Merger, Split

The authors use a dataset of 6118 exits provided by the Belgian National Bank for the years 1998–2000. The numbers are presented in Table 3.3. The most frequent events are voluntary liquidations (44%), bankruptcies (41%), and mergers (13%).

The determinants of each choice of firm exit are analyzed here using the binomial nested logit.[16] The nested logit consists of two binomial logit models: for level 1 (lower) and for level 2 (higher). The model for level 2 explains the choice between

---

[15] Cefis and Marsili (2012) use the competing risks model that is explained later in Sect. 3.5.

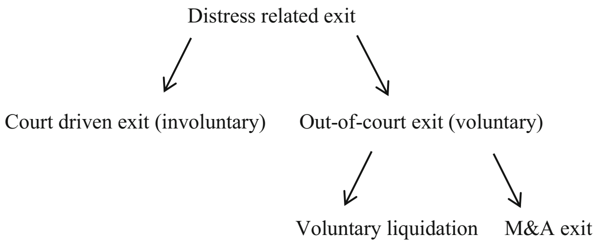
[16] See, e.g., Winkelmann and Boes (2006) and Gruszczyński (2012, Chap. 5).

Table 3.3 The number of exits by exit type and legal procedure, Belgium 1998–2000

Exit type	Legal procedure	Number of firms	Percentage
Court-driven exit	All procedures	2533	41.40
	Bankruptcy	2518	41.16
	Compulsory liquidation	4	0.07
	Moratorium on payments[a]	11	0.17
Voluntary liquidation		2700	44.13
M&A exit	All procedures	885	14.47
	Acquisition	770	12.59
	Merger	5	0.08
	Split	110	1.80
Total		6118	100.00

Source: Balcaen et al. (2012)
[a]Recalled moratorium on payments after failure of firm's restructuring plan

an out-of-court exit and a court-driven exit. The model for level 1 explains the choice between voluntary liquidation and M&A exit, conditional on an out-of-court exit. Both models use explanatory and control variables.



The model for level 1 is as follows:

$$ln \frac{P(\text{voluntary liquidation}|\text{voluntary exit})}{P(\text{restructuring exit}|\text{voluntary exit})} = \boldsymbol{x}'\boldsymbol{\beta} \tag{3.13}$$

where $\boldsymbol{x}$ is the vector of the explanatory variables and $\boldsymbol{\beta}$ is the parameter vector (P denotes probability). The model for level 2 is

$$ln \frac{P(\text{voluntary exit})}{P(\text{involuntary exit})} = \boldsymbol{z}'\boldsymbol{\gamma} \tag{3.14}$$

where $\boldsymbol{z}$ is the vector of the explanatory variables and $\boldsymbol{\gamma}$ is the parameter vector. Variables $\boldsymbol{z}$ include the so-called *inclusive value* that is calculated from the level 1 model. It is the estimate of the expected utility of the level 1 choice conditional on the choice on level 2. If the inclusive value is significant, the nested model is more beneficial than the standard multinomial logit model.

Table 3.4 Determinants of exit types (Belgian firms 1998–2000)

Distress related exit [level 2 model]	
↙	↘
Court-driven exit (involuntary exit)	Out-of-court exit (voluntary exit)
↑	↑
small cash holdings	large cash holdings
high leverage	low leverage
presence of secured debts	absence of secured debts
low level of secured debt	high level of secured debt
independent firm	member of a group
weak group relations	strong group relations
younger firm	older firm

Out-of-court exit [level 1 model]	
↙	↘
voluntary liquidation	M&A exit
↑	↑
large cash holdings	small cash holdings
absence of secured debts	presence of secured debts
high level of secured debt	low level of secured debt
independent firm	member of a group
smaller firm	larger firm

Source: Balcaen et al. (2012)

The results of the study by Balcaen et al. (2012) are presented in Table 3.4 listing the significant determinants of each choice of exit. The variables that emerge as determinants of exit are

- CASH—The percentage of cash and cash equivalents to total assets
- LEVERAGE—The ratio of the book value of long- and short-term debt to total assets
- D_SECURED—The presence of secured debts: = 1 when at least some debts are guaranteed by business securities, = 0 otherwise
- SECURED—Percentage of the total debt guaranteed by business securities to total assets
- D_GROUP—= 1 if the firm belongs to a group of companies, = 0 if it is independent

- GROUP—Strength of group relations measured by the level of financial interactions with related firms and firms with holding interests as a percentage of total assets
- AGE—Firm age: the number of years of operational activity at the first sign of distress
- SIZE—Firm size: the natural log of the book value of total assets (in €1000)

The results show that cash holdings are positively related to the choice of a voluntary exit, while they are negatively related to an involuntary exit (court-driven exit). Similarly, the company's potential resources expressed by group relations and the level of secured debt are positively related to a voluntary exit.

A financially distressed firm is more inclined to exit by voluntary liquidation if the relative efficiency of liquidation is high compared to voluntary restructuring (M&A exit). This relative efficiency is associated with factors determining the probability of a successful liquidation and a successful M&A exit, such as cash holdings, group participation, firm size, and the level of secured debt.

■

3.5 Models of Firm Survival

Microeconometric Models of Firm Survival

A model of firm duration (survival) was already shown in Chap. 2 in Table 2.4. Part of that table is dedicated to the duration model.

Duration model	Esteve-Pérez and Mañez-Castillejo (2008)
Model of firm duration: The probability that a company exists in year t assuming it existed in year t-1 (hazard function model) ($n = 14{,}193$ observations, 2028 industrial firms, Spain 1990–2000).	$\boldsymbol{X}$ = firm size (= 1 for firms with more than 200 employees, = 0 otherwise); advertising expenditures (1 = yes, 0 = no); R&D strategy (3 categories); industry technological intensity (3 categories); export intensity (3 categories); productivity (3 categories); limited liability company (1 = yes, 0 = no); and other variables.

This is an example of modeling a firm's exit (voluntary or not) in which time to exit is considered crucial. We are interested in a firm that experiences exit in a given year, assuming that it survived until that year.

Models of firm survival are popular in the field of industrial organization. Firm survival also belongs to the domain of business demography.[17] In Eurostat,

[17]The term "business demography" has been criticized as the wrong use of word "demos" (from ancient Greek: "the people"). In Poland see, e.g., Domański and Szreder (2010) and the discussion by Paradysz (2011).

"business demography" data encompass the active population of enterprises, their birth, survival (followed for up to 5 years from birth), and death (exit).

Econometric methods dedicated to survival belong to duration analysis or survival analysis. It is microeconometrics because it is rooted in the microdata of firms throughout their lifetime. The models describe the survival processes with explanatory variables (determinants, predictors). Manjón-Antolín and Arauzo-Carod (2008) published a survey of firm survival literature, and papers on firm survival also appear in journals on corporate finance, accounting, and corporate governance (some titles are mentioned in Chap. 1). There is also a series of firm survival research papers on Academia.edu.

Firm survival (or firm duration) is examined through the probability that the firm with characteristics X terminates its operations in year t assuming that the firm survives to year t. For small increments of time t, the probability can be regarded as the value of the so-called hazard function.

The variable that is studied in survival analysis is the time between "entry" and "exit"—i.e., *duration* (spell) from the firm's creation (birth) to its termination (exit). Duration may be observed as a *complete spell* (in full) or as an *incomplete spell*—due to left censoring or right censoring.

*Durati*on is the nonnegative random variable T. In economics, duration is usually discrete. The basics of survival theory is usually considered for the continuous case (Cameron and Trivedi 2005). If the cumulative distribution function of T is denoted as $F(t)$ and the density function as $f(t) = dF(t)dt$, then the probability that duration is not longer than t is given by $Pr(T \leq t) = F(t)$.

The probability that duration exceeds t is called the *survivor function* and is equal to $S(t) = 1 - F(t)$. This is the probability that exit will take place after t (i.e., the firm survives until t). The survivor function decreases monotonically from 1 to 0.

The *hazard function* is defined as

$$\lambda(t) = \lim_{\Delta t \to \infty} \frac{Pr\left(t \leq T < t + \Delta t \middle| T \geq t\right)}{\Delta t} = \frac{f(t)}{S(t)} \tag{3.15}$$

The value of the hazard function is the instantaneous (for small Δt) probability of leaving a state conditional on survival to time t. It can be shown that hazard $\lambda(t)$ equals the following change of log-survivor function: $\lambda(t) = -dln(S(t))/dt$.

Regression (econometric) analysis in survival analysis concentrates on the conditional hazard rate $\lambda(t|x)$ in which the vector x represents the explanatory variables. This contrasts with the more standard approach in which the conditional mean $E(T|x)$ is of more interest (Cameron and Trivedi 2005).

The most popular models are the *Cox proportional hazard* (CPH) and the *accelerated failure time model*. Here is a short exposition on the CPH. In the proportional hazard model, the conditional hazard $\lambda(t|x)$ can be factored into separate functions

$$\lambda(t|x, \beta) = \lambda_0(t)\phi(x, \beta) \tag{3.16}$$

where $\lambda_0(t)$ is called the *baseline hazard* and is the function of t alone and $\phi(x, \beta)$ is the function of the x variables (with parameters β). All conditional hazard functions $\lambda(t|x)$ of type (Eq. 3.16) are proportional to the baseline hazard—with the scaling factor $\phi(x, \beta)$ that is not an explicit function of t.

In the case of CPH, the semiparametric model is considered in which the form $\lambda_0(t)$ is not specified and the functional form for $\phi(x, \beta)$ is fully specified. Usually for $\phi(x, \beta)$, the exponential function is chosen.

$$\phi(x, \beta) = \exp(x'\beta) \tag{3.17}$$

It is this form that is convenient for interpretation of the parameters. If the m-th explanatory variable X_m increases by 1 unit (with other covariates unchanged), then

$$\lambda(t|x_{new}, \beta) = \exp(\beta_m)\lambda(t|x, \beta) \tag{3.18}$$

The increase of variable X_m by 1 unit is associated with a change in hazard by $\exp(\beta_m)$ times. The value $\exp(\beta_m)$ is called the hazard ratio.

As mentioned above, the parameters β may be estimated in the CPH model with no reference to the baseline hazard $\lambda_0(t)$. A typical CPH model is estimated employing the method of maximum likelihood.

Single Spell Duration Models

The typical duration models presented in corporate finance studies are single hazard (single spell) models like the Cox proportional hazard (CPH). The subject of analysis is survival until year t and termination in year t for one reason—e.g., exit by bankruptcy.

The example from the study of Esteve-Pérez and Mañez-Castillejo (2008) mentioned above is a single hazard model. The following is another example, also from Spain.

Example 3.7 Survival of Spanish Aquaculture Firms
Lagares et al. (2018) estimated the Cox proportional hazard model for the data from 247 aquacultural firms in Spain whose principal activity was feedlot operations and breeding fish, crustaceans, and mollusks. The companies were followed from 1997 to 2010. The predictors in the CPH model are as follows:

- AGE = Firm age measured by the number of years since the company was established.
- ENVIRODUM = 1 if the firm does not report its environmental commitment on its website; = 0 otherwise.
- RDIDUM = 1 if the firm does not play a role in research, development activities, or in research, development, and innovation activities; = 0 otherwise.

Table 3.5 Cox proportional hazard model for Spanish aquacultural firms 2007–2014

Cox proportional hazard model: explains the probability of a company entering the "exit" state after being in the "normal" state.		
Explanatory variables	Parameter estimates	Hazard ratios
AGE	0.035**	1.035
ENVIRODUM	1.501**	4.487
RDIDUM	0.539	1.714
EXPIMPDUM	0.293	1.341
ROADUM	0.643**	1.901
SOLRDUM	0.674***	1.962

$n = 247$ firms; data for 2007–2014; log-likelihood $= -402.657$
***significance at 0.01; ** significance at 0.05
Source: Lagares et al. (2018)

- EXPIMPDUM = 1 if the firm does not engage in export/import operations; = 0 otherwise.
- ROADUM = 1 if the firm has economic profitability that is negative or equal to zero— profitability is measured by ROA; = 0 if ROA is positive.
- SOLRDUM = 1 if the firm has a solvency ratio that is negative or equal to zero— the solvency ratio is equal to shareholder funds/total assets; = 0 if solvency ratio is positive.

Table 3.5 presents the estimation results of the survival model.

The hazard ratios are the values of $\exp(\beta_m)$. For example, exit hazard is 1.96 times higher for firms that have a solvency ratio negative or equal to zero (compared to firms with a positive solvency ratio). Survival probability decreases with firm age (AGE): a 1-year increase in age increases the exit hazard of marine aquaculture firms by 3.5%. Other estimates may be interpreted similarly.

The authors positively verified the assumption of proportional hazards and overall model adequacy.

■

The application of the Cox proportional hazard models for firm survival analysis may be found in diverse journals. For example, Matsuno et al. (2017) estimated the Cox proportional hazard model for data on 334 firms from the Japanese information service industry (1997–2010), and Jung et al. (2018) applied the CPH model to 588 Korean small- and medium-sized enterprises for 2008–2014.

Competing Risks Models

We may model only one type of firm exit (termination of activities) using single hazard. Such models are relevant when the research target concentrates on finding determinants of further operation or non-operation of the company. For many states of firm exit (bankruptcy, voluntary liquidation, restructuring, etc.), the appropriate hazard model is the *competing risks model*. In this setup, every company is concurrently under risk for all states of financial distress.

There are two outcomes observed: duration to exit and exit state (cause of exit). The causes of terminating the firm's activity are "competing" in the sense that only the cause with the shortest spell (duration) is realized. In biostatistics, competing risks models are applied to study determinants of death from competing causes. Death is the result of the cause with the shortest spell. Firm termination (exit) has the same narrative: if termination is due to liquidation, then other causes are not valid.

In a simple case, the competing risks model assumes that the risks (causes of exit) are independent. In the popular *Cox competing risk model,* there are m proportional hazards

$$\lambda_j(t|\boldsymbol{x},\boldsymbol{\beta}) = \lambda_{0j}(t)\exp\left(\boldsymbol{x}_j'(t)\boldsymbol{\beta}_j\right) \text{ for } j = 1, \ldots, m \tag{3.19}$$

where both the baseline hazard λ_{0j} and $\boldsymbol{\beta}_j$ are specific to type j hazard and $t_{j1} < \ldots < t_{jk_j}$ denote ordered distinct time points in which exit by any cause occurs (k_j is the exit of type j) (Cameron and Trivedi 2005). "Exit" may mean here either the type of firm termination or the continuation state.

Model (Eq. 3.19) is estimated with maximum likelihood. The results are estimates of the J vectors $\boldsymbol{\beta}_j$ (regression coefficients). Interpretation of the coefficients is not straightforward. The probability of "exiting" the given state via exit r is given by

$$Pr(r|t,\boldsymbol{x},\boldsymbol{\beta}) = \frac{\lambda_r(t|\boldsymbol{x}_r,\boldsymbol{\beta}_r)}{\sum_{j=1}^{J}\lambda_j\left(t|\boldsymbol{x}_j,\boldsymbol{\beta}_j\right)} \text{ for } r = 1, \ldots, J \tag{3.20}$$

where one of the "exits" is the continuation of firm activity, the remaining J–1 exits are states of termination of the firm's activity. It is always interesting to see what the marginal effect of one single variable X (covariate) is on this probability. Since the covariates "appear in both the numerator and the denominator, and moreover the denominator is the sum of all hazards, the sign of the partial derivative $\partial\ Pr\ (r|\ t, \boldsymbol{x}, \boldsymbol{\beta})/\partial x_{rk}$ depends on all parameters in the model" (Cameron and Trivedi 2005). It may be shown that if $j \neq r$ and $\beta_{rk} > \beta_{jk}$, then the sign of $\partial\ Pr\ (r|\ t, \boldsymbol{x}, \boldsymbol{\beta})/\partial x_{rk}$ is positive. This means that an increase of variable x_k is associated with an increase of the conditional probability of the r-th exit if the parameter estimate by this variable in $\lambda_r(t|\ \boldsymbol{x}_r, \boldsymbol{\beta}_r)$ is higher than the parameter estimates for the same variable in all the other hazard functions.

Formula (3.20) is similar to the formulae for multinomial models. The difference is that here we model a firm's survival up to a specific year, not only the fact that the firm terminates its activity in that year. With the independence assumption, the Cox competing risks model may be estimated separately for each risk, treating the remaining states jointly as one state. For J risks, we estimate J–1 models of the CPH type.

Example 3.8 Survival of Firms in Australia
Chancharat et al. (2010) estimated competing risks models for data on 1081 public companies in Australia over the period 1989–2005. The explanatory variables are financial ratios, market indicators, and other firm characteristics. Three states are considered:

- 0 = *Active companies*
- 1 = *Distressed external administration companies*—according to Australian law, companies file for external administration in the following categories: voluntary administration, a scheme of arrangement, receivership, or liquidation
- 2 = *Distressed takeover, merger, or acquisition companies*—delisted from the ASX (Australian Stock Exchange) because they were subject to one of those events

In the final sample, there are 891 active listed companies, 50 companies in state 1, and 140 companies in state 2. Two types of models are estimated in this study: a single hazard model and a competing risks model. For the single hazard model, the states 1 and 2 have been merged into one state of exit. Survival time for the distressed companies is the total number of years from the first year when data are available to the year of financial distress.

Explanatory variables for the competing risks model are:

- EBT—EBIT margin = EBIT/operating revenue
- ROE—Return on equity
- ROA—Return on assets
- CUR—Current ratio
- WCA—Working capital/total assets
- DET—Debt ratio = total debts/total assets
- CPT—Capital turnover = operating revenues/operating invested capital before goodwill
- TAT—Total asset turnover = operating revenues/total assets
- SIZE—The size of the company = the natural log of sales
- SIZE2—The square of SIZE
- AGE—The age of the company = the number of years since registration
- EXR—Excess return = company's stock return in year t–1 minus ASX index return in year t–1

The estimation results of the competing risks model are presented in Table 3.6.

Several predictors are not statistically significant. This may be the result of including three profitability ratios, two liquidity ratios, and two activity ratios in the model. Ratios within groups are usually correlated and one ratio from a group of ratios is enough for inclusion in a model. Interpretation of results is given by the authors as an examination of the signs of the estimated coefficients without reference to the marginal effects of the variables on the probabilities of exit.

■

Table 3.6 Survival model for Australian public companies in 1989–2005

The Cox competing risks model explains the probability of an exit from the "normal" state (state 0) to "risk" represented by two states: 1 and 2				
	1: bankruptcy, restructuring liquidation		2: merger or acquisition	
Explanatory variable	Estimate	Hazard ratio	Estimate	Hazard ratio
EBT	−0.0006	0.999	−0.0019	0.998
ROE	−0.0805	0.923	0.0195	1.020
ROA	−0.4143	0.661	−0.3871	0.679
CUR	−0.6156	0.544	−0.1787	0.836
WCA	0.9740*	2.649	−0.3987	0.671
DET	0.9205**	2.510	−0.7975*	0.450
CPT	−0.0053	0.995	0.0131*	1.013
TAT	−0.1919	0.825	−0.1554	0.856
SIZE	0.8393*	2.315	1.6956**	5.450
SIZE2	−0.0223	0.978	−0.0412**	0.960
AGE	−0.0014	0.999	−0.0028	0.997
EXR	−0.7538**	0.471	0.1167	1.124
Number of cases	50		140	

Termination event: Two states, 1 = bankruptcy, restructuring, or liquidation; 2 = merger or acquisition

n = 1081 Australian public companies 1989–2005

** significance at 0.05; * significance at 0.10

Source: Chancharat et al. (2010)

Competing risks models are popular in applications. For example, Amendola et al. (2015) present a model for Italian firms operating in the construction sector in 2004–2009. The sample consists of 221 companies that went bankrupt, 129 that entered voluntary liquidation, 228 that were inactive, and 884 companies in an active state. The authors estimated the competing risks model (with three states of exit) and a single risk model for comparison purposes. A similar study by Cressy and Farag (2014) examines the risks of firms delisting from the Hong Kong Growth Enterprise Market (GEM) in the years 2000–2012. The primary reason for delisting is transfer to the Main Board. The authors apply a range of survival models, including the competing risks model.

The issues of financial distress and bankruptcy modeling presented in this chapter complement the material on financial microeconometrics from Chaps. 1 and 2, particularly in regard to these important areas of applications. Information on financial distress, bankruptcy, and other types of cessation of company activities is of crucial importance for owners of equity, management, lenders, investors, and other parties.

As we mentioned at the beginning of this chapter, modeling bankruptcy and financial distress is the most popular area of statistical-econometric research in empirical corporate finance. The modern history of this research dates from the 60s of the twentieth century.

In this chapter, we have tried to enumerate the essential issues of microeconometric modeling in this field, including the selection of predictions, unbalanced samples, modeling for distress versus bankruptcy, as well as modeling firm exit and firm survival.

References

Altman EI (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J Financ 23(4):589–609

Altman EI (2018) Applications of distress prediction models: what have we learned after 50 years from the Z-score models? Int J Financ Stud 6(3):70

Altman EI, Eisenbeis RA (1978) Financial applications of discriminant analysis: a clarification. J Financ Quant Anal 13(1):185–195

Altman EI, Fargher N, Kalotay E (2011) A simple empirical model of equity-implied probabilities of default. J Fixed Income 20:71–85

Altman EI, Iwanicz-Drozdowska M, Latinen EK, Suvas A (2017) Financial distress prediction in an international context: a review and empirical analysis of Altman's Z-score models. J Int Financ Manag Acc 28:131–171

Altman EI, Hotchkiss E, Wang W (2019) Corporate financial distress, restructuring, and bankruptcy: analyze leveraged finance, distressed debt, and bankruptcy, 4th edn. Wiley, Hoboken, NJ

Amendola A, Restaino M, Sensini L (2015) An analysis of the determinants of financial distress in Italy: a competing risks approach. Int Rev Econ Financ 37(2015):33–41

Anderson JA (1972) Separate sample logistic discrimination. Biometrika 59:19–35

Aziz MA, Dar HA (2006) Predicting corporate bankruptcy: where we stand? Corp Gov 6(2):18–33

Balcaen S, Ooghe H (2004) 35 years of studies on business failure: an overview of the classical statistical methodologies and their related problems. Working Paper No. 248/2004, Universiteit Gent

Balcaen S, Manigart S, Buyze J, Ooghe H (2012) Firm exit after distress: differentiating between bankruptcy, voluntary liquidation and M&A. Small Bus Econ 39:949–975

Barboza F, Kimura H, Altman E (2017) Machine learning models and bankruptcy prediction. Expert Syst Appl 83(Oct.):405–417

Beaver WH (1966) Financial ratios as predictors of failure. J Account Res 4:71–111

Beaver WH, Correia M, McNichols M (2010) Financial statement analysis and the prediction of financial distress. Found Trends Acc 5(2):99–173

Bishop YM, Fienberg SE, Holland PW (2007) Discrete multivariate analysis: theory and practice. Springer Science & Business Media, New York (First published 1975 by MIT Press, Cambridge, MA)

Bodle KA, Cybinski PJ, Monem R (2016) Effect of IFRS adoption on financial reporting quality: evidence from bankruptcy prediction. Account Res J 29:292–312

Boratyńska K, Grzegorzewska E (2018) Bankruptcy prediction in the agribusiness sector: lessons from quantitative and qualitative approaches. J Bus Res 89:175–181

Cameron AC, Trivedi PK (2005) Microeconometrics: methods and applications. Cambridge University Press, New York

Campbell JY, Hilscher JD, Szilagyi J (2008) In search of distress risk. J Financ 63:2899–2939

Campbell JY, Hilscher JD, Szilagyi J (2011) Predicting financial distress and the performance of distressed stocks. J Invest Manag 9(2):14–34

Cefis E, Marsili O (2012) Going, going, gone: exit forms and the innovative capabilities of firms. Res Policy 41(5):795–807

Chancharat N, Tian G, Davy P, McCrae M, Lodh S (2010) Multiple states of financially distressed companies: tests using a competing-risks model. Australasian Account Bus Finan J 4(4):27–44
Chava S, Jarrow RA (2004) Bankruptcy prediction with industry effects. Rev Finan 8:537–569
Chen J, Marshall BR, Zhang J, Ganesh S (2006) Financial distress prediction in China. Rev Pac Basin Financ Mark Policies 9(02):317–336
Cheng WY, Su E, Li SJ (2006) A financial distress pre-warning study by fuzzy regression model of TSE-listed companies. Asian Acad Manag J Account Finan 2(2):75–93
Choi H, Son H, Kim C (2018) Predicting financial distress of contractors in the construction industry using ensemble learning. Expert Syst Appl 110(Nov.):1–10
Ciesielski P (2005) Prognozowanie upadłości podmiotów gospodarczych w Polsce (Predicting bankruptcy of companies in Poland), Rachunkowość (Accounting), No. 8/2005
Cramer JS (1999) Predictive performance of the binary logit model in unbalanced samples. The Statistician 48(Part 1):85–94
Cressy R, Farag H (2014) Stairway to heaven or gateway to hell? A competing risks analysis of delistings from Hong Kong's growth Enterprise market. Int Rev Financ Anal 36(2014):195–205
Dahiya S, Saunders A, Srinivasan A (2003) Financial distress and bank lending relationships. J Financ 58(1):375–399
Domański C, Szreder M (2010) Demografia oderwana od populacji ludzkiej? (is demography out of touch with human population). Wiadomości Statystyczne 12(2010):36–39
Esteve-Pérez S, Mañez-Castillejo JA (2008) The resource-based theory of the firm and firm survival. Small Bus Econ 30:231–249
Gerakos JJ, Hahn R, Kovrijnykh A, Zhou F (2016) Prediction versus inducement and the informational efficiency of going-concern opinions, Chicago Booth Research Paper No. 16-01 (Jan.)
Gruszczyński M (2004) Financial distress of companies in Poland. Int Adv Econ Res 10(4): 249–256. Also available as Gruszczyński M (2004) Financial distress of companies in Poland. Working Paper No. 1-04, Department of Applied Econometrics, SGH Warsaw School of Economics
Gruszczyński M (2005) Zalety i słabości modeli bankructwa (Advantages and weaknesses of bankruptcy models). In: Kuciński K, Mączyńska E (eds) Zagrożenie upadłością (Danger of bankruptcy), IFGN Working Papers, SGH Warsaw School of Economics
Gruszczyński M (ed) (2012) Mikroekonometria, 2nd edn. Wolters Kluwer, Warszawa
Gruszczyński M (2015) Issues in modelling the financial distress and bankruptcy of companies. Quant Methods Econ XVI(1): 96–107. Also available as Gruszczyński M (2015) Issues in modelling the financial distress and bankruptcy of companies. Applied Econometrics Papers, vol. 2(1). Department of Applied Econometrics, SGH Warsaw School of Economics
Gruszczyński M (2017) Błędy doboru próby w badaniach bankructw przedsiębiorstw (sample bias in the research on corporate bankruptcy). Kwartalnik Nauk o Przedsiębiorstwie 44(3):22–29
Gruszczyński M (2019) On unbalanced sampling in bankruptcy prediction. Int J Financ Stud 7 (2):28
Hensher DA, Jones S (2008) Mixed logit and error component model of corporate insolvency and bankruptcy risk. In: Jones S, Hensher DA (eds) Advances in credit risk modeling and corporate bankruptcy prediction. Cambridge University Press, Cambridge, pp 44–79
Hillegeist SA, Keating EK, Cram DP, Lundstedt KGL (2004) Assessing the probability of bankruptcy. Rev Acc Stud 9:5–34
Hotchkiss E, John K, Mooridian RM, Thorburn KS (2008) Bankruptcy and the resolution of financial distress. In: Eckbo BE (ed) Handbook of corporate finance: empirical corporate finance, North-Holland handbook of finance series, vol 2. Elsevier, Amsterdam, pp 235–290
Jones S, Johnstone D, Wilson R (2017) Predicting corporate bankruptcy: an evaluation of alternative statistical frameworks. J Bus Financ Acc 44(1–2):3–34
Jung H, Hwang JT, Kim B-K (2018) Does R&D investment increase SME survival during a recession? Technol Forecast Soc Change 137:190–198
Kaiser U (2001) Moving in and out of financial distress: evidence for newly founded services sector firms. ZEW Discussion Paper Nr 01-09, Zentrum für Europäische Wirtschaftsforschung, Mannheim

King G, Zeng L (2001) Logistic regression in rare events data. Polit Anal 9(2):137–163
Kukuk M, Rönnberg M (2013) Corporate credit default models: a mixed logit approach. Rev Quant Finan Acc 40:467–483
Lagares EC, Ordaz FG, del Hoyo JJG (2018) Innovation, environmental commitment, internationalization and sustainability: a survival analysis of Spanish marine aquaculture firms. Ocean Coast Manag 151:61–68
Lau AH-L (1987) A five-state financial distress prediction model. J Account Res 25(1):127–138
Lennox C (1999) Identifying failing companies: a reevaluation of the logit, probit and DA approaches. J Econ Bus 51:347–364
Linden H (2015) Synthesis of research studies examining prediction of bankruptcy. Master Thesis, Aalto University School of Business
Long JS, Freese J (2014) Regression models for categorical dependent variables using Stata, 3rd edn. Stata Press, College Station, TX
Lopucki LM, Doherty JW (2015) Bankruptcy survival. 62 UCLA Law Rev 970:970–1015
Maalouf M, Homouz D, Trafalis TB (2018) Logistic regression in large rare events and imbalanced data: a performance comparison of prior correction and weighting methods. Comput Intell 34 (1):161–174
Maddala GS (1983) Limited dependent and qualitative variables in econometrics. Cambridge University Press, Cambridge
Maddala GS (1991) A perspective on the use of limited-dependent variables models in accounting research. Account Rev 66(4):788–807
Makridakis S, Spiliotis E, Assimakopoulos V (2018) The M4 competition: results, findings, conclusion and way forward. Int J Forecast 34(4):802–808
Manjón-Antolín MC, Arauzo-Carod J-M (2008) Firm survival: methods and evidence. Empirica 35:1–24
Manski CF, Lerman SR (1977) The estimation of choice probabilities from choice based samples. Econometrica 45(8):1977–1988
Matsuno S, Uchida Y, Ito T (2017) A survival analysis of the Japanese information service industry. Procedia Comput Sci 121:291–296
Merton R (1974) On the pricing of corporate debt: the risk structure of interest rates. J Financ 29:449–470
Moody's (2000) Moody's Investor Service, Moody's approach to evaluating distressed exchanges. White Paper, July 2000
Nehrebecka N (2018) Predicting the default risk of companies. Comparison of credit scoring models: logit vs support vector machines. Econom Adv Appl Data Anal 22(2):54–73
Ohlson JA (1980) Financial ratios and the probabilistic prediction of bankruptcy. J Account Res 18 (1):109–131
Outecheva N (2007) Corporate financial distress: an empirical analysis of distress risk. Doctoral Dissertation, University of St. Gallen HSG
Paradysz J (2011) Apologia demografii czyli istnienie demografii bez ludności (apology of demography or demography without humans). Wiadomości Statystyczne 2:27–35
Platt HD, Platt MB (1991) A note on the use of industry-relative ratios in bankruptcy prediction. J Bank Financ 15:1183–1194
Platt HD, Platt MB (2002) Predicting corporate financial distress: reflections on choice-based sample bias. J Econ Financ 26(2):184–199
Platt HD, Platt M (2006) Comparing financial distress and bankruptcy. Rev Appl Econ 2 (2):141–157
Prusak B (2018) Review of research into enterprise bankruptcy prediction in celected central and eastern European countries. Int J Financ Stud 6(3):60
Ross SA, Westerfield RW, Jaffe J, Jordan B (2015) Corporate finance, 11th edn. McGraw Hill Irwin, New York
Senbet LW, Wang TY (2010) Corporate financial distress and bankruptcy: a survey. Found Trends Financ 5(4):243–335

Shin KS, Lee TS, Kim HJ (2005) An application of support vector machines in bankruptcy prediction model. Expert Syst Appl 28:127–135

Shumway T (2001) Forecasting bankruptcy more accurately: a simple hazard model. J Bus 74:101–124

Skogsvik K, Skogsvik S (2013) On the choice-based sample bias in probabilistic bankruptcy prediction. Invest Manag Financ Innovat 10(1):29–37

Śmigielski J, Majdzińska A, Śmigielski W (2010) Using ROC curves to find the cut-off point in logistic regression with unbalanced data samples. Stat Trans 11:381–402

Staszkiewicz P, Witkowski B (2018) Failure models for insolvency and bankruptcy, Contemporary trends and challenges in finance. Springer proceedings in business and economics ed. by Jajuga K, Locarek-Junge H, Orlowski L. Cham: Springer

Tinoco MH, Wilson N (2013) Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. Int Rev Financ Anal 30:394–419

Tinoco MH, Holmes P, Wilson N (2018) Polytomous response financial distress models: the role of accounting, market and macroeconomic variables. Int Rev Financ Anal 59:276–289

Train K (2009) Discrete choice methods with simulation, 2nd edn. Cambridge University Press, Cambridge

Wagenmans F (2017) Machine learning in bankruptcy prediction. MSc thesis, ICA-3870154, Utrecht University Repository

Winkelmann R, Boes S (2006) Analysis of microdata. Springer, Heidelberg

Zhou L (2013) Performance of corporate bankruptcy prediction models on imbalanced dataset: the effect of sampling methods. Knowl-Based Syst 41:16–25

Zięba M, Tomczak SK, Tomczak JM (2016) Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. Expert Syst Appl 58:93–101

Zmijewski M (1984) Methodological issues related to the estimation of financial distress prediction models. J Account Res 20:59–82

Chapter 4
Accounting Research and Disclosure Microeconometrics

This chapter is devoted to methodological applications in accounting research, beginning with a survey of topics and methods rooted in microeconometrics. Financial microeconometrics methodology appears appropriate in most studies where researchers in accounting employ large datasets of companies, of financial statements, of business events, etc. A major portion of the chapter presents in detail an area of accounting research called "disclosure microeconometrics"—analyses of the associations between the level of corporate disclosure and various categories relating to investor protection.

4.1 Topics in Empirical Accounting Research and Sources of Knowledge

Introduction

Empirical research in accounting uses a methodology and data sources analogous to those in corporate finance. The research questions are also parallel. For example, the authors of three papers in the 2018/5 issue of the *Journal of Accounting Research* discuss (1) how the risk assessment by auditors is related to the previous year's assessment from the standpoint of audit workpapers (prepopulated or not), (2) how managerial compensation relates to the manipulation of short-term stock prices, and (3) how negative news transmitted by CEOs on Twitter and then repeated by the company's Investor Relations Twitter account relate to investor perceptions. The topics of empirical accounting concentrate on the audit (external and internal), accounting standards, risk assessment, disclosure, cost allocation, valuation, among many other subjects, and generally with extensive use of quantitative methods.

M. Gruszczyński, *Financial Microeconometrics*,
https://doi.org/10.1007/978-3-030-34219-7_4

Research works in empirical accounting are reported in journals, working papers, books, etc. Notable journals are *Accounting and Finance*, *Accounting Review*, *Advances in International Accounting*, *Contemporary Accounting Research*, *European Accounting Review*, *International Journal of Accounting*, *Journal of Accounting and Economics*, *Journal of Accounting and Public Policy*, *Journal of Accounting Research*, *Journal of Management Accounting Research*, *Review of Quantitative Finance and Accounting*. While these are all journals with "accounting" in their titles, there are other influential journals that also publish accounting papers, including those devoted to corporate finance. Likewise, accounting publications collected by repositories are worth mentioning—e.g., the Accounting Research Network in SSRN (especially the *Research Methods & Methodology in the Accounting eJournal*).

In addition, we should also mention the following selected books on accounting research methodology: *Handbook of Management Accounting Research* (Chapman et al. 2006a, b, 2009), *Methodological Issues in Accounting Research* 2nd ed. (Hoque 2018), *Advances in Quantitative Analysis of Finance and Accounting* (Lee 2004–2008), *Research Methods for Accounting and Finance* (Paterson et al. 2016), and *Research Methods in Accounting* 4th ed. (Smith 2017).

Categories of Research Topics and Methods

The subject of empirical studies in accounting can be seen in more detail in the list composed by the *Journal of Accounting Research* showing research topics historically published in that journal. In 2018 these were:

– The impact of financial reporting and disclosure on stock prices.
– The economics of auditing, enforcement, and audit oversight.
– The use of accounting information in contracting in the debt, labor, supply, and other markets.
– The role of accounting in compensation and in corporate governance.
– The role of managerial accounting in internal decision-making such as budgeting, costing, and transfer pricing.
– The real effects of financial reporting and disclosure.
– The economics of the regulation of financial reporting and disclosure including bank regulation.
– International differences in financial reporting and the role of reporting standards in international capital markets.
– The political economy of accounting standard setting.
– The use of accounting information in public finance.
– The impact of tax regulation on transaction structuring.

Australia

Other catalogues of accounting research topics have been published in several survey articles. In the anniversary issue of *Accounting and Finance,* Chenhall and Smith (2011) assess the papers of Australian researchers in 10 leading management accounting journals during the period 1980–2009. Out of the 231 papers published

– The most frequent topics are management control systems (65 papers), budgeting (49), performance measurement (38), and methodology issues (13).
– The most commonly used settings are the manufacturing and service sectors (a total of 109 papers), followed by hospitals/healthcare (18), and the public sector (13).
– Methodology-wise, papers presenting surveys are most common (93 papers), followed by case studies/interviews (56), literature reviews (32), and experiments (24).
– In terms of theories, organizational behavior theory attracts the most papers (60), followed by contingency theories (48), psychology (39), and sociology (34).
– The most popular data analysis technique is regression (70 papers), then analysis of variance (24), and structural equation modeling/path analysis (20).

The palette of topics, theories, and methods in Australian accounting research is broad and comprehensive. Benson et al. (2015) convey an analogous message in their review of accounting research in the Asia Pacific region.

USA

Coyne et al. (2010) in *Issues of Accounting Education* provide a ranking of accounting research programs based on the publications of 517 academics, members of the American Accounting Association, in 11 leading accounting journals during the period 1990–2009. The ranking itself is not that important here, but what is interesting is the distribution of publications in terms of methodology and topic. The authors classify the papers in terms of methodology into:

– *Analytical* (A) *studies* "whose analyses and conclusions are based on the act of formally modeling theories or substantiating ideas in mathematical terms"
– *Historical* (H)/*archival studies* in which the researchers create their own repositories of data and base their reasoning on this information
– *Experimental* (E) *studies* "whose analyses and conclusions are based on data the researcher gathered by administering treatments to subjects"
– *Other* (O)

In terms of topic, the papers are divided into:

– *Accounting information systems*: Systems for gathering, storing, and generating information for accounting

– *Auditing*: The audit environment (external and internal), auditor decision-making, auditor independence, the effects of auditing on the financial reporting process, and auditor fees
– *Financial studies*: The content of financial accounting, financial markets, and decision-making based on financial accounting information
– *Managerial studies*: Issues regarding budgeting, compensation, incentives, and the allocation of resources within a company
– *Tax studies*: Issues related to taxpayer decision-making, tax allocations, tax computations, structuring of accounting transactions to meet tax goals, tax incentives, and market reactions to tax disclosures
– *Other studies*

Table 4.1 shows the distribution of papers in this survey. It emerges that American authors use mostly methods that are based on observational (H) and experimental (E) data. The ranking itself gave the overall lead to researchers at Stanford University and the University of Texas (Austin).

An analysis by Dunbar and Weber (2014), also in *Issues of Accounting Education*, focuses on papers in nine leading accounting journals during the period 1996–2011 from the standpoint of "individual antecedent works that have been cited the most often by accounting research." Their primary table shows the distribution of 3538 papers classified into topics and methods. The named topics are audit, financial, managerial, tax, and others. As for methods, the categories are archival (AR), experimental (EX), theoretical (TH), and other (O).

Table 4.2 presents the classification of the accounting papers from the study by Dunbar and Weber (2014). As above, most papers are based on the observational (AR) and experimental (EX) approaches.

Table 4.1 Classification of accounting papers in the survey by Coyne et al. (2010)

Topical area/method	A (%)	H (%)	E (%)	O (%)
Information systems	4	12	36	48
Audit	8	32	38	23
Financial studies	12	76	7	5
Managerial studies	22	22	16	40
Tax studies	13	58	19	10
Other	4	22	16	59

The rows sum roughly to 100%
Type of studies: *A* analytical, *H* historical, *E* experimental, *O* other

Table 4.2 Classification of accounting papers in the survey by Dunbar and Weber (2014)

Topical area/method	AR (%)	EX (%)	TH (%)	O (%)
Audit	43	34	7	16
Financial studies	77	6	11	5
Managerial studies	21	17	23	38
Tax studies	67	14	11	7
Other	12	2	0	86

The rows sum roughly to 100%
Methods: *AR* archival, *EX* experimental, *TH* theoretical, *O* other

Europe

Papers presented at the 2003–2008 annual congresses of the European Accounting Association (EEA) were examined by Fülbier and Sellhorn (2008). Table 4.3 shows the topical breakdown of papers delivered at those conferences. The topics are considered here in greater detail than above. Researchers in Europe cover primarily management accounting, financial reporting, corporate governance, international accounting, and financial accounting—a combined 51% of the papers.

Table 4.4 presents Fülbier and Sellhorn's (2008) list of research methods in accounting compiled from the papers presented at the annual congresses of the European Accounting Association in 2000 and 2005. It shows that research methods classified as "empirical archival methods" are the most popular, as some 70% of the papers at the EAA congress in 2005 were so classified. This means that research is based largely on statistical-econometric-type analyses such as those under the financial microeconometrics label.

Table 4.3 Topical breakdown of papers presented at the EAA Annual Congresses (2003–2008)

Subject area	% of papers	Subject area	% of papers
Acct. and capital markets	4	Finance/financial management	1
Acct. education and research	3	Financial acct. (capital markets)	8
Acct. history	3	Financial reporting	11
Acct. and information systems	2	Corporate governance	8
Acct. and strategy	1	International acct.	8
Acct. theory	3	Management acct.	16
Auditing	6	Behavioral aspects of acct.	5
Critical perspectives on acct.	3	Public sector and nonprofit acct.	7
Economic analytical acct.	2	Social and environmental acct.	5
Financial analysis	3	Acct. and taxation	2

Source: Fülbier and Sellhorn (2008)
acct. accounting

Table 4.4 Research methods of papers presented at the EAA Annual Congresses: 2000 and 2005

Methods	2000 (%)	2005 (%)
Empirical archival—Database or archive	51	70
Empirical experiment	2	0
Empirical field or case study	4	1
Empirical survey	7	6
Nonempirical—Analytical	0	1
Nonempirical—Theory	7	6
Other	21	10
Ambiguous	9	6
Total	100	100

Source: European Accounting Association, Fülbier and Sellhorn (2008)

Table 4.5 Classification of accounting papers by topic in the Machado and Ribeiro survey (2016)

Subject area	% of papers
Financial accounting	42
Management accounting	26
Auditing	11
Other	21

Table 4.6 Classification of accounting papers by data collection method in the Machado and Ribeiro survey (2016)

Method of data collection	% of papers
Existing databases	40
Literature review	28
Case studies	14
Surveys	10
Interviews	8

We close this selection of reviews with an example of a survey devoted to just one journal. Papers that appeared in the *European Accounting Review* during the period 2007–2011 were classified by Machado and Ribeiro (2016). The set contains 127 papers, mostly devoted to financial and managerial accounting (Table 4.5). As for methodology, the authors provide only a breakdown of "data collection methods" (Table 4.6).

From the evidence in this section, we may state that accounting papers in major journals and at conferences worldwide extensively exploit statistical-econometric methodology. There is a growing trend in using various techniques of advanced data analysis. What is more, the topics of "accounting" papers are parallel or analogous to the subjects elaborated in "corporate finance" and "corporate governance" literature.

Probability Expressions in Accounting

Before embarking on other topics, let us mention an issue that is an integral part of accounting/auditing and fits perfectly with quantitative discussions. The issue is an auditor's assessment of the probability of an event occurring, the best example of which is the going-concern opinion issued by a company's auditor.

Researchers have attempted to reveal what level of probability might be associated with the auditor-specific statement. Table 4.7 compiled by Silska-Gębka (2017) shows the results of five studies based on surveys of accountants, auditors, and people engaged in creating accounting standards.

Table 4.7 demonstrates that there is no agreement as to the probability level understood by various respondents. Studies reveal cultural, language, and translation differences in understanding the same standard expressions. Recent research on this includes Huerta et al. (2016) and Du et al. (2016).

Table 4.7 Interpretation of probability phrases in five studies: average probability values

	Reimers (1992)	Amer et al. (1994)	Laswad and Mak (1997)	Simon (2002)	Doupnik and Richter (2003)
Virtually certain	–	–	0.96	0.94	0.92
Highly probable	0.84	0.87	–	–	–
Expected	0.84	–	0.72	0.79	0.80
Reasonably certain	–	–	0.79	0.85	–
Probable	0.78	0.79	0.65	0.73	0.71
Likely	0.78	0.75	0.67	0.74	0.71
Reasonably possible	0.58	0.59	0.43	0.51	–
Possible	0.52	0.49	0.33	0.42	–
Unlikely	–	0.20	0.18	0.18	0.27
Highly unlikely	0.14	–	–	–	–
Remote	0.09	0.12	0.07	0.06	0.16

Source: Silska-Gębka (2017)

Qualitative variables that represent the statements of accountants and auditors can be modeled using the methodology of microeconometrics. An example of such research is shown in Gerakos et al. (2016), who model the probability of the going-concern audit opinion and its relation to the probability of bankruptcy.

4.2 Microeconometric Methodology in Accounting Research

The quantitative methods employed in accounting research—including those of microeconometrics—are not much different from those applied in corporate finance. In both areas, there are similar problems with the quality of samples, the adequacy of a specific model, the selection of covariates, correlation versus causation, etc. These issues were initially mentioned in the set of good practices in Sect. 2.13.

In this section, we discuss a few methodological issues, especially those mentioned in the surveys of Ge and Whitmore (2010), Cram et al. (2007), and Lennox et al. (2012).[1] We also present two examples of applied accounting studies.

[1]For more on these issues, see Gruszczyński (2009).

Surveys by Maddala (1991) and Ge and Whitmore (2010)

Arguably the most important paper introducing microeconometric issues in accounting is that of Maddala (1991). This is the first survey-like presentation of the application of the limited-dependent and the qualitative variables models in accounting research. The author discusses binomial models (logit, probit) and linear discriminant analysis, as well as tobit and truncated regression. The issues of sample selection and self-selectivity are also discussed.

The numerous attempts to use microeconometric models in accounting have come with mixed results in terms of the quality of the research. Ge and Whitmore (2010) reviewed more than 30 articles published in accounting journals. The classic Maddala (1991) paper on the dangers of modeling qualitative variables in accounting research is still valid. Most of the articles researched by Ge and Whitmore (2010) have ambiguities and errors in the presentation of the logistic regression model. The authors point out that the incorrect presentation of a model—even in conjunction with a correct analysis—may lead to serious misinterpretation of the research findings.

Choice-Based and Matched Samples (Cram et al. 2007)

In their paper entitled "Three threats to validity of choice-based and matched sample studies in accounting research," Cram et al. (2007) discuss the dangers connected with the use of some popular nonrandom sampling schemes. The authors reviewed 73 audit research papers and pointed out that most of them are exposed to at least one of the threats mentioned in the article's title.

Choice-based samples occur when groups making different choices are sampled at different rates. This issue was thoroughly discussed in Chap. 3. Matched samples are two samples in which the members are clearly paired or are matched explicitly by the researcher. This question was examined in Chap. 2.

The three threats mentioned in the paper's title are:

1. *The use of unconditional analysis when an analysis conditional upon the results of matching variables is needed*. "Researchers believe that the selection of a matched sample already controls for the matching variables. An unconditional analysis is performed, rather than the conditional one that is justified. Failure to account for industry, size, and other matching variables may have driven incorrect findings in many research studies." The authors' guidance to researchers is either to avoid use of matching or to take matching into account when analyzing the data.
2. *The failure to control for the result of imperfectly matched variables*. "Where matching is by 'closest' size or other continuous measure, the matching is imperfect, and there remains the possibility that case vs. control differences in this matching variable could be the cause of differences in outcome, so

researchers must evaluate that possibility and perhaps control for it." The authors advise researchers either to avoid imperfect matching or to perform and report sensitivity analyses on how imperfection in the matching might have influenced the outcomes.

3. *The failure to re-weight observations according to differing sampling rates*. "The disproportionate sampling for different population strata (implicit in the choice-based and matched sample selection) would usually necessitate weighting data by the sampling rates in each stratum, but re-weighting or other appropriate adjustment to the analysis is often not implemented." The authors suggest that choice-based and matched sampling should be avoided unless explicit sampling rate information can be obtained (allowing for explicit re-weighting) or unless the logit regression will suffice to analyze the research questions (taking advantage of the logit exemption to the need for reweighting). This very issue has been discussed under "prior correction" in Chap. 3.

Sample Selection (Lennox et al. 2012)

Sample selection methods were introduced in Chap. 2. They are just as applicable in accounting research as any other microeconometric methodology.

As we have shown previously, selectivity occurs whenever observations self-select into two discrete groups, resulting in potential coefficient bias. The traditional approach to controlling for self-selection bias is the two-step procedure developed by Heckman (1976) and known as the Heckit method (see Sect. 2.10). In the first step, the *selection equation* (binomial probit) is estimated on the full sample with the explanatory variables Z and is used to estimate inverse Mills ratios (*IMR*s) for each observation selected into one group. In the second step, the *IMR*s are included in the *outcome equation* (the primary model of interest) as a control for the results of selection. The selection equation is estimated only for observations from one group (selected) with the explanatory variables X and the *IMR* as an additional covariate. Variables X and Z may overlap. Those Z variables which are not overlapping with X in the outcome equation are called *exclusion restrictions* because the researcher assumes that they have no direct impact on the outcome.

Lennox et al. (2012) analyzed 75 papers that use selection models and were published during the 2000–2009 period in *The Accounting Review*, *Journal of Accounting and Economics*, *Journal of Accounting Research*, *Review of Accounting Studies*, or *Contemporary Accounting Research*. The survey shows that many researchers apply the Heckman methodology mechanically, without proper care.

For example, the explained variable in the selection equation equals $Y = 1$ if the company chooses an auditor from the Big N, otherwise $Y = 0$. In the outcome equation, we may be interested in modeling the cost of equity or the audit fee. To successfully identify selectivity, the researcher should include in the first stage choice model at least one exogenous independent variable that has no direct effect on the dependent variable in the second stage regression (i.e., can be validly excluded from it). The authors of the survey point out that such exclusion restrictions are rarely recognized in the accounting literature. In 14 out of 75 articles, there were no exclusion restrictions; moreover, seven papers did not report the results of the first-stage estimation (the selection equation).

There are two important observations made by the survey's authors:

- Claims about the existence and direction of the selection bias can be sensitive to the researcher's set of exclusion restrictions.
- The selection model is vulnerable to high levels of multicollinearity; it may cause the insignificance of *IMR* and unreliably indicate the absence of selection bias.

On this point, the authors state that "researchers need to critically appraise the quality of their exclusion restrictions and assess whether there are problems of fragility and multicollinearity in their specific empirical setting that might limit the effectiveness of selection models relative to OLS."

Financial Microeconometrics in Accounting: Two Examples

Microeconometric research in accounting covers many areas that have been mentioned in the previous section. Below we indicate several issues connected with the use of microeconometric methodology. Let us turn to two examples connected with auditors: a study on changing auditors and research into the determinants of going-concern opinions.

Example 4.1 Choice of Audit Office After Auditor Change: Audit Analytics' Auditor Changes Data for US Firms

Li et al. (2017) present research into the influence of peer (i.e., similar) firms' prior choices of audit firms as to whether a client company chooses to engage a "social norm" audit office in its metropolitan area as the replacement for its previous auditor (i.e., auditor change). Specifically, does a company choose an auditor in the "neighborhood" (in one of four dimensions, including geographic) after deciding to change the auditor firm? "The office in a metro area auditing the largest number of peer firms along a given similarity dimension is considered to be the social norm office" for that dimension. The authors recognize four dimensions of similarity: client geographic location, industry affiliation, client size (filing status), and departing auditor type (Big N versus non-Big N).

The authors use US data from the database of Audit Analytics' Auditor Changes for the years 2001–2012. There are 4074 firm-year observations. Research reveals that

> controlling for the determinants of auditor selection documented in prior literature, the propensity of a client to select a social norm audit office in a given year (after auditor turnover) is positively associated with (1) the proportion of its peers who are affiliated with that office in the prior year, consistent with the prediction of social norms theory, and (2) the proportion of its auditor-switching peers who selected that office in the prior year, consistent with the prediction of social learning theory.

The study is based on theories of social learning and social norms. These theories, originally explaining the behavior of individuals, postulate that, for example, an agent observes and then imitates the behavior of other agents experiencing a similar situation. In the case of firms switching auditors, this means that "the norm" would be to follow peer firms' behavior in similar circumstances—i.e., to choose the "social norm audit office."

The authors introduce two measures (variables) representing social learning and the social norm:

- SL = *social learning*: The number of listed peer (similar) client companies *choosing* that audit office in the prior year, divided by the total number of listed peer client companies choosing any office in the metro area (single dimension) in the prior year.
- SN = *social norm*: The number of listed peer (similar) client companies *affiliated* with that audit office in the prior year, divided by the total number of listed peer client companies affiliated with any office in the metro area (single dimension) in the prior year.

Single dimension means one of (1) the same location, (2) the same industry, (3) the same size, and (4) the same previous auditor type. In the models estimated in this study, the dependent variable is

- NORMOFFICE(i, t) = 1 if a client's succeeding auditor is the social norm audit office (i.e., the office having the largest number of peer clients in the metro area where the ith client is headquartered in year t), = 0 otherwise.

The social norm audit office is identified separately for each of the four proxies of peer firms: geographic location, industry affiliation, client size, and departing auditor type (Big N versus non-Big N).

Two major explanatory variables for testing the authors' research hypotheses are

- NORM(j, $t-1$) = measure SN (above) in jth peer group proxy
- LEARN(j, $t-1$) = measure SL (above) in jth peer group proxy.

The authors also consider variables DISNORM and DISLEARN to indicate "distant peers"[2]:

- DISNORM(j, $t-1$) = the proportion of listed firms (clients) audited by the office having the most such clients in year $t-1$, but being different in terms of the jth dimension (different metro area, different industry, opposite size, or opposite auditor type)

[2]The authors state that it is unlikely that the switching ith client would use social learning (or social norm) evidence from different peer clients (another metro area, another industry, etc.). Such distant peer analysis should provide information about the divergent validity of the social norm and social learning test variables—i.e., a falsification test (Li et al. 2017).

Table 4.8 The binomial probit model of audit office choice 2001–2012

Dependent variable: *NORMOFFICE*(i,t) = 1 if client's succeeding auditor is the social norm audit office (i.e., the office having the largest number of peer clients in the metro area where the ith client is headquartered in year t and that is of the same auditor type as client's ith former auditor); = 0 otherwise.			
Explanatory variable/probit estimates	**Regression 1**	**Regression 2**	**Regression 3**
Constant	−1.425***	−1.323***	−1.481***
Test variables:			
NORM	1.583***		1.578***
DISNORM		0.342**	0.303**
Control variables: Omitted			
Test of coefficient difference			41.32***
			(F-stat)
Number of observations	4074	4074	4074
N for *NORMOFFICE* = 1	668	668	668
Year effect	Yes	Yes	Yes
Wald chi^2	307.2	150.5	316.7
p-value	0.000	0.000	0.000
Pseudo R^2	0.095	0.043	0.096

Relation to social norm: previous auditor type similarity
Source: Li et al. (2017)
*** means statistical significance at 0.01; ** significance at 0.05

– DISLEARN(j, $t-1$) = the proportion of auditor-switching listed firms (clients) choosing the audit office having attracted the most new listed clients in year $t-1$, but being different in terms of the jth dimension.

In addition, there are 25 control variables representing the financial, market, and audit characteristics of the firms (i.e., the clients of audit firms). Also, fixed year effects are included.

In Table 4.8, we present the estimation results of one of the models from this study (Table 4.6). This is a binomial probit in which the dependent variable is the choice of the social norm audit office along one of the four proxies for peer firms, namely the similarity of the previous auditor type (Big N versus non-Big N).

This particular binomial model has significantly positive coefficients of variables NORM and DISNORM. The authors state that in regression 3 the variable NORM has "much larger positive effect than DISNORM (F-stat = 41.32, $p < 0.01$)." In their opinion, the norm effect of the firm's choice of auditor from its close peers—who are audited by the same type of auditor as its prior auditor—is much stronger than the choices of distant peers (audited by auditors of the type different from its prior auditor).

Analyses by Li et al. (2017) support the hypothesis which predicts that "established audit office affiliations of close, similar peers in a metro area have a significant impact on an individual client's auditor selection, consistent with the implications of social norms theory."

■

Example 4.2 Auditors' Going-Concern Opinions and Managerial Earnings Forecasts

Feng and Li (2014) examine the relationship between managerial earnings forecasts and auditors' going-concern decisions. They find that management earnings forecasts are negatively associated with auditors' going-concern opinions. The study also includes modeling of bankruptcy. All in all, the authors provide evidence that auditors are professionally skeptical about earnings forecasts when making going-concern decisions.

The authors' study involves data for financially distressed firms from Audit Analytics, First Call, and Compustat for the fiscal years 2000–2010. A firm is financially distressed if it reports either negative net income or negative operating cash flow in the current fiscal year. The initial sample comprised 24,370 firm-year observations. After adjusting for the availability of management earnings forecasts issued prior to the going-concern decisions, the final sample for the examination of the association between managerial forecasts and auditors' going-concern opinions comprises 1054 firm-year observations.

The research strategy amounts to estimating the Heckman sample selection model. Why? The study investigates how auditors view earnings forecasts when assessing a firm's going-concern status. Since providing a managerial forecast is voluntary, it is possible that the sample is biased due to self-selection. Thus, in the first stage of the Heckman procedure, the selection equation is estimated. It is a binomial probit with the dependent variable $GUIDANCE_t$ which equals 1 if a manager issues at least one earnings forecast in year t, and equals 0 otherwise. In the outcome equation, the modeled variable is the binomial GC_t which is equal to 1 if a firm receives a going-concern opinion in year t, and equals 0 otherwise. Obviously, the outcome equation includes an inverse Mills ratio (*IMR*) variable estimated from the selection equation (see Sect. 2.10 on the Heckit procedure).

The model of the outcome equation is the binomial logit with *GC* as the dependent variable. The list of explanatory variables is as follows:

- $FORECAST_{t+1}$ = Management's EPS forecast for year $t + 1$ scaled by assets per share at the end of year t.
- $SIZE_t$ = The natural log of the client's total sales at the end of year t.
- $ZSCORE_t$ = The bankruptcy probability score measured at the end of year t, where the bankruptcy score is calculated as $-4.3 - 4.5 \times$ (net income/total assets) + $5.7 \times$ (total debt/total assets) $- 0.004 \times$ (current assets/current liabilities).
- $LEVERAGE_t$ = Total liabilities divided by total assets at the end of year t.
- $PRLOSS_t$ = 1 if the company reported negative net income in year $t-1$, = 0 otherwise.
- $CHGDT_t$ = The change in long-term debt from year $t-1$ to year t divided by total assets at the end of year $t-1$.
- CFO_t = Cash flow from operations divided by total assets at the end of year t.

- $LIQUIDITY_t$ = The sum of the firm's cash and investment securities (long- and short-term) divided by total assets at the end of year t.
- $MKBK_t$ = The market value of the company's common stock divided by the book value of the common stock at the end of year t.
- $LITIGATE_t$ = 1 if the company is in a litigious industry, = 0 otherwise.
- $REPORTLAG_t$ = The number of days between fiscal year-end and the date of signing the auditor's report for year t.
- $PRIORGC_t$ = 1 if the company receives a going-concern opinion in year $t-1$, = 0 otherwise.
- $NEWEQUITY_{t+1}$ = New equity issued in year $t + 1$ divided by total assets at the end of year t.
- $NEWDEBT_{t+1}$ = New debt issued in year $t + 1$ divided by total assets at the end of year t.
- $DEPENDENCE_t$ = The ratio of the client's total audit fees divided by the total audit office revenue for year t.
- $BIG\ N_t$ = 1 if the auditor is Big 4 (5) at the end of year t, = 0 otherwise.
- $INDEXPERT_t$ = 1 if the auditor is the industry leader at both national and city levels at the end of year t, = 0 otherwise.
- $TENURE_t$ = The natural log of the total number of years that the client has engaged the auditor at the end of year t.
- IMR_t = The inverse Mills ratio (*IMR*) generated from the first-stage model.

Table 4.9 presents the estimation results of the outcome equation. The significant estimate by the inverse Mills ratio (IMR_t) validates the sample selection approach. The estimate of the coefficient by *FORECAST* is negative and statistically significant. This means that a higher earnings forecast for year $t + 1$ is associated with a lower probability of a going-concern opinion issued in year t, after controlling for other determinants of the going-concern opinion. The results signify that auditors might take managerial forecasts into consideration while deciding on a going-concern opinion.

■

Examples 4.1 and 4.2 illustrate the specifics of constructing an empirical model in accounting, using the methodology of microeconometrics and directing special attention to the defining variables and the availability of data on the companies and the events.

The next sections are entirely devoted to one area of applied accounting research—i.e., the role and the impact of financial disclosures.

Table 4.9 The binomial logit model of management forecasts in the going-concern model

Dependent variable: $GC_t = 1$ if a firm receives a going-concern opinion in year t, $= 0$ otherwise		
Explanatory variable	**Sign prediction**	**Parameter estimate**
Constant		−17.945
$FORECAST_{t+1}$	–	−4.277**
$SIZE_t$	–	−0.138
$ZSCORE_t$	+	0.277*
$LEVERAGE_t$	+	3.140**
$PRLOSS_t$	+	2.583**
$CHGDT_t$	+	9.065***
CFO_t	–	−9.440***
$LIQUIDITY_t$	–	−8.878***
$MKBK_t$	–	−0.510***
$LITIGATE_t$	+	0.478
$REPORTLAG_t$	+	3.002***
$PRIORGC_t$	+	11.613***
$NEWEQUITY_{t+1}$	–	−2.805*
$NEWDEBT_{t+1}$	–	1.736
$DEPENDENCE_t$	–	−5.804***
$BIG\ N_t$	+	−0.897
$INDEXPERT_t$	+	0.721
$TENURE_t$	+	0.257
IMR_t	?	1.612***
Year effect		Yes
Total observations		1054
GC observations		39
Likelihood ratio		223.91
Pseudo R^2		0.705

Outcome equation in the study of auditors' going-concern opinions versus managerial forecasts, 2000–2010

Source: Feng and Li (2014)

*, **, *** denote a p-value of less than 0.10, 0.05, and 0.01, respectively (one-tailed if there is a sign prediction, two-tailed otherwise); all continuous variables are winsorized at the 1st/99th percentile

4.3 Financial Disclosure, Investor Protection, and Disclosure Indices

Accounting Disclosure and Corporate Governance

One major aim of accounting disclosure is to inform current and prospective investors about a company's accounting strategies and the methods used to prepare and present the periodic corporate financial statements. Disclosure on the financials is directly linked to corporate transparency and—more broadly—to corporate governance. Corporate transparency is the accessibility of firm-specific information by the public, specifically by parties outside of the publicly traded firms. Investors and other followers of companies' governance and finances rely on disclosure, transparency, and the quality of both.

As research on disclosure belongs to research into both accounting and corporate governance. Papers on financial disclosure may be found in journals related to accounting and to corporate governance. These include the *International Journal of Disclosure and Governance* published by Palgrave Macmillan or the series "Corporate Governance: Disclosure, Internal Control & Risk Management" published in the Social Studies Research Network (SSRN).

From the standpoint of economics, transparent financial disclosure minimizes agency problems (between owners and managers) by reducing the asymmetry of information between management and shareholders. Conversely, poor financial disclosure may deceive shareholders and lead to unfavorable effects on company valuation and, consequently, on the value of shareholders' wealth. Healy and Palepu (2001) argue that transparency may enhance welfare by improving the efficiency of capital allocation in the economy.

Beekes et al. (2007) show that companies with better corporate governance deliver more disclosure to the market. In their study, the quantity of disclosed information is measured as the logarithm of "document count"—i.e., the number of documents released by the company over the 250 trading days ending 10 trading days after the company's fourth quarter earnings report. This quantity is explained in a model that includes governance level as an explanatory variable.

Financial statement disclosures are the most important for market participants. However, the increasing prominence of corporate governance in the perception of companies by all parties—company insiders and outsiders—calls for the broadest possible disclosure, not just financial. Such expanded understanding of corporate disclosure is needed as much in corporate governance as in accounting. Corporate governance issues and financial microeconometrics are presented in Chap. 5.

Investor Protection, Disclosure, and Legal Systems

The concept of investor protection encompasses all activities aimed at observing, guarding, and enforcing the rights and claims of investors. These include investors' accessibility to legal counsel and legal proceedings. Accordingly, the exact meaning of investor protection in a specific country depends on the extent to which local laws protect the rights of investors as well as on the strength of the institutions implementing those laws. The matter of investor protection is decisive for the quality of corporate governance in companies operating within a country. A healthy economic system requires equal treatment and equal availability of information for all groups of investors. Individual investors in companies are inherently weaker and less informed than institutional investors such as banks. For example, banks providing various financial products to companies are potentially more advantaged than other investors in terms of information, professional expertise, and experience.

What is the connection between investor protection and the level of disclosure? Rationally, it seems to be simple: the greater the disclosure, the better the protection for investors. On the other hand, disclosure may be unfavorable for the company—including its investors. Francis et al. (2001) indicate the following information order: (1) investor protection laws → (2) development of financial markets → (3) role of accounting and auditing in corporate governance → (4) observed characteristics of country-specific financial statements. Thus, financial statements may reflect the requirements of investor protection, but it is not clear that this is always the case.

Accordingly, it is always worth determining whether the degree of financial disclosure is adequate for investor protection or, perhaps, insufficient or excessive. If such findings are to be expressed quantitatively, they should be based on some reasonable measure of the level of disclosure.

Primary research in the area of investor protection has focused on showing the association between the level of investor protection and the type and efficiency of laws in various countries. The evidence reveals that there is a significant distinction between *civil law* countries and *common law* countries.

In civil law (*code law*) countries, the legal system is founded on codes—some dating back to the Roman Empire. Judgments are based only on these codes or on relevant previous interpretations of them. This system is representative of countries in continental Europe including France, Germany, and Spain. On the other hand, common law countries have legal systems based on past judicial opinions (legal precedents); laws are enacted over time and may be interpreted via individual court rulings (case law). Major common law countries are the UK, the USA, Canada, Australia, and New Zealand.

In their comprehensive cross-country research, La Porta et al. (1998, 1999a, b) have shown that the countries having stronger investor protection are countries of common law origin. Civil law countries generally have weaker investor protection, resulting in less developed financial markets and in less timely and less transparent accounting. The systems in civil law countries generate less demand for accounting and auditing to be mechanisms for introducing the changes necessary for better investor protection.

This observation has been evidenced in many subsequent studies. Beekes et al. (2016) in their study of 5000 firms from 23 countries find greater disclosure in those founded on common law compared to civil law countries. Nevertheless, they also present evidence that better governed firms—in whichever system—make more frequent disclosures to the market.

Using a simple theoretical structure, Shleifer and Wolfenzon (2000) proved that countries with better legal protection for outside investors—compared to those with less protection—have:

– Higher stock market capitalization (value)
– More companies listed on stock exchanges
– Larger listed companies (in terms of sales and/or assets)
– Higher valuation of listed firms in relation to their assets
– Higher dividend payouts

– Less concentration of ownership and control
– Less private benefits of control
– Higher correlation between investment opportunities and actual investments.

In the foreword to the book *Investor Protection and Corporate Governance: Firm-Level Evidence across Latin America* (Chong and López-de-Silanes 2007), Shleifer indicates that investor protection is weaker on undervalued and less developed markets than on mature markets. While the issues of investor protection vary across markets, the significant issue in Latin America is the problem of the hidden expropriation of investor assets. This is signaled by a concentration of ownership, low dividend payouts, and a considerable discrepancy between cash flow ownership by dominant shareholders and their share of votes at the annual general meeting. The author believes that the solution to these problems lies in enforcing legal mechanisms (such as corporate law, bankruptcy law, securities law) including implementation of rules addressing the problem of self-dealing by corporate insiders and on disclosure of information.

Disclosure Indices

The issue of measuring the quantity and quality of disclosure has been studied in the literature for some time. The primary idea is to create an index of disclosure founded on the details of accounting laws in a specific country or based on international accounting principles. Disclosure indices are methodologically close to the corporate governance indices that are presented in the next chapter.

Disclosure indices are constructed for a single country, for group of countries, for specific areas of disclosure, etc. A recent survey of disclosure indices (compliant with IFRS-mandatory disclosures) was presented by Devalle et al. (2016).

An example of a country index is the Brazilian Corporate Disclosure Index (BCDI) (Lopes and de Alencar 2010). The Brazilian index measures disclosure in six areas:

1. General information about the firm, its market, and major events over the last year.
2. Relations to employees and managers regarding compensation and policies.
3. Nonfinancial information about markets, sales, and products.
4. Information about forecasts of sales, cash flow, and earnings.
5. Discussion and analysis of financial data including time series information about performance and explanations of past behavior.
6. Other information.

The score is measured by means of 47 questions with binary answers—1 for answers considered to be good disclosure, and 0 otherwise. The data are derived from the 50 companies whose shares had the highest liquidity on the Saõ Paulo Stock Exchange (BOVESPA) in December 2005.

Other examples are:

- The Chinese index of disclosure and transparency (Cheung et al. 2010), which is based on the OECD corporate governance index that includes disclosure as one of five characteristics of transparency. The Index was constructed for the 100 largest Chinese companies in *Fortune's* ranking in 2004–2007. It is composed of two subindices: a voluntary disclosure index and an involuntary disclosure index. All three indices were applied as explanatory variables in regression models with Tobin's q as the dependent variable.
- The Portuguese index of financial instruments disclosure (Lopes and Rodrigues 2007), which concerns disclosure on financial statements consistent with IAS 32 and IAS 39. The statements for the year 2000 of 55 companies quoted on Euronext Lisbon were examined in terms of disclosure. Eleven primary categories of information on financial statement disclosure covered 54 items in these subcategories:
  - Accounting policies (7 items)
  - Fair value and market value (9)
  - Securitization and repurchase agreements (5)
  - Derivatives—risks (4)
  - Derivatives—hedging (10)
  - Derivatives—fair value (4)
  - Interest rate risk (2)
  - Credit risk (3)
  - Collateral (2)
  - Other (3)

Disclosure regarding an item is counted as 1, nondisclosure equals 0. The total score for each company is the sum of all the disclosures. The items are unweighted, and the total is adjusted for non-applicable items. The authors use the index as the dependent variable in models searching for the determinants of disclosure.

- The Italian disclosure index on intangible assets was proposed by Devalle et al. (2016). The index is based on 141 items suggested by the KPMG disclosure checklist. Each company receives a score that is the weighted average of the individual item scores, with the weight reflecting the relevance of each item. The disclosure index is then the dependent variable in regression with covariates representing possible determinants of compliance with mandatory disclosure.
- In the disclosure "index" proposed by Beuselinck et al. (2008), disclosures are represented by a dummy variable equal to 1 if a company discloses its entire financial report even though it is permissible to disclose only an abridged version of the report; equal to 0 otherwise.
- The international disclosure index CIFAR (the Center for International Financial Analysis and Research) measures the comprehensiveness (intensity of disclosure) of corporate annual reports. CIFAR checks how many of 90 selected items are included in a company's annual report, according to the law. This measure has been used to measure cross-country differences in accounting standards and disclosure intensity (e.g., Hope 2003; Białek-Jaworska 2017).

Disclosure Ratings and Rankings

A company's disclosure ratings are becoming an important element of information about the quality of corporate governance, especially in the USA. Commercial companies creating such ratings include[3] GovernanceMetrics International, Audit Integrity, The Corporate Library, RiskMetrics, among others. The disclosure ratings they assign to companies are sometimes the object of criticism. Daines et al. (2010) have shown that the actual corporate governance ratings for 2005 are below the level produced by Audit Integrity, RiskMetrics, GovernanceMetrics International, and The Corporate Library. Such firms (known as proxy advisory firms) provide voting recommendations to shareholders in general assembly meetings.

Corporate governance rating firms provide indices to evaluate the effectiveness of a firm's governance and claim to be able to predict future performance, risk, and undesirable outcomes such as accounting restatements and shareholder litigation. It appears that these recommendations are not adequately founded on companies' predictions. Daines et al. (2010) examined about 15,000 ratings in 2005–2007 for some 7000 companies. Most ratings did not show significant correlation with the following outcomes: accounting restatements (should signify weak governance), class action lawsuits (also a "bad" outcome), future operating performance (measured by ROA), company value (measured by Tobin's q), and future stock returns (measured by excess stock returns, *alpha*).

Researchers also used the AIMR analyst disclosure ratings database that contains ratings for US companies (e.g., Huang and Zhang 2012). The Association for Investment Management and Research (AIMR) produced analyst ratings using data collected from three disclosure areas: (1) the level of optional disclosure in the official annual report, (2) the level of optional disclosure in quarterly reports and other company reports, and (3) the level of informal contacts with analysts. In 2004, the AIMR was renamed the Chartered Financial Analyst (CFA) and no longer issues disclosure ratings (Scaltrito 2015). Another similar example is the Standard & Poor's Transparency and Disclosure Rating (Enikolopov et al. 2014; Patel and Dallas 2002; Huang and Zhang 2008).

Text Analyses for Disclosure Research

Other disclosure indices can be created directly from financial reports like the one proposed by Grüning (2011), who introduced the AIMD—the artificial intelligence measurement of disclosure. The AIMD derives disclosure from English-language annual reports for 10 different information dimensions. A fundamental term used in

[3]In 2010, Governance Metrics International, Audit Integrity and The Corporate Library merged into GMI Ratings. In 2014, GMI Ratings was acquired by MSCI.

this study is the "N-gram," an ordered sequence of N words. The principal assumption is that the meaning of all N-grams related to a given dimension of disclosure is identical. The dimensions are information about (1) sales markets and customers, (2) employees, (3) information about the corporate environment, (4) finances, (5) corporate governance, (6) R&D, (7) social and environmental responsibility, (8) capital markets, (9) corporate strategy, and (10) the value chain.

Research into the analysis of texts on disclosure is rapidly developing. Guay et al. (2016) have studied the complexity of financial statements and voluntary disclosure—using a readability index (ReadIndex) that is the primary component of the following measurements of readability: Flesch–Kincaid readability, LIX readability, RIX readability, ARI readability, and SMOG readability. These measurements are functions of word complexity and sentence length—higher values correspond to less readable text.

A survey on textual analysis in accounting and finance was presented by Loughran and McDonald (2016). The authors examine the growing literature on applying this methodological approach in accounting and finance. In the conclusion to their deep analysis, the authors point out several "tripwires" associated with textual analysis methodologies. Generally, this type of analysis is less precise than the quantitative methods traditionally used in accounting and finance. Therefore, the research outcomes should be treated with appropriate caution.

Jaeschke et al. (2018) examine management's use of language in the financial disclosures of corrupt companies. The methodology they adopt uses tools of text analysis. Management's language about disclosures can be negative, litigious, complex, and conservative. It is measured by several word lists and file characteristics. For example, negative language is a net measure of negative and positive words divided by the square root of the total number of nonnumerical words used in 10-K filings (annual financial reports). The authors found that the managers of FCPA (Foreign Corrupt Practices Act) violators use—in comparison with non-violators—more negative, less litigious, more complex, and less conservative language when disclosing financial information.

4.4 The Microeconometrics of Disclosure

Research Questions

Accounting disclosure is a popular subject of scientific research in accounting and in corporate finance. It is especially important when financial crises teach us all (governments, investors, clients alike) to demand greater corporate transparency. Company owners believe that they disclose sufficient information duly produced to satisfy the requirements of accounting law and other regulations. On the other hand, regulating bodies guard the public interest including the interest of investors and, therefore, there is always room for new requests for disclosure. It is not clear what constitutes the "optimal" disclosure on a specific market and during a specific period. The observed outcome of this tussle is just one possible solution.

Researchers usually concentrate on single research questions and explore them from a historical perspective. Healy and Palepu (2001) have designed an interesting catalogue of research problems in accounting disclosure, which is presented in Table 4.10.

Table 4.10 illustrates the research range in the context of disclosure. Beattie (2005) presents a similar classification. It is also worth mentioning the survey by Laidroo (2006), who categorizes disclosure in market-based accounting research (MBAR) in Central and Eastern Europe.

Table 4.10 Research questions in a disclosure framework (Healy and Palepu 2001)

1. Regulation of disclosure	– Why is there a need for regulation of disclosure in capital markets? What types of disclosures should be regulated, and which should not be? – How effective are accounting standards in facilitating credible communication between managers and outsider investors? What factors determine their effectiveness? – Which mandated disclosures should be recognized directly in the financial statements and which should be included as supplemental disclosures?
2. Auditors, intermediaries, and disclosure	– How effective are auditors in enhancing the credibility of financial statements? What factors influence auditors' effectiveness? – How effective are financial analysts as information intermediaries? What factors influence their effectiveness? – How does corporate disclosure affect analyst coverage of firms?
3. Managers' disclosure decisions	– What factors affect management's disclosure choices? – What is the relation between disclosure, corporate governance, and management incentives? What role do boards and audit committees play in the disclosure process?
4. Capital market consequences of disclosure	– How do investors respond to corporate disclosures? Are firm disclosures made outside the financial statements credible? – Do investors evaluate disclosures that are included directly in the financial statements differently from those that are included as supplemental disclosures? – What factors influence investors' perception of the quality of capital market disclosures across economies? – How does disclosure affect resource allocation in the economy?

Source: Healy and Palepu (2001)

Studies on the Association Between Disclosure and Investor Protection

The relationship between disclosure and investor protection is being studied extensively. Along with the research discussions to be followed in the journals quoted in Sect. 4.1, it is worth noting the surveys by Leuz and Wysocki (2008, 2016), Roberts et al. (2008), and Healy and Palepu (2001).

In their 98-page paper, Leuz and Wysocki (2016) survey the empirical literature on the economic consequences of disclosure and financial reporting regulation in the USA and internationally.

Their conclusions are as follows:

1. "Evidence on the causal effects of disclosure and financial reporting regulation is often difficult to obtain and still relatively rare"
2. "There is a paucity of evidence on market-wide effects from regulation, especially on externalities"
3. "The empirical literature exhibits a heavy focus on disclosure regulation in the United States; each major U.S. regulatory change has been studied extensively; there is much less evidence for major changes in disclosure and reporting regulation in other countries"
4. "In contrast to the work on disclosure regulation, there is a huge literature on the effects of reporting standards internationally; the worldwide adoption of IFRS is arguably one of the largest regulatory events in accounting history and not surprisingly has spawned a large literature on the economic consequences of financial reporting standards"
5. "To make significant progress with respect to the (causal) estimation of regulatory effects and cost–benefit analysis, researchers likely need help from legislators and regulators; for example, one major issue for empirical studies is that most regulation is required as of a particular date, which makes the analysis susceptible to confounding effects, be they other concurrent institutional changes, economic shocks, or market responses to the events that gave rise to the regulation; to mitigate this issue, new regulation could stipulate that rules be implemented in a staggered fashion, which would greatly facilitate ex post economic analysis." (Leuz and Wysocki 2016, 529–531)

Thus, according to this survey, researchers can rarely find evidence of causal relationships between disclosure and a specific disclosure-related issue/phenomenon (first point above). That is unless regulators provide researchers with some unexpected regulation (final point) that may serve as a natural experiment.

The relationship between disclosure and investor protection that has been studied by researchers also fits with the above description. Disclosure regulations and disclosures as such may not have an unquestionably positive impact on investor protection. Research on the quantitative approach to disclosure and investor protection covers a wide range of issues, including the theoretical considerations on mathematical economics trying to prove that globalization influences investor behavior regarding disclosure (e.g., Stulz 2009) and ranging to time series analyses of returns showing that a new disclosure regime significantly alters the properties of such time series (Goto et al. 2009).

Quantitative research on disclosure and investor protection can be grouped into cross-country comparative analyses, research on a specific type of disclosure, or single country studies. The following subsections present examples of studies from each group.

Comparative Cross-Country Studies

Arping and Sautner (2013) made use of a "natural experiment" created accidentally by the cross-listing of European stock in the USA after the Sarbanes–Oxley Act (SOX). The changes in the disclosure of these companies were compared with similar European firms not listed in the USA. Altogether 2500 companies from the EU-15 were examined. The transparency of a company was measured by the accuracy and the dispersion of revenue forecasts by analysts of the companies. It emerged that European stocks quoted in the USA became more transparent after the introduction of the SOX provisions, and this was especially visible for companies in the financial services and technology sectors. This analysis "suggests that SOX had a positive effect on corporate disclosure quality."

Using a sample of 951 companies from 38 countries, Huang and Zhang (2008) showed that a higher level of disclosure is associated with a diminished possibility of insiders accumulating cash and expropriating minority shareholders' assets. The disclosure variable was represented by Standard & Poor's Transparency and Disclosure Rating. The dependent variable in this research is Tobin's q—i.e., the ratio of the market value of assets to their book value where the assets' market value is calculated as the assets book value minus the book value of capital plus the market value of capital.

Hope (2003) found that the level of disclosure in each country (represented by the CIFAR index) might be determined not only by its legal system but also by the national culture. The sample represented 39 countries. Culture was measured by four specific variables used previously in research on the dimensions of national culture. It has been shown that both legal origin (code law or common law) and the culture of the country have an association with disclosure level. In addition, common law countries have higher disclosure indices than code law countries.

Renders and Gaeremynck (2007) examined how the earlier introduction of the IFRS (International Financial Reporting Standard) by European countries is tied with investor protection. The research outcome shows that the IFRS is more likely adopted in countries with strong laws protecting investors and/or extensive corporate governance recommendations where the loss of private benefits following IFRS adoption is lower. The results confirm that corporate governance recommendations are as effective as tough laws in stimulating IFRS adoption. Therefore, by improving corporate governance codes, countries can reduce the extraction of private benefits by managers and improve the quality of financial information.

Bushman et al. (2004) examined corporate transparency in 45 countries. The variables included six disclosure variables (including the CIFAR value), three variables regarding private information, acquisition, and communication, and one variable showing the extent of information dissemination. Factor analysis of the data revealed two major factors. The first may be regarded as representing financial transparency since it captures the intensity and timeliness of financial disclosures, their interpretation, and dissemination. The second factor represents governance transparency and captures the intensity of governance disclosures used by outside investors to hold management accountable. The governance transparency factor is related to the country's legal/judicial regime—this transparency is higher in countries with a common law legal origin and high judicial efficiency. The financial transparency factor is primarily related to the political regime—this transparency is higher in countries with low state ownership of enterprises and banks, and a low risk of state expropriation of firms' wealth.

Enikolopov et al. (2014) used Standard & Poor's Transparency and Disclosure score to represent firm-level corporate governance disclosures by 842 firms in 38 countries during the 2007–2009 financial crisis. The authors show that "the drop in firm value during the financial crisis of 2007–2009 was significantly more sensitive to firm-level transparency in countries with better investor protection. In other words, country- and firm-level institutions become complements when we consider their effect on the decline in firm value." The study is founded on a new theoretical model proposed by the authors.

Research on a Specific Type of Disclosure

The going-concern uncertainty disclosure is perhaps the most important single disclosure arising as a result of examining a company's financial statements. The topic of relating going-concern opinions to various outcomes and determinants has been presented in previous sections—e.g., Gerakos et al. (2016) in Sect. 3.1 (on going-concern and bankruptcy) and Feng and Li (2014) in Example 4.2 (going-concerns versus earnings forecasts). Examples are presented below of research related to disclosures related to the auditor's opinion, CSR (corporate social responsibility), business ethics, and ecology.

Martin (2000) compared a sample of 61 French and German companies with a sample of 61 US firms (for the period 1987–1991), both samples having been chosen from a set of equities experiencing rather low returns. Going-concern uncertainty rates (in the auditor's report or in the board report) were significantly higher for the US firms than for the French or German firms. The control variables included firm characteristics, which may be tied with the going-concern opinion. The chapter concludes that even though prescriptions for the going-concern uncertainty disclosure in Europe and in the USA are the same, the actual disclosures across countries may not have the same meaning.

Dhalival et al. (2011) examined factors determining voluntary CSR reporting. The CSR reports selected for the sample represent 196 US companies and comprise 679 observations during the 1993–2008 period. The explained variable is a dummy and equals 1 if the firm discloses a stand-alone CSR report. The model employed in the study is the binomial logit, also the binomial logit with 1-period lag—for factoring in possible endogeneity of equity capital. Firms usually disclose CRS because they expect a reduction in the firms' costs of equity capital, so the voluntary disclosure of CSR should be associated with lowering this cost—but, at the time of disclosure, the association between the fact of disclosure and the cost of equity capital should be positive. The lagged model made it possible to show that the previous cost of equity capital is positively related to the probability of disclosing CSR for the company. Firms with a high cost of equity capital tend to release corporate social responsibility reports. Also, reporting firms with relatively good social responsibility performance (higher than average) enjoy a reduction in the cost of equity capital. Furthermore, a firm with exceptional CSR performance attracts dedicated institutional investors and the coverage of analysts.

Adams and Kuasirikun (2000) examined disclosures connected with respect to ethical issues by chemical and pharmaceutical companies headquartered in either the UK or Germany. From the examination of the 1995 annual reports of companies listed in *The Times*, it emerged that German firms were better than the British in terms of reporting ecological and ethical issues. Holland and Foo (2003) analyzed ecological disclosures from annual reports for the year 2000 for 37 companies from the UK or the USA. The British companies usually reported on their environmental protection policy and on the awards received for ecological projects. On the other hand, the US firms reported primarily on environmental expenditures and discussion about the risks associated with environmental protection.

Single Country Studies

Numerous studies deal with single-country disclosure issues. Lopes and de Alencar (2010)—who created the *BCDI*, the Brazilian Corporate Disclosure Index (see Sect. 4.3)—estimated models regressing the *BCDI* against a number of variables. The data comprise a panel of 50 stocks having the highest liquidity on the São Paulo Stock Exchange (BOVESPA) on December 31, 2005. The key result is that there is a significant negative association between disclosure and the cost of equity capital. An increase of one point in the *BCDI* relates to a decrease of 14 basis points in this cost. This relationship is more pronounced for firms that receive less attention from analysts and have dispersed ownership structures: an increase of one point in the *BCDI* is associated with reductions in the cost of capital of 26 basis points (for firms with less coverage) and of 27 (for firms with dispersed ownership).

Ben Ali (2009) concentrated on those elements of corporate governance that influence disclosure quality in the context of ownership concentration in which the main agency conflict is between the controlling and the minority shareholders. For a sample using 2004 data on 86 French firms, the author found a negative association between disclosure quality and family control, double voting shares (often encountered in France), and both concentrated ownership and control. Positive associations were revealed between disclosure quality and the existence of executive stock option plans and the proportion of independent members on the supervisory board.

Beekes et al. (2007) (also referred to in Sect. 4.3), working with a sample of Canadian firms, found that companies with a higher level of governance release more documents to the stock market. The sample comprised 216 firms rated in the November 2004 Board Shareholder Confidence Index—established by the *Clarkson Centre for Business Ethics and Board Effectiveness* at the University of Toronto. The explained variable in one of the models is the natural log of the number of documents released by a company over the 250 trading days ending 10 trading days after issuance of the company's fourth quarter earnings report. The explanatory variables are the BSCI corporate governance rating, the firm size (measured by the log of capitalization), and the dummy variable "good news" equal to 1 if the return on the company's share price outperforms the market over the 250-trading-day period, and 0 otherwise. Other models examined timeliness (i.e., the speed with which the released information is reflected in the equity price). The results indicate that better governed Canadian firms release more documents and that value-relevant information is integrated into share prices more rapidly.

Patel and Dallas (2002) showed that Standard & Poor's Transparency and Disclosure Ratings for US companies are correlated with determinants of expected returns such as market risk, price/book value ratio (*P/BV*), and firm size (expressed by market capitalization). The correlation of disclosure with market risk was found to be negative, also negative with *P/BV* but nearly equal to zero, and highly positive with firm size.

Ahmad-Zaluki and Wan-Hussin (2010) analyzed the relationship between the level of corporate governance and the quality of financial information disclosed for a sample of 235 Malaysian companies that went public (IPO) during the period 1999–2006. The quality of disclosure was measured by the error of revenue forecasts prepared by company management. It was shown that companies with higher percentages of nonexecutive directors on the audit committees and larger audit committee size exhibit greater forecast accuracy (i.e., a higher quality of this disclosure). The results also suggest that effective corporate governance is a credible signal of improving the quality of financial information.

Devalle et al. (2016) examined the 2010 financial statements of 189 Italian-listed companies and their compliance with the mandatory disclosure of intangible assets. The disclosure scores or indexes (dscores) were calculated by examining items relevant to disclosure for each company, in four variants. Dscore indexes were then related to several variables in a linear regression model. It appeared that "the only significant variable for all Dscore indexes is the weight of interests on revenues and this result is a distinctive feature of the Italian market where the role of the banking systems is more important than in other countries."

An example of a study on disclosure for companies listed on the Warsaw Stock Exchange (WSE) is presented in Sect. 4.5. The study introduces the *PCDI*: Polish Corporate Disclosure Index. A recent paper on disclosure for WSE-listed companies by Białek-Jaworska (2017) uses an index based on the CIFAR methodology.

4.5 The Polish Corporate Disclosure Index (*PCDI*) and Investor Protection

Composition of the PCDI

The Polish Corporate Disclosure Index (*PCDI*) is the product of research conducted at SGH Warsaw School of Economics in 2009–2010. This was the first published index of disclosure for Polish companies (Świderska 2010; Gruszczyński 2012b). The *PCDI* represents the disclosure quality of three annual reports: the financial statement, the company report (management report), and the corporate social responsibility (CSR) report. Nine areas of reporting are considered with several disclosure issues examined in each, altogether 28 types (see Table 4.11). For each company disclosure quality was composed from the answers to 172 questions (items).

The Polish corporate disclosure index (*PCDI*) is calculated as the weighted average of all 172 disclosure items examined for each company. The weights were

Table 4.11 Disclosure types examined for the *PCDI*

Area of reporting	The disclosures concerned the following types are examined in each area
I. Financial statement (weight of 66%) 1. Nonmaterial assets 2. Financial risk 3. Fair value 4. Accounting for derivatives 5. Leasing 6. Segments of activity 7. Reserves II. Management report (weight of 24%) III. Corporate social responsibility (CSR) report (weight of 10%)	I. Financial statement disclosure (in 7 areas) – Accounting policy – Additional information and disclosure: Mandatory – Additional information and disclosure: Voluntary II. Management report disclosures – Financial and nonfinancial data – Prospective information – Data on management and shareholders – Information about nonmaterial assets III. CSR report disclosures – Communication – Credibility and reliability – Completeness

Source: Gruszczyński (2012a)

generated by a group of experts composed of 12 accounting specialists, including five certified accountants. The following weights are applied:

1. Weights s_f, s_m, s_c for three reports (financial statement, management report, and CSR report): $s_f = 66\%$, $s_m = 24\%$, $s_c = 10\%$
2. Weights w_i for nine areas of reporting (including seven areas in the financial statement); weights indicate how important the disclosures are in this area: $w_i = 1$ ("not important") to $w_i = 5$ ("very important")
3. Weights g_{ij} for each type of disclosure in each of nine areas: for each area the sum of g_{ij} is 1 ($\sum g_{ij} = 1$); there are 28 weights g_{ij}
4. Each of 172 questions (items) is assigned a rank from 0 to 4 (from 0 = "no disclosure," 1 = "poor disclosure" to 4 = "very good disclosure")
5. Average rank for *ij*th type of disclosure is r_{ij}; each company is represented by 28 values of r_{ij}.

The formula for *PCDI* is

$$PCDI = s_f \sum_{i=1}^{7} w_i \sum_{j=1}^{3} g_{ij} r_{ij} + s_m w_8 \sum_{j=1}^{4} g_{ij} r_{ij} + s_c w_9 \sum_{j=1}^{3} g_{ij} r_{ij} \tag{4.1}$$

After applying the appropriate weights *s*, *w*, and *g*, the interval of possible values of *PCDI* can be obtained. This is the range from 0 to 74.13. "The best" disclosure would be associated with a *PCDI* value of 74.13 when all $r_{ij} = 4$. Lower values of *PCDI* are "good disclosure" for *PCDI* = 55.60 (i.e., all $r_{ij} = 3$), "average disclosure" for *PCDI* = 37.06 (i.e., all $r_{ij} = 2$), "poor disclosure" for *PCDI* = 18.53 (i.e., all $r_{ij} = 1$), and "no disclosure" for *PCDI* = 0 (i.e., all $r_{ij} = 0$). Obviously, the range from 0 to 74.13 might be rescaled to a more suitable interval (e.g., from 0 to 100).

PCDI for Companies Listed on the Warsaw Stock Exchange

The disclosure items were examined carefully for 48 companies listed on the WSE for the years 2005–2007. The companies selected for the sample were intentionally chosen from the banking, chemicals, media, food processing, and telecom sectors.

It was found that the minimum *PCDI* obtained for a company is 14.63 and the maximum is 58.53. This may indicate that the *PCDI* measure discriminates companies adequately—from "poor disclosure" to "very good disclosure." The average *PCDI* for the companies chosen for the sample equals 49.24 for 2005, 50.69 for 2006, and 58.63 for 2007. This, in turn, may mean that Polish-listed companies significantly improved the degree of disclosure in their annual reports over the 2005–2007 period.

PCDI and Investor Protection

In order to investigate the relationship between the *PCDI* and investor protection, regression models with the *PCDI* as the only explanatory variable and the investor protection variables as dependent variables were examined. Those proxy investor protection variables are stock price volatility, audit quality, and the Pentor index. In detail they are defined as follows:

- *zmiennosc* = The coefficient of variation of all stock price quotations in a given year.
- *kat_aud* = The category of the company auditor: 1 = Big 4, 2 = major regional and domestic auditors, 3 = other auditors.
- *pentor* = Value of the Pentor index (published by the journal "Puls Biznesu") for each company; the index is the outcome of a questionnaire completed by stock exchange analysts, advisors, and brokers as to how the company is perceived by the market, what are the strengths of its managers, what is the quality of investor relations, what are the company's prospects, etc.

We expect that these variables are related to investor protection in the following ways:

- *zmiennosc*: The higher the volatility of equity prices, the lower the investor protection.
- *kat_aud*: The better the auditor (in terms of its international and national reputation), the better the investor protection.
- *pentor*: The higher the value of the index representing market sentiment toward the company, the better the investor protection.

The quantitative variables *zmiennosc* and *pentor* served as dependent variables in the linear regression models with *PCDI* as the only regressor. The qualitative *kat_aud* variable was used as the explained variable in the ordered logistic regression, again with the *PCDI* as the only regressor. The estimation results for the appropriate models with *PCDI* as the regressor variable are shown in Table 4.12.

The principal outcome of this research attempt suggests that *PCDI* is significantly linked with the variables representing investor protection. A higher value of *PCDI* is associated with a lower variability of stock prices as well as with a better category of auditor. The better the management, the better the company's market perception—i.e., better corporate governance is associated with higher *PCDI*.

Disclosure Types and Market Sentiment[4]

This study next concentrates on revealing the groups of disclosures that have association with market sentiment expressed by experts and investors and measured

[4]Marcin Owczarczuk is the coauthor of this subsection.

Table 4.12 Disclosures on 48 WSE-listed companies in Poland (2005–2007) related to variables representing investor protection

Explanatory variable: PCDI (Polish Corporate Disclosure Index)				
		Sign of parameter estimate		
Model	Dependent variable	2005	2006	2007
Model 1 Simple linear regression	*zmiennosc* = stock price volatility	Minus	Minus***	Minus***
Model 2 Trinomial ordered logit	*kat_aud* = audit quality	Minus***	Minus***	Minus***
Model 3 Simple linear regression	*pentor* = Pentor index	Plus***	Plus***	Plus***

Source: Gruszczyński (2012a)
$n = 48$ companies listed on the Warsaw stock exchange in 2005–2007; *** $p < 0.01$, ** $p < 0.05$

Table 4.13 Disclosure types in reports of Polish-listed companies

1. Nonmaterial assets	5. Leasing
1.1 Accounting policy	5.1 Accounting policy
1.2 Mandatory disclosure	5.2 Mandatory disclosure
1.3 Voluntary disclosure	5.3 Voluntary disclosure
2. Financial risk	6. Segments of activity
2.1 Accounting policy	6.1 Accounting policy
2.2 Mandatory disclosure	6.2 Mandatory disclosure
2.3 Voluntary disclosure	6.3 Voluntary disclosure
3. Fair value	7. Reserves
3.1 Accounting policy	7.1 Accounting policy
3.2 Mandatory disclosure	7.2 Mandatory disclosure
3.3 Voluntary disclosure	7.3 Voluntary disclosure
4. Accounting for derivatives	8. Management report
4.1 Accounting policy	8.1 Financial and nonfinancial data
4.2 Mandatory disclosure	8.2 Prospective information
4.3 Voluntary disclosure	8.3 Data on management and shareholders
	8.4 Information about nonmaterial assets
	9. CSR report
	9.1 Communication
	9.2 Credibility and reliability
	9.3 Completeness

Source: Gruszczyński (2012a)

by the *Pentor* index. The index is the dependent variable and all 28 types of disclosure shown in Table 4.11 are potential covariates. They are now grouped in Table 4.13.

The list contains 28 types of disclosure that represent possible explanatory variables. In this model, there are no weights attached to specific disclosure types. Each variable is represented by r_{ij}—i.e., the score obtained for all 48 companies for this (ijth) type of disclosure.

The objective is to find which of these 28 variables have the most significant association with the *Pentor* variable. As there are only 48 observations (companies) each year, we shall appropriately select explanatory variables since the possible number of them is large. In our model

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \cdots + \beta_k X_{ki} + \varepsilon_i \tag{4.2}$$

y_i is the value of the *Pentor* variable for the *i*th company ($I = 1, 2, \ldots, n$; $n = 48$), X_{ji} is the score of the *j*th disclosure for the *i*th company ($j = 1, 2, \ldots, k$), ε_i is the random error. Initially $k = 28$.

In order to reduce the number of explanatory variables, we propose the use of *lasso* (*least absolute shrinkage and selection operator*). Lasso is based on minimization of squared errors with an additional penalty term on the parameters

$$min_{\beta_0,\ldots,\beta_k}\left(\sum_{i=1}^{n}\left[y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 X_{1i} + \widehat{\beta}_2 X_{2i} + \cdots + \widehat{\beta}_k X_{ki}\right]^2 + \lambda \sum_{j=1}^{k}\left|\widehat{\beta}_j\right|\right) \tag{4.3}$$

To some extent, we want to achieve the best fit—i.e. smallest sum of squared errors. By increasing the value of λ, we increase the influence of the penalty term. This causes the parameters to become smaller and smaller (in absolute values). In order not to disturb the sum of squared errors, the parameters of the insignificant variables are reduced. If the value of λ is sufficiently large, the parameters of the insignificant variables are equal to zero. Of course, if the value of λ is too high, the significant variables also have zero coefficients. In order to determine the optimal value of λ, cross-validation is used. For a given λ, the model is estimated on a subsample and its performance (i.e., the mean squared error of prediction) is measured on the remaining subsample, which was not part of the estimation phase. In order to achieve some stability of results, this procedure is repeated for different partitions into training and testing subsamples. The sample is divided into q blocks, and the model is estimated using observations from $q - 1$ blocks and its performance is measured on the q-th block. Then the roles of the blocks are rotated, so we achieve q values of errors (each value corresponds to the situation in which a particular block acts as a test part). Finally, the results are averaged.

The whole procedure is repeated for different values of λ and we finally choose a value λ and a corresponding model that has the smallest value of the mean squared error. So, we choose a model that had the best prediction abilities. The variables included in this optimal model are those that best explain the dependent variable.

The *lasso* technique was introduced by Tibshirani (1996). Since then, it has been in widespread use in advanced data analysis applications, including corporate finance research (e.g., Pereira et al. 2016; Lin et al. 2014).

The *lasso* procedure was applied to our dataset. In order to have an idea to what extent the results are reproducible, we repeated the experiment with a few variants: for the whole sample (i.e., years 2005, 2006, and 2007), each year separately, and by

Table 4.14 Disclosure types in reports of Polish-listed companies—mostly recognized by analysts and investors

Dependent variable: *Pentor* = the value of the market sentiment index in 2005–2007					
2005	**2006**	**2007**	**2005/6**	**2006/7**	**2005/6/7**
Variables selected by lasso as explanatory disclosure types					
1.1	1.1	2.2	1.1	1.1	1.1
1.3	1.3	6.2	1.3	1.3	1.3
2.2	2.2	9.3	2.2	2.2	2.2
4.3	9.1			4.1	4.1
6.1	9.3			6.2	6.1
7.1				9.1	6.2
7.3				9.3	9.3
8.4					

Disclosure classification numbers from Table 4.13
$n = 48$ companies listed on the Warsaw Stock Exchange
Source: Gruszczyński (2012a)

joining neighboring years (i.e., 2005 with 2006; 2006 with 2007). The results are presented in Table 4.14.

All selected variables have positive parameter estimates, except variables 4.3 and 7.3 in the 2005 model that have negative parameter estimates.

The results presented in Table 4.14 suggest that the following disclosures are most welcomed by investors and analysts:

- 2.2: Financial risk management—Mandatory disclosure (appearing 6 times as an explanatory variable in the models)
- 1.1: Nonmaterial assets—Accounting policy (5 times)
- 1.3: Nonmaterial assets—Voluntary disclosure (5 times)
- 9.3: CSR report—Completeness (4 times)
- 6.2: Segments of activity—Mandatory disclosure (3 times)
- 4.1: Accounting for derivatives—Accounting policy (2 times)
- 6.1: Segments of activity—Accounting policy (2 times)
- 9.1: CSR report—Communication (2 times)

It may be stated that the foregoing disclosures drive the market's perceptions of companies as represented by analysts and investors. Our study indicates that disclosures regarding "financial risk management" (mandatory) and "nonmaterial assets" (accounting policy and nonmandatory disclosure) as well as the completeness of the "CSR report" are most important in the market's perception of companies. However, this study does not indicate the market importance of disclosures in the areas of fair value, leasing, reserves, and the management report.

The study shows how to determine the types of disclosure that are most relevant to market perception. This attempt to relate "soft" information about market sentiment to "soft" characteristics of the company's reporting of information seems to be valuable as a technique to analyze the quality of company reports vis-à-vis the company's quality perceived by the market.

This chapter, entirely devoted to research questions in accounting, has concentrated on the methodology of microeconometrics. We have attempted to discuss the entire spectrum of topics and methods in accounting research. Financial microeconometrics methodology appears to be appropriate in most cases where researchers use large datasets of companies, of financial statements, of business events, etc.

Corporate disclosure—including financial disclosure—is a component of corporate governance, an area of management rapidly gaining importance, including as a subject of research. Topics of corporate governance research are presented in the next chapter.

References

Adams CA, Kuasirikun N (2000) A comparative analysis of corporate reporting on ethical issues by UK and German chemical and pharmaceutical companies. Eur Account Rev 9(1):53–79

Ahmad-Zaluki NAA, Wan-Hussin WN (2010) Corporate governance and earnings forecasts accuracy. Asian Rev Account 18:50–67

Amer T, Hackenbrack K, Nelson M (1994) Between-auditor differences in the interpretation of probabilities phrases. Auditing J Pract Theory 13:126–136

Arping S, Sautner Z (2013) Did SOX section 404 make firms less opaque? Evidence from cross-listed firms. Contemp Account Res 30(2013):1133–1165

Beattie V (2005) Moving the financial accounting research front forward: the UK contribution. Br Account Rev 37(1):85–114

Beekes W, Brown PR, Chin G (2007) Do better-governed firms make more informative disclosure? Canadian Evidence. Available at SSRN: https://ssrn.com/abstract=881062

Beekes W, Brown PR, Zhan W, Zhang Q (2016) Corporate governance, companies disclosure practices and market transparency: a cross country study. J Bus Financ Acc 43(3–4):263–297

Ben Ali C (2009) Disclosure quality and corporate governance in a context of minority expropriation. Available at SSRN: https://ssrn.com/abstract=1406149

Benson KL, Clarkson P, Smith T, Tutticci I (2015) A review of accounting research in the Asia Pacific region. Aust J Manag 40(1): 36–88. Available at SSRN: https://ssrn.com/abstract=2575706

Beuselinck C, Deloof M, Manigart C (2008) Private equity investments and disclosure policy. Eur Account Rev 17(4):607–639

Białek-Jaworska A (2017) Determinants of information disclosure by companies listed on the Warsaw Stock Exchange in Poland. Glob Bus Econ Rev 19(2):157–175

Bushman RM, Piotroski JD, Smith AJ (2004) What determines corporate transparency. J Account Res 42(2):207–252

Chapman CS, Hopwood AG, Shields MD (eds) (2006a) Handbook of management accounting research, vol 1. Elsevier, Oxford, pp 1–478

Chapman CS, Hopwood AG, Shields MD (eds) (2006b) Handbook of management accounting research, vol 2. Elsevier, Oxford, pp 479–1207

Chapman CS, Hopwood AG, Shields MD (eds) (2009) Handbook of management accounting research, vol 3. Elsevier, Oxford, pp 1207–1410

Chenhall RH, Smith D (2011) A review of Australian management accounting research: 1980–2009. Account Finance 51:173–206

Cheung SY-L, Jiang P, Tan W (2010) A transparency disclosure index measuring disclosures: Chinese listed companies. J Account Public Policy 29:259–280

Chong A, López-de-Silanes F (eds) (2007) Investor protection and corporate governance: firm-level evidence across Latin America, Latin American development forum. Inter-American Development Bank, Washington, DC

Coyne JG, Summers SL, Williams B, Wood DA (2010) Accounting program research rankings by topical area and methodology. Issues Account Educ 25(4): 631–654. Available at SSRN: http://ssrn.com/abstract=1337755

Cram DP, Karan V, Stuart I (2007) Three threats to validity of choice-based and matched sample studies in accounting research. Available at SSRN: http://ssrn.com/abstract=955031

Daines R, Gow ID, Larcker DF (2010) Rating the ratings: how good are commercial governance ratings? J Financ Econ 98:439–461

Devalle A, Rizzato F, Busso D (2016) Disclosure indexes and compliance with mandatory disclosure—the case of intangible assets in the Italian market. Adv Account 35(2016):8–25

Dhalival DS, Li OZ, Tsang A, Yang YG (2011) Voluntary nonfinancial disclosure and the cost of equity capital: the initiation of corporate social responsibility reporting. Account Rev 86 (1):59–100

Doupnik T, Richter M (2003) Interpretation of uncertain expressions: a cross national study. Acc Organ Soc 28(1):15–35

Du N, Stevens K, Ahern J, Shigaev A (2016) Cross-cultural differences in interpreting IAS 37 probability phrases. Int J Financ Res 7(1):1

Dunbar AE, Weber DP (2014) What influences accounting research? A citations-based analysis. Issues Account Educ 29(1):1–60

Enikolopov R, Petrova M, Stepanov S (2014) Firm value in crisis: effects of firm-level transparency and country-level institutions. J Bank Financ 46:72–84

Feng M, Li C (2014) Are auditors professionally skeptical? Evidence from auditors' going-concern opinions and management earnings forecasts. J Account Res 52(5):1061–1085

Francis JR, Khurana IK, Pereira R (2001) Investor protection laws, accounting and auditing around the world. Available at SSRN: http://ssrn.com/abstract=287652

Fülbier RU, Sellhorn T (2008) Approaches to accounting research – evidence from EAA Annual Congresses. Available at SSRN: http://ssrn.com/abstract=985119

Ge W, Whitmore GA (2010) Binary response and logistic regression in recent accounting research publications: a methodological note. Rev Quant Finan Acc 34:81–93

Gerakos JJ, Hahn R, Kovrijnykh A, Zhou F (2016) Prediction versus inducement and the informational efficiency of going-concern opinions, Chicago Booth Research Paper No. 16-01 (Jan.)

Goto S, Watanabe M, Xu Y (2009) Strategic disclosure and stock returns: theory and evidence from US cross-listing. Rev Financ Stud 22:1585–1620

Grüning M (2011) Artificial intelligence measurement of disclosure (AIMD). Eur Account Rev 20 (3):485–519

Gruszczyński M (2009) Quantitative methods in accounting research. Metody ilościowe w badaniach ekonomicznych/Quant Methods Econ X(1): 76–87. also available as Gruszczyński M (2009) Quantitative methods in accounting research, Working Paper No. 6-09, Department of Applied Econometrics, SGH Warsaw School of Economics

Gruszczyński M (2012a) Empiryczne finanse przedsiębiorstw. Mikroekonometria finansowa [Empirical corporate finance. Financial microeconometrics]. Difin, Warszawa

Gruszczyński M (2012b) Investor protection and disclosure: quantitative evidence. FindEcon. Forecasting financial markets and economic decision-making, no. 10: 19–36. Available at SSRN: https://ssrn.com/abstract=1806003; also available as Gruszczyński M (2010) Investor protection and disclosure. Quantitative evidence. Working Paper No. 06-10, Department of Applied Econometrics, SGH Warsaw School of Economics

Guay W, Samuels D, Taylor D (2016) Guiding through the fog: financial statement complexity and voluntary disclosure. J Account Econ 62(2016):234–269

Healy PM, Palepu K (2001) Information asymmetry, corporate disclosure and the capital markets: a review of the empirical disclosure literature. J Account Econ 31(1–3):405–439

Heckman JJ (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Ann Econ Soc Meas 5:475–492

Holland L, Foo YB (2003) Differences in environmental reporting practices in the UK and the US: the legal and regulatory context. Br Account Rev 35:1–18

Hope O-K (2003) Firm-level disclosures and the relative roles of culture and legal origin. J Int Financ Manag Acc 14(3):218–248

Hoque Z (ed) (2018) Methodological issues in accounting research, 2nd edn. Spiramus Press, London

Huang P, Zhang Y (2008) Does enhanced disclosure really reduce agency costs? Evidence from the value of corporate cash holdings and dividends. China international conference in finance. Available at: http://www.ccfr.org.cn/cicf2008/download.php?paper=20080114101835.PDF (Access in 2012)

Huang P, Zhang Y (2012) Does enhanced disclosure really reduce agency costs? Evidence from the diversion of corporate resources. Account Rev 87(1):199–229

Huerta E, Petrides Y, Braun GP (2016) Interpretation of probability expressions in accounting: the effects of frame switching. Journal of International Accounting, Auditing and Taxation 27 (2016):1–12

Jaeschke R, Lopatta K, Yi C (2018) Managers' use of language in corrupt firms' financial disclosures: evidence from FCPA violators. Scand J Manag 34(2018):170–192

La Porta R, Lopez-de-Silanes F, Shleifer A, Vishny RW (1998) Law and finance. J Polit Econ 106:1113–1155

La Porta R, Lopez-de-Silanes F, Shleifer A, Vishny RW (1999a) Investor protection and corporate governance. Available at SSRN: http://ssrn.com/abstract=183908

La Porta R, Lopez-de-Silanes F, Shleifer A, Vishny RW (1999b) Investor protection: origins, consequences, and reform. NBER Working Paper No. 7428

Laidroo L (2006) Review of empirical studies on disclosure and MBAR in European countries. Tallinn University of Technology. http://www.emselts.ee/konverentsid/EMS2006/2_Rahandus_ja_pangandus/Laivi_Laidroo.pdf

Laswad F, Mak Y (1997) Interpretations of probability expressions by New Zealand standard setters. Account Horiz 11(4):16–23

Lee CF (ed.) (2004–2008) Advances in quantitative analysis of finance and accounting (6 volumes). World Scientific Publishing: River Edge, NJ

Lennox CS, Francis JR, Wang Z (2012) Selection models in accounting research. Account Rev 87:589–616

Leuz C, Wysocki PD (2008), Economic consequences of financial reporting and disclosure regulation: a review and suggestions for future research. Available at SSRN: http://ssrn.com/abstract=1105398

Leuz C, Wysocki PD (2016) The economics of disclosure and financial reporting regulation: evidence and suggestions for future research. J Account Res 54(2):525–622

Li X, Sun L, Ettredge M (2017) Auditor selection following auditor turnover: do peers' choices matter? Acc Organ Soc 57:73–87

Lin L, Shuang W, Yifang L, Shouyand W (2014) A new idea of study on the influence factors of companies' debt costs in the big data era. Procedia Comput Sci 31:532–541

Lopes AB, de Alencar RC (2010) Disclosure and cost of equity capital in emerging markets: the Brazilian case. Int J Account 45(4):443–464

Lopes PT, Rodrigues LL (2007) Accounting for financial instruments: an analysis of the determinants of disclosure in the Portuguese stock exchange. Int J Account 42:25–56

Loughran T, McDonald B (2016) Textual analysis in accounting and finance: a survey. J Account Res 54(4):1187–1230

Machado MJ, Ribeiro JL (2016) European accounting review: the profile of accounting research in Europe. WSEAS Trans Bus Econ 13:119–128

Maddala GS (1991) A perspective on the use of limited-dependent variables models in accounting research. Account Rev 66(4):788–807

Martin RD (2000) Going-concern uncertainty disclosures and conditions: a comparison of French, German and US practices. J Int Account Audit Tax 9:137–158

Patel SA, Dallas GS (2002) Transparency and disclosure: overview of methodology and study results – United States, Standard & Poor's. Available at SSRN: http://ssrn.com/abstract=422800

Paterson A, Leung D, Jackson WJ, MacIntosh R, & O'Gorman KD (eds.) (2016) Research methods for accounting and finance, (Global Management Series). Goodfellow

Pereira JM, Basto M, da Silva F (2016) The logistic lasso and ridge regression in predicting corporate failure. Procedia Econ Finan 39(2016):634–641

Reimers J (1992) Additional evidence on the need for disclosure reform. Account Horiz 6(1):36

Renders A, Gaeremynck A (2007) The impact of legal and voluntary investor protection on the early adoption of IFRS. De Economist 155(1):49–72

Roberts C, Weetman P, Gordon P (2008) International corporate reporting: a comparative approach, 4th edn. Prentice Hall, New York

Scaltrito D (2015) Assessing disclosure quality: a methodological issue. J Modern Account Audit 11(9):466–475

Shleifer A., D. Wolfenzon (2000), Investor protection and equity markets. Harvard Institute of Economic Research Paper No. 1906 and NBER Working Paper No. 7974

Silska-Gębka S (2017) Słowne określenia prawdopodobieństwa jako przyczyna różnic w interpretacji Międzynarodowych Standardów Sprawozdawczości Finansowej – ku badaniom naukowym. Zeszyty Teoretyczne Rachunkowości 92(148):131–150

Simon J (2002) Interpretation of probability expressions by financial directors and auditors of UK companies. Eur Account Rev 11(3):601–629

Smith M (2017) Research methods in accounting, 4th edn. Sage, London

Stulz RM (2009) Securities laws, disclosure and national capital markets in the age of financial globalization. J Account Res 47(2):349–390

Świderska GK (ed) (2010) Wpływ zakresu ujawnianych informacji na poprawę ochrony inwestorów oraz pozycję konkurencyjną emitentów papierów wartościowych [The extent of disclosure, protection of investors and competitive position of equity issuers]. SGH Publishing House, Warsaw (Authors: S. Borowski, M. Gruszczyński, M. Kariozen, M. Karwowski, M. Owczarczuk, M. Pielaszek, P. Roszkowska, G.K. Świderska, M. Świderska)

Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc Ser B Methodol 58:267–288

Chapter 5
The Microeconometrics of Corporate Governance

This chapter presents topics regarding corporate governance, an assessment of its level for companies, and quantitative research aimed at disclosing associations between corporate governance and firm performance and other company characteristics. The methodology of corporate governance research is mainly microeconometrics. We present several examples of microeconometric studies devoted to corporate governance issues such as female representation on management boards, CEO change, and earnings management. We also discuss corporate governance ratings and indices along with the debate on their use and misuse.

5.1 Sources of Knowledge and Areas of Corporate Governance

Corporate governance (CG) refers to the practices of companies that lead to an independent and efficient supervising body, transparent and accurate books, strong shareholder rights, and equal treatment of all ownership groups. In theory, the mechanism of CG minimizes agency costs—i.e., reduces the company's loss of market value resulting from a potential conflict between management and owners (Shleifer and Vishny 1997). Corporate governance rules are usually expressed in the form of codes (principles), both domestic and international. An investigation of the quality and/or level of CG involves legal, managerial, accounting, and financial issues. Naturally, the area of corporate governance has also been studied using quantitative approaches, including microeconometrics.

The title of this chapter is exactly the same as the monograph by Bhagat and Jefferis (2002), in which they cover only the aspect of mergers and acquisitions in the context of corporate governance. In this chapter, we concentrate on a variety of facets of CG research in terms of scope and methodology, especially those related to corporate finance, as in special issues on corporate governance in the *Journal of*

M. Gruszczyński, *Financial Microeconometrics*,
https://doi.org/10.1007/978-3-030-34219-7_5

Corporate Finance published in 2014, 2011, 2008, and 2006. Additional detail on the methodology of contemporary CG research can be found in special issues of *Corporate Governance. An International Review*, in particular one published in 2017 (vol. 25, issue 6). The subjects discussed in this chapter are selected from the comprehensive sphere of CG problems, research questions, and practical issues.

Literature on CG

Several publications and published series on contemporary corporate governance research should be mentioned. First are these journals: *Corporate Governance. An International Review*, *Corporate Governance. The International Journal of Business in Society*, *International Journal of Disclosure and Governance*, and *Journal of Business Ethics*. Current research is also published in repositories like SSRN that presents a comprehensive collection of "Corporate Governance Subject Matter eJournals." Series of papers are provided as well by the IRRC Institute,[1] the European Corporate Governance Institute (ECGI), and the Rock Center for Corporate Governance at Stanford University. Quality discourses may also be followed on the Harvard Law School Forum on Corporate Governance and Financial Regulation.

As an example of country-level publications on CG, we highlight the German journals that have been mentioned by Eulerich et al. (2013). These are *Zeitschrift für Betriebswirtschaft*, *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung*, *Die Betriebswirtschaft*, *Betriebswirtschaftliche Forschung und Praxis*, *Die Unternehmung*. The bibliometric analysis performed by the authors shows that these are the journals that most frequently publish articles on CG topics.

Subjects of corporate governance cover all fields that are studied in regard to companies: finance, management, economics, law, etc. The authors of books on CG concentrate primarily on selected issues. We mention here only a few such authors: Larcker and Tayan (2015a), Leblanc (2016), Goergen (2018), Padgett (2011), and Thomsen and Conyon (2012).

Collections of papers on CG are also published in books—e.g., the comprehensive *Oxford Handbook of Corporate Law and Governance* (Gordon and Ringe 2018), which covers issues of CG primarily related to law and the regulation of companies. Another collection of papers is the book *Corporate Governance and Corporate Finance. A European Perspective* (Frederikslust et al. 2008), which not surprisingly focuses on CG in Europe. The series of books published by Emerald and entitled *Developments in Corporate Governance and Responsibility* is also worth noting. The series is composed of 12 volumes (the last published in 2017) with papers on a variety of CG topics.

[1]Investor Responsibility Research Center Institute; since 2018 the John L. Weinberg Center for Corporate Governance at the University of Delaware.

Research on corporate governance is particularly demanding since it concerns such diverse disciplines as law, economics, accounting, finance, and management. Filatotchev and Wright (2017) state: "Corporate governance has become a truly interdisciplinary area compared to the dominance of economics and finance in the field 25 years ago." Obviously, CG is very much linked to both commercial and securities law as well as to the legal aspects of accounting. The practical exercise of the provisions of CG is embedded more in law and accounting than in management and finance.

The Areas of CG

While different authors have classified the key topics of corporate governance in various manners, we present here a selection of lists of issues representative of the CG landscape.

Klausner (2018) discusses the empirical literature on corporate law and governance in the USA in four areas:

1. State competition to produce corporate law
2. Independent boards
3. Takeover defenses
4. Corporate governance indices

The author points out that the empirical analysis of CG problems, especially when performed by economists and finance researchers, "has taken the field beyond the exchange of theoretical assertions and ideological pronouncements that often characterize legal scholarship, and it has the potential to take us farther." This view may be representative of legal scholars, who highlight misunderstandings in the institutional and legal contexts of research performed by economists.

In a special issue of the *Journal of Corporate Finance,* Gillan (2006) discusses CG mechanisms as those internal to firms and those external to firms. Internal governance is performed by management acting as the shareholders' agents, deciding on the investment of assets, and financing those investments, while the board of directors (the supervisory board) advises and monitors management, and holds responsibility for hiring, firing, and compensating senior management. The author illustrates this with a graph ("balance sheet model") adopted from slides accompanying an earlier version of Ross et al. (2015), in which the left-hand side represents internal CG and the right-hand side represents external CG. It is depicted here in Fig. 5.1.

External governance (the right-hand side of Fig. 5.1) represents elements "arising from firm's need to raise capital . . . and highlights that in the publicly traded firm, a separation exists between capital providers and those who manage the capital." The author distinguishes the following categories of internal governance:

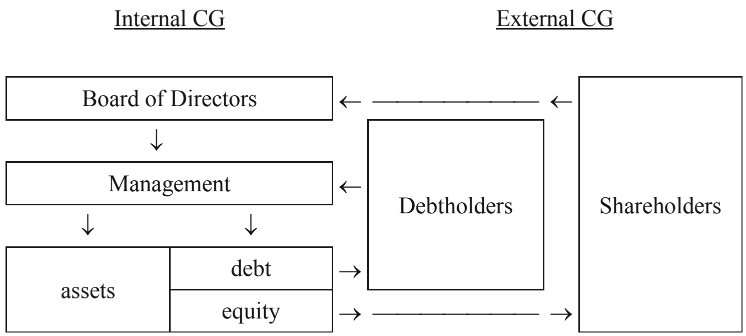


Fig. 5.1 Internal and external corporate governance. Source: Gillan (2006)

1. The board of directors (role, structure, and incentives)
2. Managerial incentives
3. Capital structure
4. Bylaws and charter provisions (or anti-takeover measures)
5. Internal control systems

Similarly, external governance includes:

1. Law and regulation—Federal law, self-regulatory organizations, and state law
2. Markets 1—Capital markets, market for corporate control, labor markets, product markets
3. Markets 2—Providers of capital market information (such as that provided by credit, equity, and governance analysts)
4. Markets 3—Accounting, financial, and legal services from parties external to the firm (auditors, directors' and officers' liability insurance, investment banking advice)
5. Private sources of external oversight—Media and external lawsuits

Corporate Governance and Corporate Finance. A European Perspective (Frederikslust et al. 2008) contains reprints of papers published in 1993–2004 in 28 chapters that are grouped into six parts:

1. Alternative perspectives on corporate governance systems
2. Equity ownership structure and control
3. Corporate governance, underperformance, and management turnover
4. Directors' remuneration
5. Governance, performance, and financial strategy
6. Takeover as a disciplinary mechanism

Our praise for the editors arises from their recognition of specific groups of topics that were dominant in European CG literature in 1993–2004. In the introduction to this volume, the editors state: "Continental Europe is characterized by a governance system in which large block shareholders play a key role, whereas the US system is driven by the needs of dispersed shareholders articulated through the stock market." The editors also state that CG systems vary across countries but consist broadly of internal and external control and monitoring mechanisms.

The areas of CG presented in the review paper by Brown et al. (2011) are different. Their survey is directed to the difficulties in CG research methodology. The authors believe that there is no theory uniting all elements of corporate governance; therefore, researchers concentrate on rather narrow questions and topics. They call for an interdisciplinary approach to developing a better theory. The survey begins with questions concerning measurement of the level of CG, including problems like the "stickiness" of data on CG (in specialized databases on companies, most CG elements do not change over time), the number of CG dimensions, the construction of CG indices, and the question of endogeneity in CG research. Next, the authors emphasize the international scope of corporate governance codes, and then develop ideas on the internal and external elements of CG.

The most interesting part of the survey by Brown et al. (2011) is devoted to the "finance outcomes" and the "accounting outcomes" of CG. This closely relates to the research questions presented in the next section. The authors distinguish the following subjects:

1. Finance outcomes
   (a) Firm performance
      - Board structure
      - Ownership structure
      - Outside blockholders
      - Market for corporate control
      - Legal protection
   (b) Sensitivity of CEO turnover to performance
   (c) CG and equity
      - The cost of equity
      - The payout policy
   (d) CG and debt
      - Leverage
      - The cost of debt
2. Accounting outcomes
   (e) CG and firm disclosure environments
      - Disclosures
      - Analyst forecasts
      - Regulation and compliance
   (f) CG and accounting quality
      - Conservatism and earnings timeliness

- Earnings informativeness
- Earnings management
- Restatements and fraud

Numerous examples of research on these subjects make this a valuable survey providing a picture of CG research with a financial-accounting focus.

CG in Common Law and Civil Law Countries

In Sect. 4.3, we described the significant distinctions between civil law (code law) countries and common law countries in terms of the level of investor protection and the type and the efficacy of law. This same distinction applies to corporate governance issues generally, as has been documented by many researchers beginning with the seminal studies by La Porta et al. (1998, 1999a, b). Other studies that are relevant to cite also for general corporate governance issues—not just issues of financial disclosure—are presented in Sect. 4.3 (such as Beekes et al. 2016).

This difference, however, seems to be narrowing over time. Using panel data extending back to 1970, Armour et al. (2009) showed that civil law countries have increased protection for minority investors to a greater extent than has been done by common law countries. The authors attribute this to the introduction of codes of good governance in most countries.

Codes of Good Governance

On the webpage of the ECGI (the European Corporate Governance Institute) can be found the full texts of corporate governance codes, principles of corporate governance, and corporate governance reforms for countries worldwide (https://ecgi.global/content/codes). The documents are adopted typically within a country and enumerate practices of corporate governance that are to be followed by all parties involved: companies, owners, investors, institutions, etc. Usually the codes are applicable to public companies.

Corporate governance codes are sets of standards and good practices regulating the issues of management and supervision of companies, transparency and disclosure, and relations with shareholders. The codes have now been extended beyond "behavior and structure of the board of directors" into governance characteristics and the behavior of institutional investors and intermediaries, in the form of stewardship codes (Haskovec 2012).

The first influential international code was set forth by the OECD (the Organisation for Economic Co-operation and Development) in 1999. The current version of *OECD Principles of Corporate Governance* was endorsed by the G20 leaders in 2015 (OECD 2015).

At the level of institutions of the European Union, corporate governance issues have been discussed since the year 2000. Generally, the EU has adopted principles-based comply-or-explain regime for member state-based corporate governance codes. The EU's approach focusses on creating a system in which effective and accountable companies report to responsible shareholders. This is consistent with the evidence that shareholder intervention improves economic performance (Dallas and Pitt-Watson 2016).

The newest EU Directive on CG named *Shareholder Rights Directive II* (SRD II) was passed in 2018 and expected to become effective in mid-2019. It replaces SRD I from 2007 and incorporates lessons learned from the 2008 financial crisis. The central focus of SRD II is encouraging long-term shareholder engagement and increasing transparency.

Example 5.1 Best Practice for Warsaw Stock Exchange Companies and Tobin's q

On the webpage of the ECGI, we find seven codes of corporate governance for companies in Poland: the first from 2002 and the latest from 2015. Garstka (2009) examines the 2008 code and relates its provisions to the fundamentals of companies listed on the Warsaw Stock Exchange (WSE).

The author constructs a CG score (*ICG*) that mimics companies' statements on observing specific provisions of the "Code of Best Practice for WSE-Listed Companies." Each of the 38 provisions was scored 1 if it was observed by the company and 0 if not. The *ICG* was rescaled to the interval < 0,100>. The average *ICG* is shown to be approximately 90 points. In order to examine the association between the companies' fundamentals and the *ICG,* the author attempted the regression models with Tobin's q as the dependent variable. The *ICG* variable is explanatory along with other characteristics of the companies.

Table 5.1 presents the OLS estimation results from one of the models attempted by the author. Financial data were collected for 284 WSE-listed companies from their 2008 annual financial statements. The dependent variable is Tobin's q defined

Table 5.1 Good practices on the Warsaw stock exchange in 2008: linear regression with Tobin's q

Dependent variable: Tobin's q = the market value plus the book value of liabilities and reserves divided by the book value of assets as of 31.12.2008	
Explanatory variables:	Estimates (significance)
ICG = Index (score) of corporate governance	−0.003
# 1 = Share of votes of the largest shareholder (%)	0.456**
Equal = 1 if shareholders rights are equal, 0 otherwise	−0.068
Ln(assets) = Log of total assets	−0.139***
Ln(debt ratio) = Log of (total liabilities/total assets)	0.368***
DPR = Dividend payout ratio	8.050***
Constant	4.125***

n = 284 companies listed on Warsaw Stock Exchange; data for 2008
R^2 = 0,251; *** $p < 0.01$, ** $p < 0.05$
Source: Garstka (2009)

here as the ratio of the market value plus the book value of liabilities and reserves divided by the book value of assets.

The author's attempts did not reveal a statistical association between compliance with the provisions of the CG code and the companies' financial results. He claims that constructing the *ICG* using the companies' own statements might not be appropriate. On another point, the results certainly suffer from not recognizing the possible endogeneity of the *ICG* as well as from neglecting the heterogeneity of companies listed on the WSE.

■

5.2 Research Topics in Corporate Governance

Research Questions

Scientific research on corporate governance concerns all the issues presented in the previous section. The research is naturally diverse, with various targets and theoretical underpinnings and spanning several disciplines: economics, law, management, finance, and accounting. Most studies use the methodologies of statistics and econometrics—in fact, financial microeconometrics.

The quantitative research on corporate governance may be roughly separated into two streams. The first is empirical, with statistical/econometric attempts to prove diverse hypotheses, such as associations between various CG indicators and measures of companies' performance. The second is theoretical, with solutions broadly fitting into the realm of mathematical economics. The classic theoretical foundation of corporate governance studies is agency theory, which assumes that the separation of ownership and control generates asymmetries between the goals of owners and managers. The divergent interests of both groups can be mitigated through (agency) contracts between the parties.

Quantitative studies in corporate governance include, for example, the relationships between various categories representing a company's performance and those describing the governance level, such as the ownership structure or the composition of the supervisory body. The number of papers and other contributions is overwhelming and constantly expanding, and surveys of the research in this area quickly become outdated. Today, almost all research in corporate governance has a quantitative focus. We mention here surveys by:

- Shleifer and Vishny (1997)—On worldwide CG research emphasizing the legal protection of investors and ownership concentration
- Adams et al. (2009) on research examining the director selection process and board composition, and their effect on board actions and firm performance
- Boyd et al. (2017)—An editorial to a methodological survey of CG research (in *Corporate Governance: An International Review*)

- Aguilera et al. (2016)—An editorial to a comprehensive survey of all aspects of CG research (in *Corporate Governance: An International Review*)
- Filatotchev and Boyd (2009)—An editorial to a survey of CG research (in *Corporate overnance: An International Review*)
- Bhagat and Jefferis (2002)—On US mergers and acquisitions in the context of CG
- Börsch-Supan and Köke (2000)—On econometric issues of CG research
- Gugler (2001)—On CG and economic performance worldwide

Typically, empirical attempts aim at examining such factors of CG as

- The composition of the supervisory board: Independent board members, institutional members, and the relationship of the board to the CEO
- The ownership structure: Diluted ownership, concentration of ownership, corporate owners, and managerial ownership
- Acquisitions (including management buyouts)
- CEO replacement
- The equity structure (debt structure)
- Managerial compensation

Brown et al. (2011) enumerate research topics on the financial outcomes of CG and the accounting outcomes of CG. The list can be found in Sect. 5.1.

Generally, most studies confirm that CG is related to the performance of companies. A survey of corporate governance in OECD countries (OECD 2004) indicates that "studies using what are considered to be best practice econometric techniques indicate that the corporate governance is an important determinant of performance.. . . As with all regression work, the question of 'causality' will never be resolved fully to everybody's satisfaction."

Let us also mention here that the "political" debate on how shareholder intervention improves economic performance. Dallas and Pitt-Watson (2016) indicate that "it is difficult to measure the specific effects of granting powers to shareholders, and in popular political debate, there is often an assumption that shareholder interventions are short-termist in nature and thus granting further powers to shareholders may not be good public policy." This has recently been expressed by practitioners in the form of demanding a "new paradigm" of corporate governance that focuses on long-term solutions (Lipton 2019), although academics do not support this view. For example, Bebchuk et al. (2015) found "no evidence that [shareholder] activist interventions . . . come at the expense of long-term performance."

Methodological Issues

Microeconometric research in corporate governance is intertwined with studies in corporate finance, in accounting, and especially in studies involving issues of transparency. While the research questions may vary, the methodology is common, and common also are the methodological dangers. Therefore, most methodological questions discussed previously in Chaps. 2 through 4 are valid as well here, in corporate governance studies.

A special issue of *Corporate Governance: An International Review* (Vol. 25, issue 6, 2017) entitled "Research methodology of governance studies: Challenges and opportunities" is devoted to methodology—mostly microeconometric methodology. The authors of eight papers focus on highlighting the methodological problems in corporate governance studies. Among the eight, the paper by Adams (2017) "The ABCs of empirical corporate (governance) research" was quoted in Chap. 2, along with other good practices of microeconometric research in corporate finance.

In an earlier paper, Börsch-Supan and Köke (2000) identified the following methodological questions of microeconometric research on CG:

– *Structural reverse causality*: For example, the direction of causality between the ownership structure and the firm's performance is not clear. More concentrated ownership can improve firm performance, but the reverse relation is also possible: firms well assessed by the market could also attract investors.
– *Missing variables*: It is customary in the area of corporate governance not to include major explanatory variables in the model; moreover, the linear specification of the equations excludes the presence of higher order terms.
– *Sample selectivity*: Most empirical studies on corporate governance analyze only the largest companies (usually listed). Such samples are selected by a "performance" variable and the studies effectively have sample selection bias.
– *Measurement error in variables*: For example, a company's performance can be measured by different variables (e.g., market value, ROA, ROE, EBIT, Tobin's q), but these variables are sometimes uncorrelated—i.e., measure the same performance in a different way.

In the next subsections, we present examples of CG research using the methodology of microeconometrics.

Corporate Governance and the Performance of Companies: Two Studies

Example 5.2 Tobin's q and Corporate Governance for Companies Listed on the Oslo Stock Exchange

Bøhren and Ødegaard (2006) used data on all companies listed on the Oslo Stock Exchange during the period 1989–1997—217 firms at the end of 1997. The dataset contains 868 firm-years. One of the models is shown in Table 5.2. It is the linear regression of Tobin's q against variables representing the ownership concentration, insider holdings, owner type, board characteristics, security design, financial policy, and control variables (including the company sector). The concentration variable is

Table 5.2 Tobin's q and corporate governance variables for the Oslo Stock Exchange

Dependent variable: Tobin's q	
Explanatory variables:	*Sign of estimate (p)*
Ownership concentration (the fraction of equity held by the largest owner)	−***
Insiders (fraction held by officers and directors)	+***
Squared (insiders)	−**
Aggregate state holdings	−
Aggregate international holdings	+
Aggregate individual holdings	+***
Aggregate nonfinancial holdings	−
Ln (board size)	−**
Fraction of voting shares	+**
Debt to assets	−***
Dividends to earnings	−**
Industrial company	−***
Transport/shipping company	−***
Offshore company	−***
Investments to income	−
Ln (equity value)	+**

$n = 868$; $R^2 = 0.29$; *** $p < 0.01$, ** $p < 0.05$
Linear regression model
Source: Bøhren and Ødegaard (2006)

measured as the fraction of holdings of the largest owner. In other attempts, the authors showed that the estimation results are insensitive to the choice of the concentration variable (e.g., the Herfindahl index). Also, in order to avoid perfect collinearity, there are four aggregate holdings per type included in the model (the base type is "financial owners" and it is excluded).

The authors maintain that the results confirm the agency theory: large outside owners are negatively associated with firm economic performance, while insider ownership (not very large) is related positively to performance. In particular:

– Estimation results indicate a significant inverse relationship between outside concentration and Tobin's q.
– Insider holdings are value creating up to 60% (negative sign by the estimate of "squared insiders").
– The fraction of individual investors has significant association with company performance.
– Small boards prove better than large in terms of firm performance.
– Firms issuing shares with unequal voting rights may lose market value.

The model from Table 5.2 was estimated with the OLS on pooled data from 1989 to 1997. The robustness of this model has been shown by estimating each year separately and also by using the GMM (Generalized Method of Moments) on the polled data as well as the OLS on the pooled data with fixed annual effects. The authors state that the main conclusions "survive across a wide range of equations"—the major CG variables have parameter estimates with the same sign and are "more-or-less" significant.

In order to solve the problem of possible endogeneity, the authors attempt the use of a three-equation model:

Performance = f (Concentration, Insiders, Other variables, Instruments)
Concentration = f (Performance, Insiders, Other variables, Instruments)
Insiders = f (Performance, Concentration, Other variables, Instruments)

The first equation represents the original model, while the second and the third assume the effects of "reverse causality." The authors estimate nine variants of each of the three equations: the performance equation, the concentration equation, and the insiders equation—with a choice of instruments. They find that the relationships are sensitive to the choice of instruments and conclude that "simultaneous system modeling is not necessarily superior to single–equation models when exploring the relationship between governance and performance." It may be the result of the underdeveloped theory of how governance and performance interact.

■

Example 5.3 Women on Boards and Firm Risk in the USA
Sila et al. (2016) examine the gender diversity of corporate boards and its possible effect on company risk. The sample of US firms is taken from RiskMetrics, Compustat, Execucomp, and CRSP databases with 13,581 firm-year observations in 1960 companies for the period 1996–2010.

1. *Linear regression (DPS-GMM)*

The authors propose a linear regression model with risk as the dependent variable and the proportion of women on the board as an explanatory variable, with control variables, and lagged risks among the other explanatory variables.

$$Risk_{it} = \alpha + \beta Proportion\ of\ Women_{it} + \boldsymbol{x}'_{it}\boldsymbol{\gamma} + \sum\nolimits_{s=1}^{p} \delta_s Risk_{i,t-s} + \{\eta_i + \varepsilon_{it}\} \tag{5.1}$$

where $Risk_{it}$ is the measure of the i-th company's risk in year t, vector $\boldsymbol{x}'_{it}$ represents control variables, η_i is the random term specific for the i-th company, ε_{it} is the random term for the i-th company and year t.

The dependent variable (risk) is represented by three measures: total risk (the standard deviation of daily stock returns over the last year), systematic risk (the coefficient on the stock market portfolio from a market model regression using the CRSP/NYSE/AMEX/NASDAQ/ARCA equally weighted index), and idiosyncratic risk (standard deviation of the residuals from the market model regression).

The authors considered the endogeneity issue: the board characteristics are not exogenous; boards are endogenously chosen by firms to suit their operating environments and the power of various stakeholders. Female boardroom representation is a choice made by the firm and this choice may be influenced by unobservable factors such as the CEO's abilities and corporate culture (included in η_i). Therefore, the OLS is rejected since it would give an inconsistent estimate of β. For other reasons, the fixed effects estimator is also excluded. The authors decided to use DPS-GMM—i.e., a dynamic panel system GMM (generalized method of moments) estimator.

2. *Binomial probit*

In order to confirm that the attempted dynamic model is suitable, the authors began by estimating the binomial probit explaining the probability that at least one female director is appointed in the *i*-th company for year *t*. The question is whether risk affects the appointment of female directors. The authors focus on those firm-years when at least one director is appointed. The number of observations is 7101. Table 5.3 shows the result of estimating one estimated probit models.

In Table 5.3 the variable "proportion of male directors with board connections to women" is the proportion of male directors who sit on the board of another firm with at least one female director. Variable "board independence" is the proportion of outside directors. This probit model gives evidence that the risk variable may be the determinant of a female appointment.

3. *Estimation of DPS-GMM*

The next step in the authors' study was the estimation of model (5.1) since the probit model delivered the evidence that "past risk influences the choice of selecting women into the boardroom." The authors preliminarily include two lags of risk measures in (5.1). The model is estimated with the use of DPS-GMM. The result is that the authors found no evidence "in support of the view that female directors reduce equity risk." In the three models explaining total risk, systematic risk, and idiosyncratic risk, the explanatory variable "proportion of women on the board" is not significant. In addition, one robustness check applied by the authors revealed that low-risk firms tend to have a higher proportion of female directors on their boards.

Table 5.3 Determinants of gender in director appointments: binomial probit model

Dependent variable: Female appointment = 1 when a women director is appointed, = 0 otherwise	
Explanatory variables:	*Sign of estimate (p)*
Total risk	−**
Woman departing the board (= 1 if yes, = 0 otherwise)	+***
Man departing the board (= 1 if yes, = 0 otherwise)	−
Proportion of male directors with board connections to women	+**
Proportion of women on the board	−***
Board size	+
Board independence	+
Ln (total assets)	+***
Market-to-book	+***
R&D expenditure	−
Capital expenditure	−
Leverage	−
ROA	−

$n = 7101$; pseudo-$R^2 = 0.059$; *** $p < 0.01$, ** $p < 0.05$
Linear regression model
Source: Sila et al. (2016)

But overall, “there is no robust evidence suggesting that higher female boardroom representation leads to lower equity risk.”

4. *Diff-in-Diff with matching*

Assuming that the “proportion of female directors” might not be the appropriate metric to identify gender diversity, the authors considered an alternative identification strategy in the second part of their study. The approach is the Diff-in-Diff matching estimator that is the combination of Diff-in-Diff and matching. The methodology of Diff-in-Diff (described earlier in Sect. 2.12) in the authors’ version amounts to estimating the following model:

$$\begin{aligned} Risk_{it} = \alpha_0 &+ \alpha_1 Female\ Appointment_{it} * Post\ Period_{it} \\ &+ \alpha_2 Female\ Appointment_{it} + \alpha_3 Post\ Period_{it} + \boldsymbol{CONTROL}'_{it}\boldsymbol{\gamma} \\ &+ \varepsilon_{it} \end{aligned} \tag{5.2}$$

where $Female\ Appointment_{it}$ is a binary variable that equals 1 when the firm is in the treatment group, equals 0 otherwise. The treatment group comprises the firms that appoint exactly one female director in a given year to replace a departing male director (who must be older than 60). The variable $Post\ Period_{it}$ is binary and equals 1 in the posttreatment period, equals 0 in the before-treatment period. The choice of treatment group excludes appointments that may be the result of a strategy change that might affect the risk.

The authors identified 153 appointments in the treatment group and 737 (control) cases where a male director is appointed to replace another male director (also older than 60). Then the treatment firms are matched to similar control firms with propensity score and nearest neighbor matching techniques. The Diff-in-Diff model (5.2) is estimated for the data resulting from both matching techniques. In both cases, the $Female\ Appointment_{it} * Post\ Period_{it}$ variable is not significant in the model. This means that treatment has no impact on the outcome: the difference between the changes in the two firms (treatment and nontreatment) is zero. In other words, the difference in risk of firms with female director appointments is not significantly different from firms with male director appointments.

All in all, this comprehensive study shows that “a board with a higher proportion of female directors is no more or less risk-taking than a more male-dominated board.” This result might be an example of how to conduct thorough research in examining links between demographic characteristics of decision-makers and firm outcomes. It is necessary “to causally isolate firm outcomes from between-firm heterogeneous factors that influence both the demographic characteristics in the boardroom and the firm outcomes.”

■

Firm Performance and CEO Change: Two Studies

Example 5.4 CEO Turnover in Fortune 500 Firms

Kaplan and Minton (2012) studied CEO turnover in large US companies during the period 1992–2007. The authors considered 10,715 firm-years and 1698 CEO turnovers that happened during this time span (turnover of 15.84%). The comprehensive study included examination of turnover changes over time and their relation to firm performance (especially to stock performance). It also distinguished between internal and external turnovers. External turnovers are those due to mergers, acquisitions, and delistings from a major stock exchange. Internal turnovers accounted for 1254 turnovers or 11.79% of firm-years.

One model is devoted to *Fortune* 500 firms. In the authors' dataset, the Fortune 500 firms comprise 7631 firm-years and 1148 CEO turnovers (15.04%). There are 881 internal turnovers (i.e., 11.55% of all observations).

The model is multinomial logit and is estimated on pooled annual data (firm-years). The dependent variable represents three states of CEO turnover in the *Fortune* 500 firms:

- No CEO turnover
- Unforced CEO turnover
- Forced CEO turnover

Turnover is classified as forced "if an article in the business press indicates that the CEO was fired, forced, or left following a policy disagreement or some other equivalent." In addition, the succession is classified as forced when the CEO is under 60 and "the first article reporting the announcement does not report the reason for the departure as involving death, poor health, or the acceptance of another position elsewhere." The authors do not disclose the number of cases for each state of variable "CEO turnover."

The explanatory variables for this model represent market returns with the exception of CEO age that is defined as a dummy variable equal to 1 if the lagged CEO age is greater than or equal to 60, but equal to 0 otherwise. Other regressors are:

(a) The annual stock market return on the S&P 500 index.
(b) The relative industry performance—the difference between the return on the median firm in the industry and the return on the S&P 500 index.
(c) The relative firm performance (the industry-adjusted stock return)—the difference between a firm's stock return and the return on the median firm in the industry.

Table 5.4 presents the marginal effects (labeled Δprob) of each explanatory variable on the probability that the dependent variable is in the state "Unforced CEO turnover" and in the state "Forced CEO turnover." The marginal effects are changes in the probabilities associated with unit changes in the regressors. The results show that the probabilities of forced and unforced turnover are negatively related to the components of a firm's stock performance: (a), (b), and (c) as well as

Table 5.4 Trinomial logit estimates of CEO turnover for Fortune 500 firms in 1992–2007: marginal effects

Dependent variable: CEO turnover with three states: No CEO turnover, unforced CEO turnover, and forced CEO turnover		
Explanatory variables:	Unforced CEO turnover Marginal effect Δprob	Forced CEO turnover Marginal effect Δprob
(a) Return on S&P 500	−0.1402***	−0.0341***
Lagged return on S&P 500	0.0761***	0.0182***
(b) Industry return—Return on S&P 500	−0.0202	−0.0229***
Lagged industry return—Return on S&P 500	−0.0454***	−0.0037
(c) Industry-adjusted stock return	−0.0285***	−0.0288***
Lagged industry-adjusted stock return	−0.0399***	−0.0130***
CEO age dummy	0.1517***	0.0087***

$n = 7631$; pseudo-$R^2 = 0.0878$; *** $p < 0.01$
Source: Kaplan and Minton (2012)

positively related to lagged market performance (a). Firms with the CEO age dummy equal to 1 have higher probabilities of both CEO turnover types, forced and unforced.

The estimated model confirms that both types of turnover are negatively related to market performance: the poorer the performance, the higher the probability of turnover. The authors state “this strongly suggests that a number of unforced turnovers are not voluntary.”

■

Example 5.5 CEO Turnover and Mandatory IFRS Adoption in Europe
Wu and Zhang (2019) studied CEO turnover in the context of the adoption of the International Financial Reporting Standards (IFRS). Mandatory IFRS adoption is usually related to improvement in the informative aspect of accounting earnings. Research shows that “the CEO turnover responds more to a firm’s accounting performance after adoption” and this increase in turnover-to-earnings sensitivity is evident in countries with stronger enforcement of financial reporting.

The authors concentrated on Western European countries. The sample includes 7214 firm-year observations (2082 firms) during the period 2002–2008, largely from the UK (65% of the observations) as well as 15 other countries.

The major model presented in the study is the binomial logit with $CEO\_Turnover_{it}$ as the dependent variable: $CEO\_Turnover_{it} = 1$ if there is a CEO turnover in firm i in year t, and equals 0 otherwise. The explanatory variables represent both financial-accounting characteristics and market returns. Table 5.5 presents stylized estimation results (signs of estimates).

The explanatory variables are

- $\Delta ROA = \Delta ROA_{i,\, t-1}$ = The lagged change in earnings before interest and taxes over total assets (in year $t-1$)
- $Post = 1$ for the post-adoption firm-years (2006–2008), and $= 0$ for the pre-adoption firm-years (2002–2004)
- $RET = RET_{i,\, t-1}$ = The raw stock return over the prior fiscal year

Table 5.5 Binomial logit estimates: mandatory IFRS adoption and the effect of accounting earnings on CEO turnover: European firms in 2002–2008

Dependent variable: $CEO\_Turnover_{it} = 1$ or $CEO\_Turnover_{it} = 0$ (all turnovers, including voluntary)	
Explanatory variables:	*Sign of estimate (p)*
ΔROA	$-**$
$Post \times \Delta ROA$	$-***$
RET	$-**$
$Post \times RET$	+
Age	+***
$Size$	+***
BTM	+*
$Leverage$	+*
Year, country, industry effects	Yes

$n = 7214$; number of CEO turnovers = 939;
Pseudo-$R^2 = 0.0387$; $*** \ p < 0.01$, $** \ p < 0.05$, $* \ p < 0.1$
Source: Wu and Zhang (2019)

- $Age = Age_{i,\,t-1}$ = The age of the CEO at the end of year t–1
- $Size = Size_{i,\,t-1}$ = Firm size: the natural log of lagged market capitalization (in euros)
- $BTM = BTM_{i,\,t-1}$ = The book value of equity over market capitalization (lagged)
- $Leverage = Leverage_{i,\,t-1}$ long-term debt over total assets (lagged).

The product $Post \times \Delta ROA_{i,\,t-1}$ is designed to capture incremental turnover-to-performance sensitivity following IFRS adoption, similarly $Post \times REIT_{i,\,t-1}$. The results indicate that performance measured by $\Delta ROA_{i,\,t-1}$ and $RET_{i,\,t-1}$ is negatively associated with CEO turnover. Also, the negative estimate of the parameter by $Post \times \Delta ROA_{i,\,t-1}$ is evidence that CEO turnover is more sensitive to "accounting performance" measured by $\Delta ROA_{i,\,t-1}$ following mandatory IFRS adoption than pre-adoption.

■

The Accounting Effects of CG: Two Studies

Example 5.6 CG Quality and Earnings Management: European Companies Cross-Listed in the USA

Bajra and Cadez (2018) investigated the association between earnings management and corporate governance represented by the quality of the internal audit and the quality of the board of directors. The sample includes 127 European-based publicly traded companies that are also cross-listed on US equity markets. Data for this study encompass the 14-year period 2000–2013.

The hypothesis is that the quality of the internal audit and the quality of the board of directors are negatively related to earnings management. The dependent variable is earnings discretionary accruals $EDAC_{it}$ that serves as a proxy for earnings management. As the authors state, "Unlike the nondiscretionary component, which reflects business conditions (such as growth and the length of the operating cycle) that naturally create and terminate accruals, the discretionary component identifies management choices."

The variable $EDAC_{it}$ is the difference between total accruals and nondiscretionary accruals. Total accruals $TAcc_{it}$ are estimated using two models, one of which is the following "modified Jones model":

$$TAcc_{it} = \alpha_0 + \alpha_1 \frac{1}{Toas_{i,t-1}} + \alpha_2 \frac{\Delta Rev_{it} - \Delta Rec_{it}}{Toas_{i,t-1}} + \alpha_3 \frac{PPE_{it}}{Toas_{i,t-1}} + \varepsilon_{it} \tag{5.3}$$

where

- $TAcc_{it}$ = total accruals, equals the change in current assets—the change in current liabilities—the change in cash flow—depreciation and amortization for firm i in year t
- $Toas_{it}$ = total assets
- ΔRev_{it} = the change in revenues between years t and t–1
- ΔRec_{it} = the change in accounts receivable between years t and t–1
- PPE_{it} = gross property, plant, and equipment
- ε_{it} = error term.

The value of $EDAC_{it}$ is the residual from the estimated model (5.3). The residual represents the part of total accruals that is not explained by changes in total assets, liabilities, cash, and depreciation.

The primary model is presented in stylized form in Table 5.6. This is the multiple regression estimated with the OLS on 1502 firm-years of data and the dependent

Table 5.6 OLS estimates: earnings management and corporate governance quality (modified Jones model): European firms cross-listed in the USA in 2000–2013

Dependent variable: $EDAC_{it}$ (represents earnings management) = residual from (5.3)	
Explanatory variables:	*Sign of estimate (p)*
IAFQ	−***
BoDQ	−***
IAFQ×BoDQ	+***
CFO	−***
ROA	−**
lagROA	−***
SIZE	−***
DEBT	−*
PROFIT	+**
IFRS	+
BIG4	+***
Constant	+***

n = 1502;
$R^2 = 0.24$; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Source: Bajra and Cadez (2018)

variable is $EDAC_{it}$. The explanatory variables include measures representing corporate governance in the company, as follows:

- $IAFQ_{it}$ = Internal audit function quality for the firm i in year t: measured on a scale from 0 (low) to 5 (high quality) where points (0 or 1) are given for formal existence, proficiency, size, independence, and involvement in financial statement audits
- $BoDQ_{it}$ = Board of directors' quality: measured on a scale from 0 (low) to 5 (high quality) where points (0 or 1) are given for board size, independence, frequency of meetings, financial expertise, and board rotation
- CFO_{it} = Cash flow from operations
- ROA_{it} = Return on assets
- $SIZE_{it}$ = Natural log of total assets
- $DEBT_{it}$ = Debt to total assets
- $PROFIT_{it}$ = Indicator (dummy) variable = 1 if firm i made a profit, = 0 if a loss in year t
- $IFRS_{it}$ = Indicator (dummy) variable = 1 if the firm's financial statements are prepared using IFRS, = 0 otherwise
- $BIG4_{it}$ = Indicator (dummy) variable = 1 if the firm's financial statements have been audited by one of the BIG4 firms, = 0 otherwise.

The results indicate that both measures of corporate governance internal quality (*IAFQ* and *BoDQ*) are related negatively to the measure of earnings management (*EDAC*). The interactive effect of *IAFQ* and *BoDQ* has a positive association with *EDAC* that (in the authors' statement) is not as expected. The control variables generally have the expected signs of relation to the *EDAC* variable (except for *PROFIT* and *BIG4*).

The authors underscore that their study reveals negative associations between both the mechanisms of corporate governance (internal audit function quality and board of directors' quality) and the incidence of earnings management.

■

Example 5.7 Corporate Governance and Earnings Manipulation in Spain
Osma and Noguer (2007) studied the association between earnings management and two elements of corporate governance: board composition and the existence of board monitoring committees. The country is Spain and the data were collected in 1999–2001 through a voluntary questionnaire about the degree of compliance with good governance recommendations. There are 155 firm-year observations in the sample.

The authors tested whether accounting manipulation is related to the proportion of institutional directors, the proportion of independent directors, and the existence of independent audit and nomination committees. Accounting manipulation is defined with the "Jones model," similar to (5.3), as follows:

$$\frac{TA_{it}}{A_{i,t-1}} = \frac{\beta_0}{A_{i,t-1}} + \beta_1 \frac{\Delta REV_{it}}{A_{i,t-1}} + \beta_2 \frac{PPE_{it}}{A_{i,t-1}} + \varepsilon_{it} \tag{5.4}$$

where

- TA_{it} = total accruals for firm i in year t
- $A_{i,\,t-1}$ = lagged total assets
- ΔREV_{it} = change in revenues
- PPE_{it} = gross level of property, plant, and equipment.

The residuals from this "discretionary accruals" model are taken (in absolute value) as the dependent variable in the model with the corporate governance variables as regressors.

The percentage of institutional directors on the board is shown to be negatively associated with earnings manipulation (unsigned discretionary accruals). However, the percentage of independent directors is associated positively with accounting manipulation and—as the authors' state—this coincides with the doubts expressed in Spain about "the real independence of these directors." Also, the existence of audit and nomination committees is not connected with earning manipulations—in this particular sample.

■

5.3 Indices and Ratings of Corporate Governance

Corporate governance indices and ratings have naturally emerged among researchers and practitioners as the answer to the need to express complicated CG issues in a single number. However, the abundance of proposals and applications encouraged Klausner (2018) to state that "the use—or, as I will explain, misuse—of corporate governance indices has reached epidemic proportions." Before we study this statement further, let us first identify the major proposals for CG indices, historically and geographically.

The number of elements that constitute a complete picture of corporate governance is usually large. Moreover, the elements differ in terms of substance, importance, and validity. Therefore, constructing a quality CG rating or CG index is a challenging task. Black et al. (2017) studied the construct validity of firm-level indices in four economies: Brazil, India, South Korea, and Turkey. They considered overall governance indices composed of subindices, and subindices composed of governance elements. The finding is that, at subindex level, the construct validity is reasonable. The methodology applied to confirm the construct validity is Cronbach's alpha, inter-item correlations, and principal component analysis.

Below we comment on CG indices that have been proposed at country levels, especially for the USA, as well internationally.

CG Ratings and CG Indices at the Country Level

Corporate governance ratings and corporate governance indices have been proposed by researchers, practitioners, and institutions in countries with developed capital markets. For example, in Poland the popular CG rating is recommended by the Polish Corporate Governance Forum and the Polish Institute of Directors. Another index for Polish companies was proposed by Kowalewski et al. (2007) and named the TDI: *transparency and disclosure index*.

A CG index was constructed for Greek-listed companies by Lazarides and Drimpetas (2009) and for German companies by Drobetz et al. (2003). In the book on investor protection and corporate governance in Latin America edited by Chong and López-de-Silanes (2007), various CG and transparency indices were proposed for Argentina, Brazil, Chile, Colombia, Mexico, and Venezuela. The Korean CG index (KCGI) was proposed by Black et al. (2002). The Brazilian CG index (BCGI) stemmed from the works of Carvalhal da Silva and Leal (2005) as well as Black et al. (2014). The CG index for Indian companies was presented by Varshney et al. (2012).

These are just a few examples of indices proposed at the national level. In addition, researchers tend to compose their own CG indices for the purpose of their particular studies.

Corporate governance ratings are also proposed by business magazines and newspapers like *The Globe and Mail* in Canada. Examples of analyst ratings are presented below for the US market.

International CG Ratings and Indices

The CG ratings and indices considered internationally are usually those that are the most popular in the USA (see below). In addition, the commercially constructed indices usually have international use. An example is the rating by the AllianceBernstein Capital Emerging Market Universe that is available to researchers (with some delay). Klapper and Love (2003) used the CG index formulated by Credit Lyonnais Securities Asia that covered 495 firms across 25 emerging markets and 18 sectors.

Martynova and Renneboog (2010) examined corporate governance legal regimes and their evolution over time (from 1990 to 2005) in 30 European countries and the USA. The authors developed three CG indices that show the quality of national laws aimed at protecting (i) corporate shareholders from being expropriated by management, (ii) minority shareholders from being expropriated by a large blockholder, and (iii) creditors from being expropriated by shareholders.

CG Ratings and CG Indices in the USA

The most debated "generic" CG indices have been proposed for US companies. This is because corporate governance issues are most visible on developed capital markets like the USA, especially in times of financial turmoil. New laws directed at more protection for investors (i.e., better corporate governance) usually follow periods of financial disturbance.

Here is a selection of CG ratings and indices that have emerged in the USA:

Ratings[2]:

- TCL: The Corporate Library (*Board Effectiveness Rating*)
- GMI: Governance Metrics International (*Market and Industry Indices*)
- RiskMetrics (*Corporate Governance Quotient*)
- Egan-Jones Proxy Services (*Corporate Governance Rating*)

Indices:

- *G-Index*: Gompers et al. (2003)
- *E-Index*: Bebchuk et al. 2009
- *Gov-Score*: Brown and Caylor (2006)
- Glass, Lewis & Company (*Board Accountability Index*)

As an example, the popular *G-index* is constructed on a range from 1 to 24, from the standpoint of shareholder protection: 1 means the strongest, and 24 the weakest protection.

Corporate governance indices and ratings are often intertwined with similar measures directed at another dimension. For example, the corporate disclosure indices (see Chap. 4) include many elements of CG. Other examples are multiple indices of CSR (*corporate social responsibility*) that consider governance aimed at social responsibility issues.

Discussion on Uses and Misuses of CG Indices: Klausner (2018)

Klausner (2018) discusses the *G-Index* created by Gompers et al. (2003) and the *E-Index* by Bebchuk et al. (2009). The elaboration from a legal perspective indicates that the author believes that indices do not adequately represent CG laws, bylaws, and state corporate law. An example of the author's views is as follows: "It is doubtful that anyone familiar with the details of corporate charters, bylaws and state corporate law would have initiated a research program with the hypothesis that the elements contained in the *G* Index would correlate with firm value or firm

[2]See footnote 3 in Chap. 4 on GovernanceMetrics International, Audit Integrity and The Corporate Library.

performance, let alone the possibility of a causal relationship." Overall, the author concludes that both indices (*G* and *E*) employ simplification of complex relationships that does not adequately reflect the actual functions of the index components: "Their elements do not have any justification in terms of how corporate governance works." Researchers use the indices mechanically without understanding and that creates confusion in corporate governance literature.

Six Myths: Armstrong et al. (2010), Brickley and Zimmerman (2010)

In a comprehensive report Armstrong et al. (2010) present a survey of 10 years of literature on the role of transparency in financial reporting in reducing agency conflicts. The report expresses much skepticism as to the validity of (scientific) views taken for granted in corporate governance research. In a comment to the paper, Brickley and Zimmerman (2010) follow up on these doubts and formulate "six myths" of corporate governance. These are:

1. A common definition of "corporate governance" exists.
2. A useful distinction is "internal" versus "external" governance mechanisms.
3. Outside directors perform two separable roles: to advise and to monitor managers.
4. Research has identified "good" and "bad" governance practices.
5. A "good" governance index can be constructed.
6. Corporate governance "best practices" can be deduced from peer data.

These "truths and beliefs" of researchers and other parties involved in corporate governance are very much in doubt in the authors' view. Their exposition supplies many reasons in support of this stand; however, the word "myth" seems to be an exaggeration. As for the CG indices ("myth 5"), the authors state that "the firm's governance system consists of multiple components" and CG quality does not depend on one element alone. However, a single index of CG is not feasible. Why? There are three reasons. Firstly, the entire system of corporate governance is endogenous: "Without additional structure, the driving forces behind any observed associations among endogenous variables (such as a governance index and performance) are impossible to determine." The second reason is that a distinction between "good" and "bad" features of CG is clearly not possible. Third, there is no possibility of attaching sensible weights to elements of CG to construct a reliable index—moreover, the weights may differ across firms.

Seven (Other) Myths: Larcker and Tayan (2011, 2015b)

Worldwide financial crises and growing distrust of existing legal solutions and the operations of public institutions call for new remedies that have been proposed on the basis of convictions and beliefs in the area of corporate governance. This new awareness has been uncritically applied in research and practice. Larcker and Tayan (2011) point out that in many cases the "truths" or "certainties" are unsubstantiated. The authors reveal and discuss the following seven truths/myths about CG:

1. The structure of the board = the quality of the board.
2. CEOs are systematically overpaid.
3. There is no "pay for performance" in CEO compensation.
4. Companies are prepared for CEO succession.
5. Regulation improves corporate governance.
6. Voting recommendations are based on rigorous research.
7. Best practices are the solution.

The authors dissect each of these statements with examples, cases, research outcomes, statistics, etc. The CG indices are commented under "myth (7)." The authors state that

> [t]he most destructive myth in corporate governance is the notion that best practices exist which, if uniformly followed, lead to better oversight and performance.. . .Despite the best efforts of regulatory, commercial, and academic experts, no one has yet identified standards that are consistently associated with improved corporate outcomes. This includes the recommendations of blue-ribbon panels, corporate governance ratings, and governance indices. (Larcker and Tayan 2011)

A similar argumentative tone is present in the subsequent paper by the same authors (Larcker and Tayan 2015b) on the "seven myths of boards of directors."

CG Indices and Company Performance (Bebchuk et al. 2013, Bebchuk et al. 2010)

Despite these concerns, CG indices have been used in numerous studies. One example is the *E-Index* (Entrenchment Index) by Bebchuk et al. (2009) that as of May 2015 had been applied in more than 300 research papers (Harvard Law Today, June 11, 2015).

More recent research by (Bebchuk et al. 2013, 2010) on the then-current value of CG indices was awarded the 2013 annual prize from the Investor Responsibility Research Center Institute (IRRCI)—now the Weinberg Center for Corporate Governance at the University of Delaware. The study found that the E-Index and the G-Index (of Gompers et al. 2003) no longer produce abnormal returns in trading strategies based on these indices. What had been the case in the 1991–1999 period, no longer holds for the years 2000–2008 (or for any subperiods within this period).

The authors provide evidence based on "market participants' learning to appreciate the difference between scoring well and poorly on governance indices." They document that

> (i) attention to corporate governance from the media, institutional investors, and researchers has exploded in the beginning of the 2000s and remained on a high level since then, and (ii) until the beginning of the 2000s, but not subsequently, market participants were more positively surprised by the earning announcements of good-governance firms than by those of poor-governance firms. (Bebchuk et al. 2010)

The results prove that an effective investment strategy based on indices G or E is currently not possible. However, the G and E indices are still negatively correlated with Tobin's q and, according to the authors, "thus remain valuable tools for researchers, policy makers and investors."

5.4 Corporate Governance and Firm Performance in Poland[3]

CG Ratings for Polish Companies

This chapter on financial microeconometrics in corporate governance research concludes with this section devoted to past research on Polish corporate governance issues.

Corporate governance ratings of companies listed on the Warsaw Stock Exchange (WSE) have been published since 2001. In 2006, we have attempted to compare the ratings offered by two institutions and to discover whether the ratings of the companies are related to their financial performance. The two ratings under consideration are:

- Polish Corporate Governance Forum (PFCG) ratings
- Polish Institute of Directors (PID) ratings

The PFCG uses the term "ranking" but it is, in fact, a "rating" since one category is assigned to a given company from among the few quality categories. The PFCG and PID ratings differ methodologically. The key distinction lies in the philosophy of designing the related survey and in arriving at the final rating. Both ratings evaluate the degree of compliance with the "Best practices in public companies," the WSE's corporate governance code (published in 2002 and 2005), as well as with the OECD's "Principles of Corporate Governance" (published in 1999 and revised in 2004). However, not all principles are equally important for the two rating

[3]This final section of Chap. 5 is based on the paper presented at the 63rd International Atlantic Economic Conference, Madrid, 14–18 March 2007 (Gruszczyński 2007). The models and exposition presented in this section are reflective of the then current situation. The applications, however, are still highly relevant and very much worth examining.

institutions— PFCG and PID. They also differ in the methodology of collecting and assessing the data on governance.

Polish Corporate Governance Forum and Polish Institute of Directors ratings are no longer in use in Poland. However, the codes of best practices for WSE-listed companies are regularly amended and updated. In 2019, the code to be observed is "Best Practice for WSE Listed Companies 2016." For more codes, see https://ecgi.global/content/codes. The purpose of this section is to show how the rating system based on a code of best practice might be built, and what might be a method of assessing the relationship between companies' ratings and their financials.

The PFCG Rating

The Polish Corporate Governance Forum (PFCG), allied with the Gdansk Institute for Market Economics (IBnGR), has published CG ratings of WSE-listed companies since 2001. The ratings for 2001, 2003, and 2005 were published in Tamowicz et al. (2001), Dzierżanowski et al. (2004), and Dzierżanowski et al. (2006). The ratings were also published in *Rzeczpospolita* daily. In addition, IBnGR prepared the ratings for 2004, in an unpublished research paper by Aluchna et al. (2005).

Since the second rating for 2003, the methodology of PFCG ratings has been standardized and is based on an index constructed of 60 variables covering the following nine areas:

> the general characteristics of supervisory board (competence, number of members, etc.), the institution of the independent members, functioning of the management board, regulations concerning the general shareholders' meeting, strengthening the function of audit, exposition to the external control (i.e. the lack of defences against a hostile take-over), regulations on trading in own shares, companies' declared goals and intentions, transparency of information. (Dzierżanowski et al. 2004)

Data were collected by analyzing the disclosed documents (such as statutes, company internal regulations, bylaws of supervisory boards and general meetings of shareholders) including statements of compliance with the "Best practices in public companies" (with company comments), as well as company webpages. The formal declarations on compliance are additionally monitored by checking whether appropriate provisions are included in the bylaws.

The PFCG rating assigned to a company is based on the final score obtained from nine appropriately weighted governance areas and, since 2005, plus a 30% addition attributable to the assessment of the fund managers and analysts. The range of PFCG rating covers six categories: A (highest), A–, B+, B, B–, and C+.

The PID Rating

The corporate governance rating by the PID is based on the criteria set by the Rating Chapter of Institutional Investors. The chapter was composed of seven executives (independent of PID) representing institutional investors operating in Poland. The rating methodology uses the provisions of the OECD *Principles of Corporate Governance* (Polski Instytut Dyrektorów 2006).

The PID rating criteria focus on 12 areas:

1. Company's ownership structure
   - Transparency of ownership structure
   - Ownership concentration and owners' influence
2. General meeting of shareholders, relationships between the shareholders and the other interest groups
   - Information ensuring equal treatment of shareholders at general meetings
   - Voting and procedures at general meetings
   - Shareholder rights
3. Financial transparency and the accessibility of information
   - Quality and content of information disclosed by the company
   - Timeliness and accessibility of information disclosed
   - Independence and the status of the auditor
4. Composition of the supervisory board and the processes occurring within the board
   - Composition and structure of the supervisory board
   - Efficiency of the board's activities
   - Functioning and the role of the independent board members
   - Remuneration of the management board

The method of data collection differs from the PFCG rating. The PID rating uses a questionnaire addressed to all domestic institutional investors. The company is assessed if at least five institutional investors provide answers on the company's corporate governance characteristics. In the survey for 2005, PID received evaluations from institutions representing 82% of the domestic market of institutional investors. For each of the 12 areas, the assessing investor may give the company the grade from "−5" to "5". The total translates into the rating, in this case conveyed by the number of "stars." The best companies get five stars and are distinguished with the title "a trustworthy company." The PID rating spans five quality categories (stars).

The Level of Association Between the PFCG and PID Ratings

As one can judge from the above, the PFCG and PID ratings differ to a great extent, despite the common objective. The technique of collecting data is different. Also, different bodies have designed the rating rules: for the PFCG rating these are "market observers," while the PID ratings we have "market participants." From this latter point of view, perhaps the PID rating is closer to the market than the PFCG

Table 5.7 Correlation of PFCG and PID ratings

PFCG rating for the year	2005	2004	2003
Number of companies	55	51	51
PID rating for the year	2005	2004	2002
Number of companies	65	43	35
Both ratings	2005	2004	2002–3
Number of "common" companies	44	33	31
Spearman rank correlation			
Coefficient	0.5401	0.3150	0.2343
	**	*	

Source: Gruszczyński (2012); Significantly different from zero with $p = 0.01(**)$ or $p = 0.1(*)$

rating. On the other hand, the PFCG ratings were first on the Polish "ratings market"—since 2001—and, as such, they created an initial benchmark. We believe that PFCG and PID ratings may be compared due to the commonality of their rating objective, which is finding out how well companies listed on the Polish stock exchange comply with the rules of corporate governance. In addition, it may be worth uncovering whether the direct market impression on this issue is similar to the impression of (passive) market observers.

The company ratings are conveyed by rating categories—from lowest to highest. The distances between neighboring categories are unknown. One can perhaps assume that these distances are equal but testing such an idea might not be possible. Therefore, comparing the ratings is difficult, especially because the ratings' order is important. If companies X, Y, and Z are given categories 3, 2, and 1 in one rating and B, B–, and C+ in the second, one can say that in this case both ratings give the same result: the order of the companies is consistent.

A composite comparison of ratings can be accomplished by using the Spearman rank correlation coefficient between the rating results, separately for each year. This is possible only for the companies assessed in both ratings simultaneously. There are 31 such companies for the 2002 (2003) ratings, 33 for the 2004 ratings, and 44 for the 2005 ratings. The relevant ratings have been specified as numerical codes from 1 to 6 for the PFCG rating and from 1 to 5 for the PID rating. Table 5.7 presents the appropriate values of the Spearman rank correlation.

The rank correlations suggest that the association of the PFCG and PID ratings for Polish-listed companies in 2004 and 2005 may be considered as significant and positive. Both ratings assess the quality of corporate governance in the same direction. Moreover, the significance of this association seems to be increasing over time. The "severity" of the ratings may be compared by means of the average values of the numerical codes. Table 5.8 presents these averages, with both codes scaled on the same range of 1–5.

Table 5.8 shows that the PID ratings are more stable and, perhaps, also more "severe" in assessing the corporate governance level in Polish-listed companies.

Table 5.8 Mean values for the PFCG and PID ratings (only for the "common" companies in Table 5.7)

PFCG rating for the year	2005	2004	2003
Average in the range of 1–6	3.80	3.61	3.13
Average in the range of 1–5	3.16	3.01	2.61
PID rating for the year	2005	2004	2002
Average in the range of 1–5	2.98	3.03	3.16

Source: Gruszczyński (2012)

The Governance Level and the Financial Performance of Listed Companies

An attempt to examine the relationship between the financial performance of companies and their ratings has been undertaken with the use of methodology introduced in earlier studies (Gruszczyński 2004, 2006). The approach is to employ the ordered logit model in which the endogenous variable represents the rating categories while the exogenous variables are various financial characteristics of companies.

It is assumed that a company rating for a given year (e.g., 2005) may exhibit an association with the company's financial performance in that year (i.e., 2005). This is because the most recent ratings in Poland were prepared during the year following the rated one (here: during 2006), usually in the first half of the next year. Thus, the explanatory variables in the logit model are the financial indicators (ratios) calculated from the financial statements for the year of the rating (i.e., 2005). Gruszczyński (2006) considered the PFCG rating for 2003, prepared in fact in November 2003. Therefore, in that case only 2002 financial results were the source of the explanatory variables. Here the PFCG and PID ratings for 2004 and 2005 are modeled against the financial results. Only "common" companies are considered—i.e., 33 companies for 2004 and 44 companies for 2005. Furthermore, banks and other financial institutions have been excluded, so the final number of companies in the sample is 25 for 2004 and 34 for 2005. Hence, the samples are small. For this reason, the original 5 and 6 categories for the endogenous variables were also expressed in a 3-category setup, as explained in Tables 5.9 and 5.10.

The potential explanatory variables represent 20 financial ratios calculated from the companies' 2004 and 2005 financial statements (data source: *Notoria Serwis*). These are:

- Profitability ratios: P1 = gross profit from sales margin, P2 = operating profit margin, P3 = gross profit margin, P4 = net profit margin, ROE = return on equity, and ROA = return on assets
- Liquidity ratios: L1 = current ratio, L2 = quick ratio, and L3 = acid test
- Activity ratios: A1 = amount due turnover, A2 = inventory turnover, A3 = operating cycle, A4 = liabilities turnover, A5 = cash conversion cycle, A6 = current assets turnover, and A7 = assets turnover
- Debt ratios: D1 = fixed assets cover ratio (shareholders' equity/fixed assets), D2 = debt margin, D3 = EBITDA/financial expenses, and D4 = debt/EBITDA

Table 5.9 Structure of the sample for 2004

Rating PFCG 04	Code	No. of comp.	Rating 3PFCG 04	Code	No. of comp.	Rating PID 04	Code	No. of comp.	Rating 3PID 04	Code	No. of comp.
C+	1	2									
B–	2	4	C+, B–	1	6	One star	1	4	One and two stars	1	11
B	3	5				Two stars	2	7			
B+	4	11	B, B+	2	16	Three stars	3	5	Three stars	2	5
A–	5	2				Four stars	4	3			
A	6	1	A–, A	3	3	Five stars	5	6	Four and five stars	3	9
Total		25			25			25			25

Source: Gruszczyński (2012)

Table 5.10 Structure of the sample for 2005

Rating PFCG 05	Code	No. of comp.	Rating 3PFCG 05	Code	No. of comp.	Rating PID 05	Code	No. of comp.	Rating 3PID 05	Code	No. of comp.
C+	1										
B–	2	4	B–, B	1	18	One star	1	6	One and two stars	1	16
B	3	14				Two stars	2	10			
B+	4	10	B+	2	10	Three stars	3	8	Three stars	2	8
A–	5	5				Four stars	4	7			
A	6	1	A–, A	3	6	Five stars	5	3	Four and five stars	3	10
Total		34			34			34			34

Source: Gruszczyński (2012)

Associations of Ratings and Financial Indicators

Simple correlation coefficients between the numerical codes of the endogenous variables (PFCG04, 3PFCG04, PID04, 3PID04, PFCG05, 3PFCG05, PID05, and 3PID05) and the relevant financial ratios indicated the following possible associations (the + or − sign indicates the direction, double sign indicates the association is significant at the level of 10%):

2004:

Endogenous variable	Potential explanatory variables
PFCG04	L1(+), L2(+), L3(+), A2(−−), A3(−−), A4(−−), A6(−−), D2(−)
3PFCG04	ROE(+), A3(−−), A4(−−), A6(−−), D4(+)
PID04	ROE(+), A1(−−), A5(+)
3PID04	P1(+), A1(−), D4(+)

2005:

Endogenous variable	Potential explanatory variables
PFCG05	P2(−), A1(−), A3(−−), D1(−)
3PFCG05	L2(−), A1(−−), D1(−−), D2(++)
PID05	A1(−−), A3(−−), D4(−)
3PID05	P3(−), P4(−), A1(−−), A3(−−), A4(−−), A6(−−), A7 (−−)

It is hardly possible to indicate any pattern of relationship between the financial indicators and the rating variables (PFCG, 3PFCG, PID, and 3PID). Moreover, some of the directions suggested by the correlation signs seem to be counterintuitive.

This outcome is due to the small sample sizes, as well as to the heterogeneity of the samples. The samples include companies from diverse branches in the manufacturing and service sectors, companies of various sizes and from a broad range of activity profiles. The elements influencing such non-homogeneity of the sample are so many that, given the sample size, it is not possible to account for them in properly extracting the relationship under study.

If one insists on spelling out the association “disclosed by the data,” it can be said that—with a significant degree of confidence—the ratios A1, A3, and A4 are negatively correlated with the corporate governance rating of listed companies in Poland for 2004 and 2005. Perhaps, this might indicate that companies with a higher level of receivables turnover and liabilities turnover are regarded as inferior from the standpoint of corporate governance.

Due to such meager results from the introductory data analysis, we present below only the examples of the estimated ordered logit models. The models include at most two explanatory variables. The τ_j are thresholds (cutpoints; see Sect. 2.4) indicating the intervals for categories expressed in real numbers: from the first to the fifth

Table 5.11 The ordered logit estimation results for the PFCG and PID ratings for 2004: variables PFCG04 and PID2004 (six and five rating categories)

Endogenous variable→	PFCG04		PFCG04		PID04		PID04	
Explanatory variable↓	Parameter estimate	p	Parameter estimate	p	Parameter estimate	p	Parameter estimate	p
ROE					5.109	0.07	5.245	0.09
L1	0.973	0.05						
A1					−0.019	0.07		
A4			−0.018	0.01				
A5					0.018	0.10	0.017	0.13
A6	−0.016	0.01						
τ_1	−3.725	0.02	−5.157	0.00	−2.451	0.02	−0.957	0.16
τ_2	−2.018	0.15	−3.192	0.00	−0.655	0.48	0.686	0.28
τ_3	−0.912	0.49	−2.121	0.01	0.293	0.74	1.577	0.02
τ_4	1.746	0.23	0.399	0.62	1.005	0.28	2.212	0.00
τ_5	3.035	0.07	1.650	0.15				
Log likelihood	−33.047		−32.974		−34.857		−36.603	
Forecast accuracy (%)	48		44		40		28	

Source: Gruszczyński (2012)

(or sixth) for the model with five (or six) rating categories, and from the first to the third for the model with three rating categories. Due to the multicollinearity of the explanatory variables and the low level of their association with the endogenous variables, the models have poor statistical characteristics.

Tables 5.11 and 5.12 present the estimation results for the representative models constructed for 2004.

Tables 5.13 and 5.14 present the estimation results for the representative models constructed for 2005.

The results presented in Tables 5.11–5.14 may be commented on keeping in mind the details for constructing, estimating, and interpreting the multinomial ordered variables model presented in Sect. 2.4. The sign of the parameter estimate by a variable is exactly the same as the sign of marginal effect of this variable on the probability that the dependent variable receives a higher category (i.e., a higher rating). Estimates of the cutpoints τ_j are employed in calculating the probabilities of a company having a specific rating—probabilities estimated from the model. This is then applied in the calculation of forecast accuracy. As expected, this accuracy is low.

The results of this attempt to determine the relationship between the financial performance of companies and their ratings are mixed and rather disappointing. The correlations between the ratings and performance indicated significant negative association for the activity ratios. Significant correlations between rating level and the financial indicators are scarce and sometimes counterintuitive. Therefore, the

Table 5.12 The ordered logit estimation results for the PFCG and PID ratings for 2004: variables 3PFCG04 and 3PID2004 (three rating categories)

Endogenous variable→	3PFCG04		3PFCG04		3PID04		3PID04	
Explanatory variable↓	Parameter estimate	p	Parameter estimate	p	Parameter estimate	p	Parameter estimate	p
P1					2.932	0.33		
ROE	0.949	0.11						
A4	−0.016	0.05					−0.012	0.24
A6			−0.012	0.07				
D4			0.116	0.24	0.112	0.30		
τ_1	−4.949	0.11	−2.861	0.03	0.944	0.39	−1.120	0.17
τ_2	0.016	0.05	0.796	0.48	1.824	0.10	−0.261	0.74
Log likelihood	−18.432		−19.506		−25.027		−25.509	
Forecast accuracy (%)	72		72		52		44	

Source: Gruszczyński (2012)

Table 5.13 The ordered logit estimation results for the PFCG and PID ratings for 2005: variables PFCG05 and PID2005 (six and five rating categories)

Endogenous variable→	PFCG05		PFCG05		PID05		PID05	
Explanatory variable↓	Parameter estimate	p	Parameter estimate	p	Parameter estimate	p	Parameter estimate	p
A1					−0.015	0.07		
A3	−0.010	0.02	−0.006	0.09	−0.004	0.36	−0.006	0.13
D1	−1.271	0.03						
τ_1	−5.311	0.00	−2.960	0.00	−3.356	0.00	−2.378	0.00
τ_2	−2.717	0.01	−0.736	0.25	−1.921	0.02	−1.055	0.10
τ_3	−0.985	0.35	0.773	0.26	−0.647	0.39	0.120	0.85
τ_4	0.910	0.51	2.582	0.02	1.270	0.18	0.955	0.02
Log likelihood	−37.928		−40.873		−44.223		−45.972	
Forecast accuracy (%)	42		39		48		26	

Source: Gruszczyński (2012)

effort to construct ordered logit models for explaining the rating level was at the outset designed as only a methodological example. Nevertheless, the simple representative models which have been estimated, have reasonable interpretation and, as can be expected, have rather poor ex post forecast accuracy. It turned out that the models for the PID ratings were harder to specify and have worse fit than the models for the PFCG ratings. Such weak results are mainly due to the small sample sizes and the heterogeneity of the samples (samples include companies from diverse branches, of various sizes, etc.). On the other hand, the corporate governance rating is based on

Table 5.14 Ordered logit estimation results for PFCG and PID ratings for 2005: variables 3PFCG05 and 3PID2005 (three rating categories)

Endogenous variable→	3PFCG05		3PFCG05		3PID05		3PID05	
Explanatory variable↓	Parameter estimate	p	Parameter estimate	p	Parameter estimate	p	Parameter estimate	p
A1	−0.014	0.09	−0.017	0.06	−0.013	0.19		
A3					−0.007	0.19	−0.010	0.07
D1			−1.716	0.04				
D2	3.859	0.07						
τ_1	0.719	0.54	−3.136	0.01	−2.197	0.02	−1.628	0.05
τ_2	2.371	0.06	−1.378	0.21	−0.925	0.29	−0.419	0.59
Log likelihood	−30.608		−28.952		−29.855		−30.658	
Forecast accuracy (%)	56		62		55		52	

Source: Gruszczyński (2012)

a homogenous set of criteria, applied to each company in the same manner, independently of the company's particular characteristics.

The study is an example of modeling the CG rating as a dependent variable related to regressors representing financial performance. Such research attempts are still infrequent. A similar study was presented by Lazarides and Drimpetas (2009) for Greece in which the dependent variable is binomial and represents two states of the CG index: "low" and "high."

This chapter presented topics regarding corporate governance, assessment of its level for companies, and quantitative research aimed at disclosing associations of corporate governance with firm performance and other firm characteristics. This area of financial microeconometrics is closely related to other areas that have been presented throughout the book.

While corporate governance issues are being debated globally, statistical–econometric research serves as just one tool in those debates. It may help in finding current solutions to questions posed by practitioners and analysts, but the issue of corporate governance is complex and constantly changing, in time and space. As a result, even the acclaimed indices that measure CG levels are subject to severe criticism and should only be used with great care.

References

Adams RB (2017) The ABCs of empirical corporate (governance) research. Corp Gov Int Rev 25:461–464

Adams RB, Hermalin BE, Weisbach MS (2009) The role of boards of directors in corporate governance: a conceptual framework & survey. European Corporate Governance Institute (ECGI) – Finance Working Paper No. 228/2009

Aguilera R, Florackis C, Kim H (2016) Advancing the corporate governance research agenda. Corp Gov 24:172–180

Aluchna M, Dzierżanowski M, Przybyłowski M, Zamojska-Adamczak A (2005) Analiza empiryczna relacji między strukturami nadzoru korporacyjnego a wskaźnikami ekonomicznymi i wyceną spółek notowanych na GPW. Instytut Badań nad Gospodarką Rynkową, Gdańsk

Armour J, Deakin S, Lele P, Siems M (2009) How do legal rules evolve? Evidence from a cross-country comparison of shareholder, creditor and worker protection. Am J Comp Law 57:579–629

Armstrong CS, Guay WR, Weber JP (2010) The role of information and financial reporting in corporate governance and debt contracting. J Account Econ 50(2–3):179–234

Bajra U, Cadez S (2018) The impact of corporate governance quality on earnings management: evidence from European companies cross-listed in the US. Aust Account Rev 28(2):152–166

Bebchuk LA, Cohen A, Ferrell A (2009) What matters in corporate governance? Rev Financ Stud 22(2):783–827

Bebchuk L, Cohen A, Wang CCY (2010) Learning and the disappearing association between governance and returns. NBER Working Paper No. w15912, Harvard Law School Olin Discussion Paper No. 667

Bebchuk L, Cohen A, Wang CCY (2013) Learning and the disappearing association between governance and returns. J Financ Econ 108(2013):323–348

Bebchuk LA, Brav A, Jiang W (2015) The long-term effects of shareholder activism. Columbia Law Rev 115(5):1085–1156

Beekes W, Brown PR, Zhan W, Zhang Q (2016) Corporate governance, companies disclosure practices and market transparency: a cross country study. J Bus Financ Acc 43(3–4):263–297

Bhagat S, Jefferis RH (2002) The econometrics of corporate governance studies. MIT Press, Cambridge, MA

Black BS, Jang H, Kim W (2002) Does corporate governance matter? Evidence from Korea. KDI School Working Paper No 02-04, Seul

Black B, de Carvalho AG, Sampaio JO (2014) The evolution of corporate governance in Brazil. Emerg Mark Rev 20:176–195

Black B, de Carvalho AG, Khanna V, Kim W, Yurtoglu B (2017) Corporate governance indices and construct validity. Corp Gov Int Rev 25:397–410

Bøhren Ø, Ødegaard BA (2006) Governance and performance revisited. In: Ali P, Gregouriu G (eds) International corporate governance after Sarbanes-Oxley. Wiley, Hoboken, NJ

Börsch-Supan A, Köke J (2000) An applied econometricians' view of empirical corporate governance studies. ZEW Discussion Paper No. 00-17, Mannheim

Boyd BK, Adams R, Gove S (2017) Research methodology of governance studies: challenges and opportunities. Corp Gov 25(6):381–464

Brickley JA, Zimmerman JL (2010) Corporate governance myths: comments on Armstrong, Guay, and Weber. J Account Econ 50:235–245

Brown LD, Caylor ML (2006) Corporate governance and firm valuation. J Account Public Policy 25:409–434

Brown P, Beekes W, Verhoeven P (2011) Corporate governance, accounting and finance: a review. Account Finance 51:96–172

Carvalhal da Silva AL, Leal RPC (2005) Corporate governance index, firm valuation and performance in Brazil. Revista Brasileira de Financas 3(1):1–18

Chong A, López-de-Silanes F (eds) (2007) Investor protection and corporate governance: firm-level evidence across Latin America, Latin American development forum. Inter-American Development Bank, Washington, DC

Dallas G, Pitt-Watson D (2016) Corporate governance policy in the European Union: through an investor's lens. CFA Institute, Charlottesville, VA

Drobetz W, Schillhofer A, Zimmermann H (2003) Corporate governance and expected stock returns: the base of Germany, Working Paper, University of Basel

Dzierżanowski M, Przybyłowski M, Tamowicz P (2004) A small step forward: corporate governance rating 2003 for Polish listed companies. Instytut Badań nad Gospodarką Rynkową, Gdańsk

Dzierżanowski M, Milewski G, Przybyłowski M, Tamowicz P (2006) Postęp, ale nie rewolucja. Ranking nadzoru korporacyjnego 2005. Polskie Forum Corporate Governance. Rzeczpospolita, 6 February 2006

Eulerich M, Haustein S, Zipfel S, Van Uum C (2013) The publication landscape of German corporate governance research: a bibliometric analysis. Corp Ownersh Control 10(2):661–673

Filatotchev I, Boyd BK (2009) Taking stock of corporate governance while looking to the future. Corp Gov 17:257–265

Filatotchev I, Wright M (2017) Methodological issues in governance research: an editor's perspective. Corp Gov Int Rev 25:454–460

Frederikslust R, Ang JS, Sudarsanam PS (eds) (2008) Corporate governance and corporate finance: a European perspective. Routlegde, New York

Garstka S (2009) Stosowanie dobrych praktyk przez spółki notowane na Giełdzie Papierów Wartościowych w Warszawie a wyniki ekonomiczne spółek. Analiza ekonometryczna [Observing good corporate governance practices by companies listed on the Warsaw Stock Exchange and their financial results. Econometric analysis]. Master thesis, SGH Warsaw School of Economics

Gillan SL (2006) Recent developments in corporate governance: an overview. J Corp Finan 12:381–402

Goergen M (2018) Corporate governance: a global perspective. Cengage Learning EMEA, Hampshire

Gompers P, Ishii J, Metrick A (2003) Corporate governance and equity prices. Q J Econ 118 (1):107–155

Gordon JN, Ringe W-G (eds) (2018) The Oxford handbook of corporate law and governance. Oxford University Press, Oxford

Gruszczyński M (2004) Financial distress of companies in Poland. Int Adv Econ Res 10(4): 249–256. Also available as Gruszczyński M (2004) Financial distress of companies in Poland. Working Paper No. 1-04, Department of Applied Econometrics, SGH Warsaw School of Economics

Gruszczyński M (2006) Corporate governance and financial performance of companies in Poland. Int Adv Econ Res 12(2): 251–259. Also available as Gruszczyński M (2005) Corporate governance and financial performance of companies in Poland. Working Paper No. 2-05, Department of Applied Econometrics, SGH Warsaw School of Economics

Gruszczyński M (2007) Corporate governance ratings and the performance of listed companies in Poland. Working Paper No. 4-07, Department of Applied Econometrics, SGH Warsaw School of Economics

Gruszczyński M (2012) Empiryczne finanse przedsiębiorstw. Mikroekonometria finansowa [Empirical corporate finance. Financial microeconometrics]. Difin, Warszawa

Gugler K (ed) (2001) Corporate governance and economic performance. Oxford University Press, Oxford

Haskovec N (2012) Codes of corporate governance: a review. Working Paper published by the Millstein Center for Corporate Governance and Performance, Yale School of Management

Kaplan SN, Minton BA (2012) How has CEO turnover changed? Int Rev Financ 12(1):57–87

Klapper L, Love I (2003) Corporate governance, investor protection and performance in emerging markets. J Corp Finan 195:1–26

Klausner M (2018) Empirical studies of corporate law and governance. In: Gordon JN, Ringe W-G (eds) The Oxford handbook of corporate law and governance. Oxford University Press, Oxford

Kowalewski O, Stetsyuk I, Talavera O (2007) Corporate governance and dividend policy in Poland. Wharton Financial Institutions Center Working Paper No. 07-09

La Porta R, Lopez-de-Silanes F, Shleifer A, Vishny RW (1998) Law and finance. J Polit Econ 106:1113–1155

La Porta R, Lopez-de-Silanes F, Shleifer A, Vishny RW (1999a) Investor protection and corporate governance. Available at SSRN: http://ssrn.com/abstract=183908
La Porta R, Lopez-de-Silanes F, Shleifer A, Vishny RW (1999b) Investor protection: origins, consequences, and reform. NBER Working Paper No. 7428
Larcker DF, Tayan B (2011) Seven myths of corporate governance. Rock Center for Corporate Governance at Stanford University closer look series: topics, issues and controversies in corporate governance. No. CGRP-16
Larcker DF, Tayan B (2015a) Corporate governance matters: a closer look at organizational choices and their consequences, 2nd edn. Pearson FT Press, Old Tappan, NJ
Larcker DF, Tayan B (2015b) Seven myths of boards of directors. Rock Center for Corporate Governance at Stanford University closer look series: topics, issues and controversies in corporate governance. No. CGRP-51
Lazarides T, Drimpetas E (2009) Evaluating corporate governance and identifying its formulating factors: the case of Greece. Corp Gov 11:136–148
Leblanc R (ed) (2016) The handbook of board governance: a comprehensive guide for public, private, and not-for-profit board members. Wiley, Hoboken, NJ
Lipton M (2019) It's time to adopt the new paradigm. Post on: Harvard Law School Forum on Corporate Governance and Financial Regulation
Martynova M, Renneboog L (2010) A corporate governance index: convergence and diversity of national corporate governance regulations. Center Discussion Paper Series No. 2010-17; TILEC Discussion Paper No. 2010-012. Available at SSRN: https://ssrn.com/abstract=1557627 or https://doi.org/10.2139/ssrn.1557627
OECD (2004) Corporate governance: a survey of OECD countries. Organisation for Economic Co-operation and Development (OECD), Paris
OECD (2015) G20/OECD principles of corporate governance. OECD report to G20 finance ministers and central Bank governors. Organisation for Economic Co-operation and Development (OECD), Paris
Osma BG, Noguer BG (2007) The effect of the board composition and its monitoring committees on earnings management: evidence from Spain. Corp Gov 15(6):1413–1428
Padgett C (2011) Corporate governance: theory and practice. Palgrave Finance, New York
Polski Instytut Dyrektorów (2006) Raport. III rating spółek giełdowych według oceny inwestorów instytucjonalnych
Ross SA, Westerfield RW, Jaffe J, Jordan B (2015) Corporate finance, 11th edn. McGraw Hill Irwin, New York
Shleifer A, Vishny R (1997) A survey of corporate governance. J Financ 52:737–775
Sila V, Gonzalez A, Hagendorff J (2016) Women on board: does boardroom gender diversity affect firm risk. J Corp Finan 36:26–53
Tamowicz P, Dzierżanowski M, Lepczyński B (2001) Rating nadzoru korporacyjnego w spółkach giełdowych. Instytut Badań nad Gospodarką Rynkową, Gdańsk
Thomsen S, Conyon M (2012) Corporate governance: mechanisms and systems. Mc-Graw Hill Education, Maidenhead
Varshney P, Kumar KV, Vasal VK (2012) Corporate governance index and firm performance: empirical evidence from India. Available in SSRN: http://ssrn.com/abstract=2103462
Wu JS, Zhang IX (2019) Mandatory IFRS adoption and the role of accounting earnings in CEO turnover. Contemp Account Res 36(1):168–197

Chapter 6
Topics in Empirical Corporate Finance and Accounting

This chapter presents a selection of additional research topics associated with corporate finance, corporate governance, and accounting. We begin with questions of how fundamental corporate information is translated into market returns and what is its relevance to company valuation. These are subjects of the "value relevance of companies' financial statements," "microeconometrics for equity valuation," and "fundamental strategies." The second part of this chapter presents several topics demonstrating other applications of microeconometric methodology: mergers and acquisitions, IPOs, and dividend payouts.

The selection in this chapter includes topics not mentioned in previous chapters that are also within the author's purview and which may be of interest to the reader. Most examples of research presented here have one common denominator: they are the outcome of the author's collaboration with young researchers at SGH Warsaw School of Economics who have worked on MA or PhD thesis under the author's supervision.

6.1 Value Relevance of Companies' Financial Statements

Microeconometric methodology is often employed in empirical corporate finance and applied accounting in research known as "value relevance" (of accounting numbers, of financial statements, of earnings, of book values, etc.). It is naturally the object of fundamental analysis. Typical regression models for comparative valuation relate price or return ratios ("multiples") to a selection of fundamental variables rooted in company financials (Damodaran 2012). The seminal models of Fama and French (1993, 2015) also concentrate on finding associations between returns and company financials. This section develops some thoughts expressed in Gruszczyński et al. (2016).

Valuation models use company financials in addition to market variables to explain price/return multiples. Value relevance models are targeted "only" at

M. Gruszczyński, *Financial Microeconometrics*,
https://doi.org/10.1007/978-3-030-34219-7_6

examining the association between a company's market performance and the results disclosed in its financial statements. Francis and Schipper (1999) distinguish between four interpretations of value relevance (VR). One interpretation common to most studies states that VR means an association or correlation between accounting information and stock returns and prices.

Modern research on value relevance dates to the 1990s. Collins et al. (1997) showed in their survey that earnings and book values have historically remained value relevant over the 40 years of their reported research. However, there was a shift of value relevance from earnings to book values which can be explained, for example, by an increasing frequency of negative earnings or changes in average firm size. A more recent survey by Beisland (2009) comments on more than 160 papers on this topic, most from the preceding two decades. His definition states that "accounting information is denoted as value relevant if there is a statistical association between the accounting numbers and market values of equity." The author distinguishes between the following research streams on value relevance: (1) the VR of earnings and other flow measures, (2) the VR of equity and other stock measures, (3) the VR over time, and (4) the VR of alternative accounting methods.

Numerous papers reporting value relevance research have appeared worldwide. A sample from recent years might include these articles:

- Clacher et al. (2013) on the VR of direct cash flow components since the adoption of IFRS in Australia.
- Elshandidy (2014) on the VR of accounting information from 1999 to 2012 in various segments of the Chinese stock market.
- Alali and Foote (2012) on the VR of accounting information under IFRS for companies listed on the Abu Dhabi Stock Exchange.
- Mulenga (2015) on the VR of bank financials for companies listed on the Bombay Stock Exchange.
- Sharif et al. (2015) on the VR of accounting ratios for companies listed on the Bahrain Stock Exchange.
- Zulu et al. (2017) on the VR of interim and annual financial statements for companies listed on the Johannesburg Stock Exchange.

The vast literature on VR directly applies or expands basic models like that of Ohlson (1995) in which a firm's market value is linearly related to its earnings per share and book value per share. Also, frequently applied is the approach by Easton and Harris (1991), in which stock return is related to the earnings level and the change in earnings over the previous period.

Increasingly, reports are also tackling new questions arising in VR research, both concerning the scope of analyses as well as methodological issues. For example, the latter topic is discussed in the paper on econometric methodology in VR research by Onali and Ginesti (2015), who argue that including the lag of stock price as an explanatory variable in estimating price level regressions significantly improves a model's performance. The increasing number of papers published in the VR field are evidence of the ongoing awareness of the need to constantly search for connections between company fundamentals and the financial market (see, e.g., the Value Relevance stream at www.academia.edu).

The three examples that follow present results for NYSE-listed companies and companies listed on the Warsaw Stock Exchange.

The NYSE

Example 6.1 Value Relevance Models for Earnings and Book Value for NYSE-Listed Companies

Keener (2011) estimated a number of simple regression models with the dependent variable being the share price of the i-th company in the t-th year (P_{it}) and explanatory variables including earnings per share (E_{it}) and book value per share (BV_{it}). This is the model of Collins et al. (1997) but using more recent data, specifically NYSE-listed company data for the period 1982–2001. The data sources were Compustat and CRSP. The model was estimated for each year separately, for all years combined, and for industries.

For example, the full model for all years had 98,284 observations with the regression equation

$$P_{it} = \alpha_0 + \alpha_1 E_{it} + \alpha_2 BV_{it} + \varepsilon_i \tag{6.1}$$

The α_1 estimate is 1.437 and the α_2 estimate is 0.766 (both significantly different from zero), and R-squared is 0.413. The same exercise performed for each year separately produced similar results: both explanatory variables were significant in the models, with positive values for both regression parameters.

This simple exercise confirms the association between fundamentals and market valuation; however, a more relevant methodology might include the panel regression approach and the testing of the functional form.

■

The Warsaw Stock Exchange (WSE)

We first comment on the results of Dobija and Klimczak (2010), who investigated companies listed on the WSE from 1994 to 2008. Their paper reports evidence of the relevance of earnings, although the association does not improve over time. In particular, the introduction of new accounting regulations in 2000 as well as the adoption of IFRS in Poland in 2005 did not strengthen this association. Methodologically, the authors use the approach by Easton and Harris (1991). The two results for WSE-listed companies that follow are presented in greater detail.

Example 6.2 Use of Principal Component Analysis in VR Research

The study by Kubik-Kwiatkowska (2013) aimed to empirically investigate the value relevance of the financial reports of companies listed on the WSE.[1] The database consisted of the consolidated and audited annual reports of 440 listed companies in the period 2000–2010 (11 years). Additionally, the analysis of quarterly reports was based on the unaudited financial reporting of 364 firms from Q1 2001 through Q2 2011 (42 quarters). The database has a large amount of missing data in each of the 204 categories in the financial reports, because companies often did not publish full reports. The database contained the following reports: balance sheet, income statement, and cash flow statement.

The samples were randomly divided into two sets: companies to build the model (training samples) and companies whose data were used to validate the models (holdout samples). Financial reports were used to build the model for the years 2000–2006 and from Q1 2001 through Q3 2009 (training periods), while the holdout periods covered the reports from 2007 to 2010 and from Q4 2009 to Q2 2011.

The dependent variable in all models is the measure of company market valuation, which is defined as market capitalization divided by total assets. For some models, the variable is scaled by a market adjustment—i.e., the denominator is multiplied by the WIG index (a WSE index). All items included in the companies' financial statements were selected to configure the set of explanatory variables. The principal component analysis applied to these items resulted in several components representing specific financial areas. These components are considered as explanatory variables together with additional categories such as employment level. The models were estimated primarily by means of a random effects panel regression.

The results support the main hypothesis that there is a significant relationship between the measure of company value and information from the financial reports. Although the detailed lists of significant explanatory variables differ among the range of examined models, there are a few factors that are common to them all. For the models based on both annual and quarterly financial statements, a positive relationship is shown between the company valuation variable and accounting information such as equity, (net) profit, and income tax.

Thus, factors representing the typical Ohlson model such as profit and equity have significant association with a measure of the value of companies listed on the WSE. This had been expected after the earlier findings of Dobija and Klimczak (2010). Additional factors related to market value such as income tax have proven significant in all types of models examined in the study.

Considering the variety of empirical models, it can be summarized that the following four factors are relevant to company valuation: profit, equity, tax, and company size (measured as the logarithm of asset values). The additional models have also proven that a factor called "industry" has a relationship with the value of companies.

[1] Gruszczyński et al. (2016).

For models based on annual financial reports, it was shown that the best type of model was one in which the dependent variable was a scaled value of a company with correction for the market sentiment (the WIG index). This shows that the relationships between information from the firms' financial statements and the share prices of these companies improve when market sentiment measured by a stock exchange index is taken into account.

■

Example 6.3 E/P as the Dependent Variable in VR Research
The study by Bilicz (2015) is focused on verifying whether selected financial variables are associated with the E/P ratio.[2] Data were extracted from the quarterly financial statements of WSE-listed companies.

As regards the time frame, the first quarterly reports used in this research are from Q1 2005. Poland joined the European Union on May 01, 2004, and the integration process triggered adjustments in Polish law and regulations so that the system would conform with that of the EU. The last data are from Q4 2013. The entire period 2005–2013 was split into three subperiods with supposedly different determinants that might influence the dependent variable. From the start of 2005 until mid-2007, prices on the Polish stock exchange rose rapidly (boom phase). This was followed by a downturn in the economy, which eventually ended in Q1 2009 (*downturn* phase). Then, from 2009 to 2013 the market remained flat, not recovering as did the USA and Western Europe (*stagnation* phase).

As some WSE-listed companies became targets for speculators mainly due to their low liquidity, only firms that were included in one of the three WSE indices—WIG20, WIG40, or WIG80—in 2005 were considered for the sample. Companies in the financial sector were excluded, as were firms that were delisted from the stock exchange or went bankrupt during the time frame. The final dataset comprised 80 firms.

Following the work of Huang et al. (2007), the price-earnings (P/E) ratio was initially selected as the dependent variable. However, this idea was later abandoned as there are two main drawbacks with this indicator. Under certain circumstances, the P/E ratio can have negative values, and it can also be very unstable especially when earnings for a specific period are close to zero. The most common solution is to exclude all such observations; however, this can lead to a potential bias and to an unbalanced panel. Another solution suggested by Damodaran (2001) is to use the inverse P/E ratio, which can be interpreted even if the values are negative. This approach was implemented and finally the dependent variable was set as the *trailing* P/E ratio and then inverted, resulting in what is called *earnings yield* (E/P). The same variable was used in a study by Penman and Zhang (2006).

In total, there are eight independent variables used in the study: return on equity, return on sales, book value, debt ratio, cash flow index, company size (capitalization), 10-year Polish bond yield, and the dividend ratio.

[2] Gruszczyński et al. (2016).

Four different econometric methods were applied: pooled regression, the fixed effects estimator, the random effects estimator, and the Blundell–Bond estimator.[3] For the pooled regression, the results of the Breusch–Pagan test and the Wald test revealed that individual specific effects were statistically significant, therefore, the pooled regression was not advocated. The choice between fixed and random effects estimators for a specific phase was supported by the Hausman test. The null hypothesis in the test was rejected only for the *stagnation* phase and this was a strong argument for using the random effects estimator for the first two phases and the fixed effect estimator for the last. For the dynamic version, the Blundell–Bond estimator was applied with two lags of the dependent variable. Use of the lags can be justified by the work of Campbell and Shiller (1988). After the series of tests and examining the statistically significant variables, it was shown that panel models provided better results than the standard pooled OLS regression.

It was also hypothesized that, in the *boom* and *downturn* phases, the fundamental variables were not the best suited to describe the market situation. This is due to the presumption that herd behavior and other psychological aspects may impact prices during those phases. To confirm this hypothesis, it was assumed that both lags of the dependent variable in the Blundell–Bond method should be statistically significant and with a positive estimate of the coefficient. Also, the number of other relevant fundamental variables, in the *boom* and *downturn*, should be relatively small compared with the *stagnation* phase.

For the *boom* phase, one lag of the dependent variable, the bond yield, the dividend ratio, the debt ratio, cash flow, and book value were statistically significant. For the *downturn* phase, only the lags, book value, and size were significant and for *stagnation* both lags (the second was negatively correlated with E/P), ROE, the bond yield, the dividend ratio, and the size of the company were relevant.

Comparing the results with the outcomes of popular models such as the Fed, the Fama-French three-factor model, and the Gordon model, it can be stated that strong empirical evidence was found in data for the WSE companies only for the first model. Bond yield has a positive association with the dependent variable in both *boom* and *stagnation* and during the *downturn* phase it can be said that statistical significance was not expected.

■

6.2 Microeconometrics for Equity Valuation

Experience in value relevance research leads directly to questions of equity valuation. Fundamental analysis of securities investigates the relationships between the fundamentals (i.e., solid data regarding a company's financial and strategic standing)

[3]For the application of panel models in corporate finance, see Flannery and Hankins (2013).

and the market value of the company. This is inherent in VR analyses. Here we need more specific relevance to share prices.

Fundamental Analysis

Penman (2010) presents fundamental analysis as a process with the following stages:

1. Knowing the business
   - The product
   - The knowledge base
   - The competition
   - The regulatory constraints

   The effect: Formulating the strategy
2. Analyzing information
   - In financial statements
   - Outside of financial statements
3. Forecasting payoffs
   - Measuring value added
   - Forecasting value added
4. Convert forecast to a valuation
5. Trading on a valuation
   - Outside investor: Compare value with price to BUY, SELL, or HOLD
   - Inside investor: Compare value with cost to ACCEPT or REJECT strategy.

This process of fundamental analysis may accommodate quantitative methodology, especially in analytical stages 2, 3, and 4 where relevant methods are applied to make judgements and forecasts.

The valuation of common stock is the topic of all corporate finance textbooks (e.g., Brigham and Daves 2019). Several books have been devoted to valuations only. Especially important are books by Damodaran (2012, 2014) and Penman (2010).

Relative Valuation

Although statistical-econometric methodology may be applied in various approaches to valuation, we concentrate here on methods that use datasets comprising information from many companies. It is a method of relative valuation, also called comparative valuation or valuation of comparables. The method can be applied if access is available to data on comparable companies like those from the same sector, industry, etc. Preferable are data on companies with instant market valuation—such as companies listed on stock exchanges.

The method of relative valuation involves calculating a market multiple like the P/E ratio (price/earnings ratio) for a given company and comparing it to the P/E of similar companies in the same market. If the P/E calculated for such group equals 15 and the earnings per share for the company in question is USD2, then the market valuation of one share is USD30. Such calculation seems very simple, but the relative valuation methodology is not that straightforward. Selection of comparable companies is always ambiguous. Also, characteristics of the company other than price and earnings perhaps should be considered.

Relative valuation is popular because of its simplicity (few assumptions compared to the methodology of discounted cash flow) and due to its intuitive timeliness (since it reflects current market conditions). On the other hand, methods based on multiples ignore many important features like growth potential and various risks, including the risk of over- or undervaluing due to erratic markets.

The drawbacks of relative valuation methods might be diminished by finding statistical associations between the valuation multiple and the attributes of companies in the comparison sample. The major multiples used in relative valuation are:

– *Earnings multiples* (e.g., P/E ratio = market price per share to earnings per share)
– *Book value multiples* (e.g., P/BV ratio = price to book value)
– *Revenue multiples* (e.g., P/S ratio = price to sales ratio).

Damodaran (2012) advocates the use of multiples for valuation by observing the following stages. Firstly, the multiple must have the same definition across the sample of comparable firms. The second (statistical) stage includes examining the industry- and market-wide distribution of the multiple (means, outliers, sample bias). The third (analytical) stage entails the analysis of the multiple with regard to fundamental factors and their association with the multiple. The fourth stage is the choice of comparable firms for the sample and eventually the application of regression analysis in order to include more factors as regressors explaining the multiple.

Thus, we finally have a microeconometric (regression) model with one of the multiples as a dependent variable.

Regression Models for Relative Valuation

On the webpage accompanying the Damodaran (2012) textbook on valuation, we find examples of regression models estimated in January 2011 for the purpose of multiple valuations. One of the models estimated for US data (about 6000 firms) is as follows:

$$\widehat{PE} = 6.37 + 83.55\, g_{EPS} + 5.83\, Payout + 5.06\, Beta \qquad (6.2)$$

where *PE* is the *P*/*E* ratio, g_{EPS} is the expected growth rate in *EPS* (earnings per share) for the next 5 years (analyst estimates), *Payout* = dividends/earnings, *Beta* = market beta (measure of market risk). For this model $R^2 = 0.198$. All variables are significant.

Similar regressions are given for other multiples like *PEG* = price/earnings-to-growth (*P*/*E* divided by expected growth rate in revenues); *PBV* = price to book value; *PS* = price to sales ratio; *EV*/*Invested capital* = Enterprise value/(book value of equity + book value of debt – cash); *EV*/*Sales* = Enterprise value/sales; *EV*/*EBITDA* = Enterprise value/(earnings before interest, taxes, depreciation, and amortization). Enterprise value (*EV*) is the market value of the business—more comprehensive than market capitalization.

The simplicity of relative valuation is obscured when we consider the arithmetic relationships of a multiple with the variables that characterize its numerator and denominator. For example, in theory the *P*/*E* ratio is equal to $P/E = s/(k–g)$, where *k* is the rate of return expected by shareholders, *g* is the growth rate, and *s* is the dividend rate. This is the standard dividend discount model. Therefore, we would expect that a higher dividend rate (*s*) is associated with a higher *P*/*E* ratio. Also, a higher rate of growth (*g*) should be linked to a higher *P*/*E*, and a higher *k* to a lower *P*/*E*. Variable *k* is commonly represented by a measure of risks such as the market *beta* or the volatility of returns. It might also be the risk measure generated from the company's financial statement (e.g., the debt ratio).

Moreover, the relationship between a multiple (dependent variable) and regressors might not be stable, the regressors may be highly intercorrelated (the multicollinearity problem), and the character of the relationship might be linear or not. Linear relations are commonly applied that might not be correct.

Example 6.4 Model for Comparative Valuation of Listed Companies in Poland
Stachurski (2009) attempted to estimate the regression of the *P*/*E* on the fundamentals of companies listed on the WSE. Variables representing historical growth rates (variable *g* in the notation above), risk (*beta*), and dividend rate (*s*) sometimes were shown to be unsatisfactory as regressors for the P/E (the signs of the estimates not as expected, unstable results for various sets of regressors, etc.). The inclusion of analyst forecasts did improve the results, although it reduced the number of observations. Table 6.1 presents one of the models.

The dependent variable is the "logarithm of *P*/*E* ratio." This was decided as the result of testing the model with the *P*/*E* ratio for nonlinearity. The explanatory variable *eps_2y_g_f* represents the expected growth rate for *EPS* (*earnings per share*). This is the geometric mean of the annual growth of *EPS* for the years 2008–2010 calculated with the data for Q1–Q4 2008 and with analyst forecasts of *EPS* growth for the years 2009–2010. The model has only two explanatory variables but has good statistical properties, with the expected signs of the parameter estimates. It can be used for comparative valuation with the *P*/*E* multiple.

■

Table 6.1 Regression of the *P*/*E* ratio on the growth rate and the risk factor for WSE-listed companies on 31.03.2009

Dependent variable: The log of the *P*/*E* for the company where $P =$ the closing share price on 31.03.2009 and $E =$ net earnings per share for Q1–Q4 2008.	
Explanatory variables	*Parameter estimates (p)*
eps_2y_g_f = the geometric average of the growth of earnings per share in 2008–2010 (expected *g* for *EPS*)	1.049∗∗∗
beta = the parameter beta from 52 weeks before 31.03.2009	−0.257∗∗∗
Constant	2.316∗∗∗

$n = 70$ companies listed on Warsaw Stock Exchange
$R^2 = 0.6114$; ∗∗∗ means $p < 0{,}01$
Source: Stachurski (2009)

Fundamental Strategies

Regression models applied in comparative valuation actually provide a "correction" of the valuation multiple for a single company in the sample (e.g., the correction of the *P*/*E*). Generally, *value relevance* models do not aim to find specific valuation but concentrate on revealing associations between company performance on the capital markets and company financial fundamentals disclosed in the accounting data.

Value relevance modeling may be viewed as belonging to the complex stream of research into fundamental analysis that began with Fama and French's seminal papers (Fama and French 1992, 1993). Their contribution lies in extending the classical CAPM (*capital asset pricing model*) by additional factors of fundamental origin. In the famous three-factor Fama-French model, the dependent variable is the expected rate of return of a portfolio (or of a single stock) and the explanatory variables are:

1. Market risk.
2. Return on a diversified portfolio of small stocks minus return on a diversified portfolio of big stocks (*SMB* = Small Minus Big).
3. Difference between the returns on diversified portfolios of high and low book-to-market-ratio stocks (*HML* = High Minus Low).

The model is linear and is usually presented as

$$R_{it} - R_{Ft} = a_i + b_i(R_{Mt} - R_{Ft}) + s_i SMB_t + h_i HML_t + \varepsilon_i \tag{6.3}$$

where R_{it} = the return on the security or the portfolio i for period t, R_{Ft} = the risk-free return, and R_{Mt} = the return on the value-weighted market portfolio.

The field of fundamental strategies has developed both research wise and practically to such an extent that it is beyond any reasonable survey in this book. We must, however, underscore the role of microeconometric modeling in all the attempts to sensibly apply the ideas of fundamental strategies in capital markets.

For example, in a more recent contribution by Fama and French (2015), the authors propose a 5-factor model that includes two new explanatory variables:

4. The difference between the returns on diversified portfolios of stocks with robust and with weak profitability (RMW = Robust Minus Weak).
5. The difference between the returns on diversified portfolios of low and high investment stocks (CMA = Conservative Minus Aggressive).

The complete 5-factor model is

$$R_{it} - R_{Ft} = a_i + b_i(R_{Mt} - R_{Ft}) + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + \varepsilon_i \tag{6.4}$$

The models were estimated for specially constructed portfolios of US stocks over 594 months (July 1963–December 2012). The results were mixed, especially with regard to testing portfolio efficiency by Gibbons et al. (1989).

Example 6.5 Returns Versus Fundamentals for WSE-Listed Companies
Witkowska (2006) estimated several microeconometric panel models relating stock returns to variables representing fundamental signals and risk factors for WSE-listed companies during 1999–2003. One of the models, with Fama and French factors, is presented in Table 6.2.

Table 6.2 shows parameter estimates with standard errors corrected for autocorrelation and heteroscedasticity (*PCSE*: *panel corrected standard errors*). The Fama and French factors are earnings to price ratio (variable *EP*), market value of equity (*marketequity*), and book-to-market equity (variable *BVMV*). From this set, only *EP* is significant in the model. From the fundamental factors, *grossmargin* and *expenses* are significant. All parameter estimates have the signs as theoretically expected.

■

Table 6.2 One-year ahead stock returns versus fundamentals and risk factors (Fama and French factors) for companies listed on the WSE during 1999–2003. Prais–Winsten estimation results with panel corrected standard errors

Dependent variable: *stockreturn* = one-year-ahead return = $price_{t+1+5/12}/price_{t+5/12}$; t = year		
Explanatory variables	*Estimates (p)*	*Expected sign*
grossmargin = Δ sales – Δ gross margin	−0.0912***	–
expenses = Δ sales and administrative expenses – Δ sales	−0.5618**	–
ROA = net income/total assets	0.8813	+
EP = earnings per share/market price per share	0.8563***	+
marketequity = number of outstanding shares * share $price_{t+5/12}$	−1.4e-08	–
BVMV = book value of firm's stock/market value	0.0029	+
Constant	1.2333***	

Δ = the percentage annual change in the variable from the average of prior 2 years
n = 187 WSE-listed companies; financial data from 1997–2003
$R^2 = 0.856$; *rho* (estimate of coefficient of autocorrelation ρ) = 0.0650;
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$
Source: Witkowska (2006)

6.3 Mergers and Acquisitions, IPOs, and Dividend Payouts

Mergers and Acquisitions

The final section of this chapter covers three additional topics of corporate finance and accounting that are studied with the use of microeconometrics.

Mergers and acquisitions (M&A) is the first example. As with many other topics in financial microeconometrics, it stands at the intersection of law, finance, accounting, management, and economics. An excellent set of papers devoted to research into M&A was edited by Eckbo (2010) and published in a two volume work that was mentioned in Chap. 1. As with fundamentals in the previous section, the scope of this book prevents us from delving into details or delivering a survey on research in the field of M&A.

In Chap. 2, we presented an example relating to mergers, specifically, on voting for a merger—with the use of a linear probability model (LPM).

Linear probability model Matvos and Ostrovsky (2008)	
Probability of voting for the merger from family cross ownership: $y = 1$ if the vote in the acquirer is "for", $y = 0$ otherwise ($n = 6369$ mutual fund votes in acquirer's shareholders meetings in 114 completed mergers and acquisitions of public US companies, 2003–2006).	X = holdings in the target (= 1 if the fund holds shares in the target, = 0 otherwise), family and cross (= 1 if the fund holds shares in the target and some other fund in the same family holds shares in the target as well, = 0 otherwise), family and no cross (= 1 if the fund does not hold shares but some other fund in the same family does, = 0 otherwise).

The study by Matvos and Ostrovsky (2008) aimed at determining the factors that are associated with voting for the merger of two companies by an investment fund (one of the owners of the acquirer). The estimated LPM revealed that if the fund holds shares "in the target," then the probability of voting "yes" (for the merger) is higher by 0.016 (1.6-percentage points) than the same probability in the case of other shareholders.

Example 6.6 Mergers and Acquisitions in the WSE
Lesisz (2004) considers WSE-listed companies called to sell shares. In 2000–2002, there were 85 such calls. After excluding calls from the financial and banking sector, calls on share buybacks, and repeated calls, etc., the final set consisted of 25 observations that represented calls for real control over companies by an outside investor. This selected group of companies was matched by a group of 37 listed companies that were not targeted for acquisition. Two estimated binomial logit models are shown in Table 6.3. The dependent variable Y is binomial with $Y = 1$ for companies being the target of acquisition and $Y = 1$ otherwise.

Values of the explanatory variables are calculated from the financial statements for the year preceding the acquisition year. The author of the study prefers model 2 over model 1 because both models tested equal (with the test on structure) and

Table 6.3 Binomial logit models of company acquisition on the Warsaw Stock Exchange in 2000–2002

Dependent variable: $Y = 1$ if the company is an acquisition target (to be acquired by another company), $Y = 0$ otherwise		
Explanatory variables	*Model 1 estimates (p)*	*Model 2 estimates (p)*
P/E = price/earnings ratio	0.035*	0.037**
$FCF2$ = operating cash flow/total assets	−0.680	–
$LIQ1$ = working capital/ total assets	−3.565**	−3.946**
$LEV1$ = long-term liabilities/equity	−7.393**	−7.611**
$AKT1$ = sales revenue/total assets	−0.585	–
Constant	1.192	0.508
Pseudo R-squared	0.173	0.164
Accuracy ex post	71%	64%

$n = 62$ WSE-listed companies, 2000–2002; *** $p < 0.01$; ** $p < 0.05$; *$p < 0.1$
Source: Lesisz (2004)

because all regressors are significant in model 2. Factors that are associated with a higher probability of company acquisition are a higher *P*/*E* ratio, low financial liquidity, and a low level of debt. However, only the last factor (*LEV*1) has the sign of association with *Y* as expected in theory. It means that the reality of mergers and acquisitions on the WSE during the period 2000–2002 differed from that on other markets.

■

Initial Public Offerings

IPO (*Initial Public Offering*) effects are the initial returns on the day of an IPO. The extant research on public offerings and the IPO effect is surveyed in two chapters of the *Handbook of Corporate Finance. Empirical Corporate Finance* (Eckbo 2007). The first chapter is devoted to security offerings (Eckbo et al. 2007) and the second to IPO underpricing (Ljungqvist 2007).

Example 6.7 IPOs on the Warsaw Stock Exchange
Futera (2005) followed 105 IPOs on the WSE during 1998–2004. The author's sample includes about 70% of the IPOs that occurred on the exchange during this period—other IPOs were eliminated for various reasons. The average rate of return on the day of the IPO was 7%. In 56 cases out of 105, the first-day price was higher than the issue price. Over a longer horizon (up to 2 years), the returns of IPO companies were shown to be lower than the return for the entire market. This has also been observed in other capital markets. The author estimated several linear regression models for returns as well as models for the signs of returns (linear probability models and binomial logit models). The types of models considered were as follows: the return on the IPO day, the return a week after the IPO, the returns after 1 month, 6 months, 1 year, and 2 years after the IPO.

Table 6.4 Returns on IPOs on the WSE during 1998–2004: estimation results for linear regression model, linear probability model (LPM), and binomial logit model

Dependent variable: Regression model: *DEB* = return on the IPO day (closing price on first day of quotations to issue price); LPM and logit: *SIGN_DEB* = 1 if *DEB* > 0 and = 0 otherwise			
Explanatory variables	*Regression model*	*LPM*	*Logit*
TATE	0.00389***	–	–
PZSK	0.11321**	–	–
KPKD	−0.02241*	–	–
RED	0.24422***	0.40878**	2.33516**
FTA	–	0.44665**	3.45440**
LNA	–	0.12938***	0.84400**
LNZ	–	−0.07238*	−0.46665*
Constant	−0.07931	−1.65470*	−14.0915**
R-squared	$R^2 = 0.26526$	$R^2 = 0.26364$	*pseudo* $R^2 = 0.2377$

$n = 105$ IPOs on WSE in 1998–2004
$***p < 0.01$, $**p < 0.05$; $*p < 0.1$
Source: Futera (2005)

Table 6.4 presents three models of return on the day of IPO. Explanatory variables are:

- *TATE*—Total assets/equity (*total assets to equity*).
- *PZSK*—Earnings increase: = 1 for an increase in net earnings during the fiscal year (or half year) before the IPO.
- *KPKD*—Proceeds from the issue/net assets before the issue.
- *RED*—Degree of reduction of orders for IPO shares from individual investors.
- *FTA*—Fixed assets/total assets (*fixed to total assets*).
- *LNA*—Natural log of total assets.
- *LNZ*—Natural log of average number of employees in the company.

It is especially important to explain directly the returns on two explanatory variables in the regression model: the degree of reduction of orders for IPO shares (*RED* variable) and total assets to equity (*TATE*). In both cases, higher values are associated with higher returns on the IPO day. The *RED* variable is also valid in binomial models: the higher the level of reduction, the higher the probability of a positive return on IPO day. In addition, the share of fixed assets in total assets and the size of total assets have a positive association with the probability of a positive IPO return.

■

Dividend Payout

Dividends are part of a firm's payout policy, which amounts to the return of capital by firms to the equity owners—through dividends and share repurchases. In a chapter of *Handbook of Corporate Finance. Empirical Corporate Finance* (Eckbo

2008), Kalay and Lemmon (2008) present a survey of theories and research on payout policy. In the summary, the authors state that

> [t]he modern study of payout policy is rooted in the irrelevance propositions developed by Nobel Laureates Merton Miller and Franco Modigliani. Payout policy is irrelevant when capital markets are perfect, when there is no asymmetric information, and when the firm's investment policy is fixed. Relaxing these assumptions leads to a role for payout policy to control agency problems and convey information to investors.

Example 6.8 Dividends in European Companies
Żelaszczyk (2011) uses data collected by Damodaran on dividends paid by companies in Europe during 2009–2010. Data were taken from Damodaran's webpage in 2011.

Firstly, the author estimated a probit model for 4788 companies in 2009—with a dependent variable equal to 1 if a company actually paid a dividend for 2009, and equal to 0 otherwise. There were 2079 companies that paid a dividend in 2009 with the average dividend yield equal to 4.7%. Dividend yield is the ratio of a company's annual dividend (per share) to the share price. The model estimated for 2009 helps to find relevant explanatory variables for the model for 2010.

Next, the author estimated the Heckman selection model in order to find the determinants of the dividend yield for 2010. The Heckit procedure (described in Sect. 2.10) consists of estimating two special equations: a selection equation and an outcome equation.

Out of 4818 companies in 2010, there were 1979 companies that paid a dividend in 2010, with an average dividend yield equal to 3.9%. Table 6.5 presents estimation results of the Heckit method for companies in 2010.

The upper part of Table 6.5 presents the estimates of the outcome equation. The dependent variable is dividend yield (*dividendyield*) and the explanatory variables are:

– *correlationwithmarket*—The correlation of the stock return with the market portfolio return.
– *standarddeviationinstockprice*—The standard deviation in the stock price (in USD).
– *historicalgrowthinrevenueslast5y*—The rate of growth in revenues over the last five years.
– *pretaxoperatingmargin*—The pretax operating margin.

The second part of Table 6.5 contains estimates of the selection equation (a binomial probit model). The dependent variable is the dividend payment in 2010 (*dividendyesno* $=$ 1 if the company actually paid a dividend for 2010, $=$ 0 otherwise). The explanatory variables are:

– *marketcapinus*—Market capitalization (in USD)
– *totaldebtinus*—Total debt (in USD)
– *cash*—Cash in hand (in USD)
– *cashfirmvalue*—Ratio of cash in hand to firm value
– *effectivetaxrate*—Effective tax rate for the company

Table 6.5 Heckman selection model for dividends in European companies in 2010 (printout from Stata)

Heckman selection model			
(regression model with sample selection)	Number of obs	=	4295
	Censored obs	=	2839
	Uncensored obs	=	1456
	Wald chi2(4)	=	3122.31
Log likelihood = -517.4709	Prob > chi2	=	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
dividendyi~d						
correlatio~t	-.0229497	.0095835	-2.39	0.017	-.041733	-.0041664
standardde~e	.0355656	.0105778	3.36	0.001	.0148334	.0562978
historica~5y	-.0276207	.013662	-2.02	0.043	-.0543978	-.0008437
pretaxoper~n	-.1386489	.0025562	-54.24	0.000	-.143659	-.1336388
_cons	.0531652	.0073471	7.24	0.000	.0387652	.0675653
dividendye~o						
marketcapi~s	.0000239	4.52e-06	5.29	0.000	.0000151	.0000328
totaldebti~s	-2.52e-06	8.19e-07	-3.08	0.002	-4.13e-06	-9.16e-07
cash	.0000111	7.69e-06	1.44	0.150	-4.01e-06	.0000261
cashfirmva~e	.3654202	.1192902	3.06	0.002	.1316156	.5992248
effectivet~e	3.134764	.1270713	24.67	0.000	2.885708	3.383819
revenues	2.81e-06	3.64e-06	0.77	0.440	-4.32e-06	9.94e-06
netcapex	.0000602	.0000217	2.77	0.006	.0000176	.0001028
_cons	-1.029565	.0333071	-30.91	0.000	-1.094846	-.9642848
/athrho	-.1718725	.0486059	-3.54	0.000	-.2671384	-.0766066
/lnsigma	-2.644044	.019311	-136.92	0.000	-2.681893	-2.606195
rho	-.1701999	.0471979			-.2609601	-.0764571
sigma	.0710733	.0013725			.0684335	.0738149
lambda	-.0120967	.0034291			-.0188176	-.0053757

LR test of indep. eqns. (rho = 0): chi2(1) = 11.31 Prob > chi2 = 0.0008

Source: Żelaszczyk (2011)

– *revenues*—Revenues (in USD)
– *netcapex*—Net capital expenditures (in USD)

The variable *lambda* denotes the inverse Mills ratio (*IMR*, see Sect. 2.10) and *rho* is the correlation coefficient of random errors in the outcome and the selection equations (*rho* = *lambda*/*sigma*).

The Heckman selection model is designed to accommodate sample nonrandomness. In this example, the outcome equation and the selection equation have random errors correlated. It is reported in the chi-square test in the last row of Table 6.5.

The selection equation suggests that the probability of a dividend payment is positively associated with market capitalization, with the ratio of cash in hand to firm

value, with the effective tax rate for the company, and with net capital expenditures. It is negatively associated with debt. These signs are in accord with expectations. The outcome equation shows that the dividend yield is positively associated with the dispersion of the stock price. Other variables have negative signs of parameter estimates—not as expected. The author suggests that the weak correlation with the market returns might be explained but variables representing growth and revenues are perhaps not adequate for this particular model.
■

The topics presented in this chapter represent a supplement to the subjects of financial microeconometrics discussed in previous chapters. The selection includes only a few examples out of the many possibilities. The presented examples indicate the variety of methodological approaches, although the most popular is linear regression estimated for cross-section data. A common research problem still remains—finding relevant predictors (i.e., the explanatory variables for the models).

Capital markets use countless signals. The fundamentals represented by financial statements are not necessarily the most important. Therefore, the search for statistical–econometric relationships in corporate finance and accounting is both challenging and fascinating at the same time.

References

Alali FA, Foote PS (2012) The value relevance of international financial reporting standards: empirical evidence in an emerging market. Int J Account 47(1):85–108

Beisland LA (2009) A review of the value relevance literature. Open Bus J 2:7–27

Bilicz R (2015) Cena akcji spółek notowanych na Giełdzie Papierów Wartościowych w Warszawie a wyniki kwartalne – analiza ekonometryczna [Share prices of companies listed on Warsaw Stock Exchange and their quarterly financial results – econometric analysis]. Master thesis, SGH Warsaw School of Economics (unpublished)

Brigham EF, Daves PR (2019) Intermediate financial management, 13th edn. Cengage, Boston, MA

Campbell JY, Shiller RJ (1988) Stock price earnings and expected dividends. J Financ 43 (3):661–676

Clacher I, de Ricquebourg AD, Hodgson A (2013) The value relevance of direct cash flows under international financial reporting standards. Abacus 49(3):367–395

Collins DW, Maydew EL, Weiss IS (1997) Changes in the value relevance of earnings and book values over the past forty years. J Account Econ 24:39–67

Damodaran A (2001) The dark side of valuation: valuing old tech, new tech, and new economy companies. FT Press, Upper Saddle River, NJ

Damodaran A (2012) Investment valuation: tools and techniques for determining the value of any asset, 3rd edn. Wiley, Hoboken, NJ

Damodaran A (2014) Applied corporate finance, 4th edn. Wiley, Hoboken, NJ

Dobija D, Klimczak KM (2010) Development of accounting in Poland: market efficiency and the value relevance of reported earnings. Int J Account 45:356–374

Easton P, Harris T (1991) Earnings as an explanatory variable for returns. J Account Res 29:19–36

Eckbo BE (ed) (2007) Handbook of corporate finance: empirical corporate finance, North-Holland handbook of finance series, vol 1. Elsevier, Amsterdam

Eckbo BE (ed) (2008) Handbook of corporate finance: empirical corporate finance, North-Holland handbook of finance series, vol 2. Elsevier, Amsterdam

Eckbo BE (ed) (2010) Corporate takeovers: modern empirical developments, vol. 1: takeover activity, valuation estimates and sources of merger gains, vol. 2: bidding strategies, financing and corporate control. Elsevier/Academic Press, Amsterdam

Eckbo BE, Masulis RW, Norli Ø (2007) Security offerings. In: Eckbo BE (ed) Handbook of corporate finance: empirical corporate finance, North-Holland handbook of finance series, vol 1. Elsevier, Amsterdam, pp 233–374

Elshandidy T (2014) Value relevance of accounting information: evidence from an emerging market. Adv Account 30(1):176–186

Fama EF, French KR (1992) The cross-section of expected stock returns. J Financ 47:427–465

Fama EF, French KR (1993) Common risk factors in the returns on stocks and bonds. J Financ Econ 33:3–56

Fama EF, French KR (2015) A five-factor asset pricing model. J Financ Econ 116(1):1–22

Flannery MJ, Hankins KW (2013) Estimating dynamic panel models in corporate finance. J Corp Finan 19:1–19

Francis J, Schipper K (1999) Have financial statements lost their relevance? J Account Res 37 (2):319–352

Futera M (2005) Losy debiutów na Giełdzie Papierów Wartościowych w Warszawie w latach 1998–2004 w ujęciu ekonometrycznym [Econometric analysis of IPOs on Warsaw Stock Exchange in 1998-2004]. Master thesis, SGH Warsaw School of Economics

Gibbons MR, Ross SA, Shanken J (1989) A test of the efficiency of a given portfolio. Econometrica 57:1121–1152

Gruszczyński M, Bilicz R, Kubik-Kwiatkowska M, Pernach A (2016) Value relevance of companies' financial statements in Poland. Metody ilościowe w badaniach ekonomicznych/Quant Methods Econ XVII(4): 40–49. Also available as Gruszczyński M, Bilicz R, Kubik-Kwiatkowska M, Pernach A (2016) Value relevance of companies' financial statements in Poland. Working Papers 2016-014, SGH Warsaw School of Economics, Collegium of Economic Analysis

Huang Y, Tsai C, Chen CR (2007) Expected P/E, residual P/E and stock reversal: time-varying fundamentals or investor overreaction. Int J Bus Econ 6:11–28

Kalay A, Lemmon M (2008) Payout policy. In: Eckbo BE (ed) Handbook of corporate finance: empirical corporate finance, North-Holland handbook of finance series, vol 2. Elsevier, Amsterdam, pp 3–58

Keener MH (2011) The relative value relevance of earnings and book value across industries. J Financ Account 6:1. http://www.aabri.com/jfa.html

Kubik-Kwiatkowska M (2013) Znaczenie raportów finansowych dla wyceny spółek notowanych na Giełdzie Papierów Wartościowych w Warszawie S.A. [Value relevance of financial reports for the companies listed on Warsaw Stock Exchange]. Ph.D. dissertation, SGH Warsaw School of Economics (unpublished)

Lesisz J (2004) Model prawdopodobieństwa przejęcia firmy na Giełdzie Papierów Wartościowych w Warszawie SA [Modeling probability of company acquisition on Warsaw Stock Exchange]. Master thesis, SGH Warsaw School of Economics

Ljungqvist A (2007) IPO underpricing. In: Eckbo BE (ed) Handbook of corporate finance: empirical corporate finance, North-Holland handbook of finance series, vol 1. Elsevier, Amsterdam, pp 375–422

Matvos G, Ostrovsky M (2008) Cross-ownership, returns, and voting in mergers. J Financ Econ 89:391–403

Mulenga MJ (2015) Value relevance of accounting information of listed public sector banks in Bombay Stock exchange. Res J Finan Account 6(8):222–231

Ohlson JA (1995) Earnings, book values, and dividends in equity valuation. Contemp Account Res 11(2):661–687

Onali E, Ginesti G (2015) Sins of omission in value relevance empirical studies. MPRA Paper No. 64265. Posted 12 May 2015

Penman SH (2010) Financial statement analysis and security valuation, 4th edn. McGraw-Hill, New York

Penman SH, Zhang X-J (2006) Modeling sustainable earnings and P/E ratios with financial statement analysis. Working Paper, Columbia University

Sharif T, Purohit H, Pillai R (2015) Analysis of factors affecting share prices: the case of Bahrain Stock exchange. Int J Econ Financ 7(3):207–216

Stachurski K (2009) Metody ekonometryczne w wycenie porównawczej. Zastosowanie do wyceny spółek z Giełdy Papierów Wartościowych w Warszawie SA [Econometric models for comparative valuation. Application to valuation of companies listed on Warsaw Stock Exchange]. Master thesis, SGH Warsaw School of Economics

Witkowska M (2006) Fundamentals and stock returns on the Warsaw Stock Exchange. The application of panel data models. Working Paper no. 4-06, SGH Warsaw School Economics, Department of Applied Econometrics

Żelaszczyk M (2011) Problem poziomu dywidendy dla spółek europejskich w roku 2010 [Dividend yield for European companies in 2010], project report submitted for the course "Microeconometrics". Master Study, SGH Warsaw School of Economics

Zulu M, De Klerk M, Oberholster JGI (2017) A comparison of the value relevance of interim and annual financial statements. S Afr J Econ Manag Sci 20(1):1–11