Firm Strategies and Business Models in the Software Industry: A Configurational Approach

Vom Fachbereich Rechts- und Wirtschaftswissenschaften
der Technischen Universität Darmstadt
zur Erlangung des Grades eines
Doctor rerum politicarum (Dr. rer. pol.)
genehmigte

Dissertation

vorgelegt von
Dipl.-Wirtsch.-Inform. Anton Pussep
aus Novosibirsk

Erstgutachter: Prof. Dr. Peter Buxmann
Zweitgutachter: Prof. Dr. Dirk Schiereck

Tag der Einreichung: 23.11.2016
Tag der mündlichen Prüfung: 20.04.2017

Darmstadt 2017
Hochschulkennziffer D17

To my parents, reaching this point in my life
wouldn't have been possible without their unlimited support.

Contents

Abstract

Researchers have long focused on the determinants of firm success, which is of crucial interest to practitioners as well, since being successful is at the very heart of economic activity. Extant research emphasizes three levels of analysis at which determinants occur: firm, industry, and group level. Each level has been found to affect firm success. At group level, firms choose between a limited set of competitive approaches. The resulting groups are referred to as *configurations*. The analysis of configurations, their characteristics, and effects are the particular focus of configurational research and this thesis.

Along with a multitude of other concepts, scholars have used industry-specific conceptualizations of firm strategy to derive configurations, referred to as strategic groups. Despite theoretical and methodological weaknesses in its beginnings, strategic group research has overcome initial challenges and produced a strong body of theoretical argument, methodology, and empirical evidence in the tradition of configurational research.

More recently, business models have emerged as a topic of growing interest to researchers and practitioners. Though some methods from configurational research have been applied to business models, previous studies do not nearly grasp the full potential of configurational analysis. In addition to methodological shortcomings, business model research is still under criticism for theoretical and conceptual weaknesses.

This thesis uses the theoretical and methodological body of knowledge from strategic group research and applies it to *strategies and business models of software firms*. The particular case of the software industry is chosen because of its dynamics, size, growth, and importance to other industries. In order to improve our understanding of strategies and business models in the software industry, a software-specific value chain is derived and used as the main theoretical foundation to both concepts. Building upon detailed conceptualizations, three empirical studies are presented, each using a unique dataset to analyze the concepts at hand.

The empirical studies demonstrate the applicability of configurational analysis to software firms and provide insights into their characteristics and success factors. The results indicate that the most distinctive delineators of strategies and business models determine a firm's product and market scope, such as firm size, share of international revenues, and the number of targeted industries. Strong empirical evidence suggests that broader scope is associated with higher success in terms of higher performance, higher risk-adjusted performance, and in some cases lower risk. Being consistent with the economic properties of software products and markets, such as network effects, the findings bear rich implications for researchers and practitioners, including decision makers, investors, analysts, and policy makers.

List of figures

List of tables

List of acronyms

AMERICAS North, Central-, and South America

APJ Asia, Pacific, and Japan

APN average proportion of non-overlap

AROS inflation-adjusted return on sales

ASP application service provider

CAD software computer-aided design software

CAPM capital asset pricing model

CBRET cumulative beta excess returns

CfMoS cash flow margin on sales

CRET cumulative returns

CRSP The Center for Research in Security Prices

EMEA Europe, Middle East, and Africa

ERP software enterprise resource planning software

EUR Euro

GPL GNU General Public License

HAC hierarchical agglomerative clustering

ICT information and communication technology

IPO initial public offering

IS information systems

M&A mergers and acquisitions

NA not available

NACE Statistical Classification of Economic Activities in the European Community revision 2

OP operating profit

OPM operating profit margin

OSS open source software

PBR price / book ratio

PER price / earnings ratio

R&D research and development

ROA return on assets

ROE return on equity

ROIC return on invested capital

ROS return on sales

SaaS software-as-a-service

SBMF Software Business Model Framework

SIC Standard Industry Classification

SME small and medium enterprises

SWOT analysis "strengths-weaknesses-opportunities-threats" analysis

US United States

USD United States dollars

VIF variance inflation factor

Section 1

Introduction

Apple, Google, and Microsoft are among the top five most valuable firms by market capitalization (Statista, 2016). What makes these firms more successful than others? This turns out to be a deep question well beyond the pursuit of business success. The answer can help explaining how firms behave and why they choose certain strategies (Porter, 1991, p. 95). It thus comes as no surprise that the question what determines firm success or failure preoccupies a multitude of research fields, including strategic management (e.g., McNamara et al., 2003), information systems (e.g., Grover and Saeed, 2004), and software business (e.g., Schief et al., 2012).

After much controversy and discussion, scholars have emphasized three levels at which determinants of firm success operate: individual firm level, broad industry level, and intermediate group level (McGee and Thomas, 1986; Short et al., 2007). At individual level, firm-specific characteristics determine firm success, while at industry level, industry forces affect the success of member firms. At group level, firms can choose between a limited set of competitive approaches, some of which are more successful than others. For instance, software firms are often separated in two groups, suppliers of individual and standard software. While individual software suppliers develop solutions for specific needs, standard software suppliers target the mass market with standardized solutions (Buxmann et al., 2012, p. 16-17).

Scholars have used a great variety of concepts to capture groups of firms. These include classes, configurations, gestalts, modes, archetypes, strategic groups, competitive groups, and types (Short et al., 2008, p. 1054). For example, the well-known typology by Miles and Snow (1978) describes how strategic types differ in structure, technology, and decision process. Thus, by knowing the strategic type of a firm, inferences can be made regarding a broad set of firm characteristics. Scholars have also used strategic types to analyze further implications such as their impact on sourcing of information systems (Aubert et al., 2009).

In order to reduce the confusion stemming from different terminology, Short et al. (2008) coined the term *organizational configurations* as a cover term for more specific concepts at group level. Each concept falls in one of the following categories: generic strategies, organizational forms, strategic groups, and archetypes. The categories are aligned along the two dimensions "applicability of organizational configurations" (industry-specific or generic) and "primary basis for identifying configurations" (competitive strategy or other organizational features) as illustrated in Figure 1.1. According to this classification, strategic groups are industry-specific and are derived from variables assembling competitive strategy (e.g., Cool and Schendel, 1987). Archetypes are industry-specific and derived from variables that represent organizational features in a broader sense than compet-

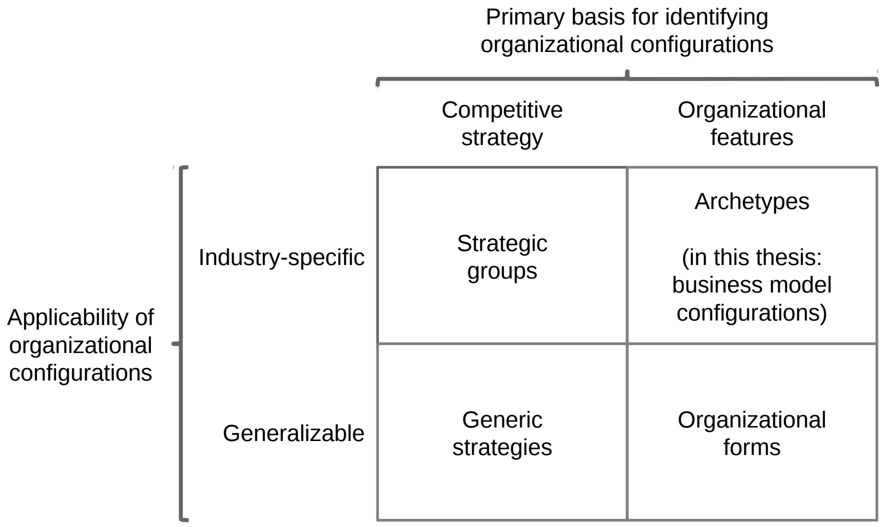

Figure 1.1: Classification of organizational configurations. (Short et al., 2008, p. 1057).

itive strategy (e.g., Bonaccorsi et al., 2006). Generic strategies are industry-independent and are derived on the basis of competitive strategy (e.g., Porter, 1980). Finally, organizational forms are industry-independent and derived from variables that represent organizational features (e.g., Zott and Amit, 2008).

The focus on configurations rather than individual firms is the main difference between conventional and configurational quantitative research. Conventional quantitative studies of organizations conceptualize a set of variables and use correlation techniques (e.g., multiple regression or structural equation modeling) to capture differences across all organizations (Short et al., 2008; Ragin, 2013). Whereas there is no doubt in the value of results derived with these methods, there are inherent shortcomings, some of which can be addressed with configurational analysis. For example, conventional quantitative research neglects interdependencies between variables, although an independent variable may have varying effects on a dependent variable in different contexts. Configurational analysis accounts for such relationships by creating different contexts for the analysis of variables. In each context, variables are regarded as interconnected elements and analyzed in concert, thus providing a systemic perspective (Fiss et al., 2013, p. 2).

Short et al. (2008, p. 1054) emphasize three particular goals of configurational research: description, explanation, and prediction. (1) Configurational research describes organizations by identifying sets of firms with unique characteristics. Methodologically, appropriate techniques are used, such as cluster analysis and Q-sort to provide a classification of organizations. As Ragin (2013) points out, there is often a limited diversity of configurations. The limitation itself is meaningful and rich in implications. (2) Configurational research explains organizational success by "identifying distinct, internally consistent sets of firms than by seeking to uncover relationships that hold across all organizations" (Ketchen et al., 1993, p. 1278). The premise is that some configurations are more successful than others. (3) Configurational research aims to predict which organizations will be more successful than others by extrapolating success factors.

In decades-long research, strategic group scholars have used configurational methods to study the group level of firm strategies (e.g., Porter, 1979; Mas-Ruiz and Ruiz-Moreno, 2011). In the mid-1990s, business models have emerged as a similar unit of analysis (Burkhart et al., 2011, p. 7–8). Scholars have since applied configurational methods to analyze the group level of the new concept. For instance, Malone et al. (2006) define a typology of sixteen business models along the two dimensions “What type of asset is involved?” and “What rights are being sold?”. They refer to configurations as a relevant research stream and use performance measures commonly used in configurational research. Valtakoski and Rönkkö (2010) use common methods from configurational analysis to derive a taxonomy of eight business models in the software industry.

Business models as well as strategies can be seen as representations of activities that firms perform to create value for their customers (Burkhart et al., 2011; Porter, 1991; Zott and Amit, 2010). Value creation, in turn, determines firm success, as customers are willing to reward firms that fulfil their needs better than others (Ketchen et al., 2007, p. 962). Both concepts are thus powerful units of analysis. This thesis uses configurational analysis to study the group level implications of these concepts. Arguably, deeper understanding can be obtained from industry-specific than generic analyzes, which is reflected in the recent shift to industry-specific inquiries in configurational research (Short et al., 2008, p. 1063). The configurational concepts of interest to this thesis are therefore strategic groups and archetypes in terms of Figure 1.1. However, the term business model configurations is preferred over archetypes, as the former is more specific about the underlying concept.

For the industry-specific analysis, the software industry is chosen because its powerful dynamics make it an outstanding and interesting subject. Software markets are shaped by disruptive technological trends such as big data, mobile computing, software-as-a-service (SaaS), and in-memory databases (Veit et al., 2014, p. 47). A unique system of economic properties such as network effects and high economies of scale (Buxmann et al., 2012) facilitates profound shifts in the software industry. Examples include the ascent of firms like Amazon, Google, and Facebook that rose to some of the most valuable firms on the planet. Such dynamics challenge the mental models of practitioners, who strive to understand the industry forces in order to respond to threats and to exploit opportunities. Rigorous empirical analysis can help to improve our understanding of the software industry beyond outdated statistics, beliefs, and industry fads.

1.1 Research questions

The overarching research goal of this thesis is to *improve our understanding of strategies and business models in the software industry.* It addresses the characteristics of software firms, their interrelationships, performance and risk implications, and tradeoffs between the different success dimensions through configurational analysis. This section breaks down the overall research goal into multiple research questions, which are then approached throughout the remainder of this thesis.

The first research question addresses the potential of configurational analysis to contribute to a deeper understanding of the software industry. It examines the concepts, theories, and methods for their applicability to software firm strategies and business models. Answering this question provides the conceptual, theoretical, and methodological foundation for the configurational analysis of the software industry and connects it to extant research.

Research Question 1 *How can configurational analysis of strategies and business models contribute to our understanding of the software industry?*

The second research question addresses the main theoretical foundation of firm strategies and business models: the value chain and its value creating activities. A software-specific value chain is methodically conceptualized that serves as a foundation to conceptualizations of software firm strategies and business models. Answering this question provides the theoretical framework for individual and conjoint analyzes of strategies and business models in the software industry.

Research Question 2 *Which activities are performed and combined by software firms to create competitive advantage?*

The third research question addresses the structure of the software industry in terms of prevailing strategies and business models. It builds upon the first two research questions that provide the necessary background for configurational analysis. Answering this question provides insights into the overall number of available competitive approaches, their distinguishing characteristics, and interrelationships between them. The characteristics are of particular interest because they indicate which adjustments need to be made to change to a different strategy or business model.

Research Question 3 *Which strategies and business models prevail in the software industry?*

The fourth research question addresses the success associated with different strategies and business models. Like the third research question, it builds upon the first two research questions. It also uses the findings from the third research question to examine how successful each configuration is in terms of different success dimensions, such as performance and risk. Answering this question helps to explain why some software firms perform better than others and provides guidance to practitioners in their decision-making. Possible conflicts along different success dimensions can help to explain why software firms choose different strategies and business models.

Research Question 4 *How successful are the various strategies and business models along different (and possibly conflicting) success dimensions in the software industry?*

1.2 Structure

This thesis combines confirmatory and exploratory research to answer the research questions presented in the previous section. Quantitative empirical methods play the major role in the course of analysis. However, qualitative methods are also used for a comprehensive literature review and advancement of concepts at hand.

The overall structure, as shown in Figure 1.2, is such that Section 2 and Section 3 lay the theoretical and conceptual foundations for configurational analyzes of strategies and business models in the software industry, focusing on Research Question 1 and 2. Section 4, Section 5, and Section 6 present the results of three quantitative studies, each addressing Research Question 3 and 4. Section 7 discusses the findings and Section 8 concludes this thesis. Each section is briefly summarized in the following paragraphs.

Section 2 provides an overview of extant research relevant to this thesis and addresses Research Question 1 regarding the potential of configurational analysis to improve our understanding of the software industry. The following topics are covered: (1) theoretical foundations of the software industry, (2) determinants of firm success, (3) strategic groups, and (4) business model configurations. For the software industry, Section 2.1 addresses the main economic properties in terms of software markets and software products. These economic properties help to understand what

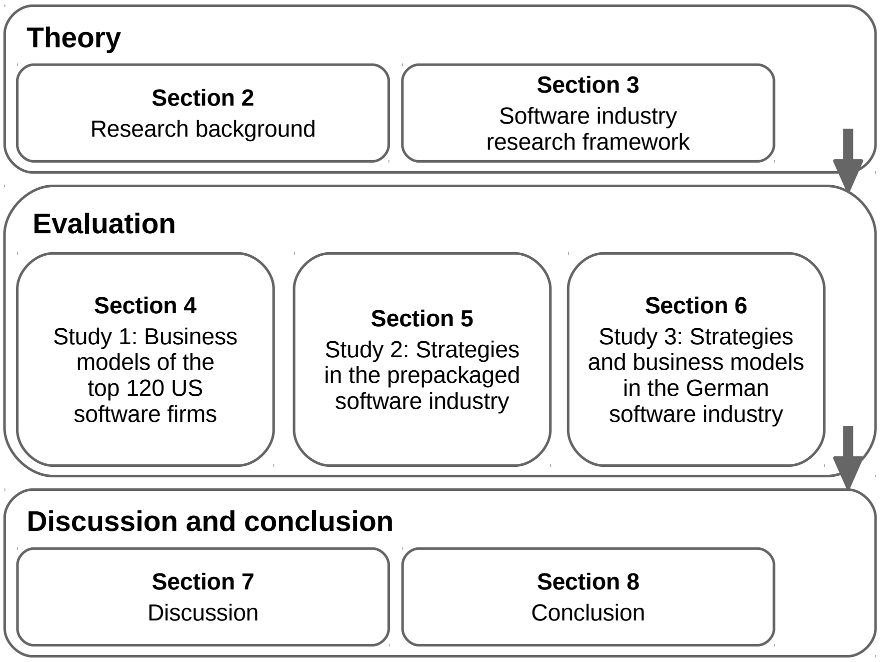

Figure 1.2: Thesis structure.

makes the software industry an interesting subject. They are also important to understand the forces affecting software firms. Section 2.2 addresses firm characteristics impacting performance, such as vertical integration, and introduces the concept of firm success as a multi-dimensional concept. The measurement of firm characteristics and success is discussed. Section 2.3 introduces the strategic group concept, its theoretical foundations, and rich methodology in the field. Section 2.4 introduces the business model concept and provides a comprehensive literature review of extant configurational research in that area. Comparisons are drawn with strategic group research to point out possibilities to develop business model research further. Section 2.5 concludes the literature reviews and develops testable hypotheses that help to answer the proposed research questions.

Section 3 addresses Research Question 3 regarding value creating activities of software firms, thereby establishing a common theoretical framework for conceptualizations of strategies and business models of software firms. It presents a software-specific value chain that is applicable to both concepts and allows for comparisons between them. The framework is based upon a literature review, an expert survey, and empirical proof of concept. As a result, a generic software value chain is presented with detailed activity descriptions. The value chain is used for analyzes in the subsequent studies.

Sections 4–6 address Research Question 3 and 4 on different datasets, because each covers desirable data characteristics which could not be covered by a single dataset:

- Section 4 presents a quantitative study of business model configurations. The sample consists of the 120 largest public software firms in the United States (US). The data is obtained from annual firm reports and financial databases. It is the only study in this thesis that uses expert classification to measure firm business models. Beyond the empirical results, this study confirms the general applicability of configurational analysis to business models in the software industry.

- Section 5 presents a quantitative study of strategic groups. The sample consists of public firms in the prepackaged software industry for which data is obtained from a financial database. This study is unique in this thesis regarding a multitude of periods and a narrowly defined, but large sample. These features allow for the most advanced configurational analysis. Moreover, this is the first study to provide a detailed analysis of strategic groups in the software industry.

- Section 6 presents a quantitative study of business model configurations and strategic groups. The sample consists of German software firms for which data was obtained via a public survey. In comparison to the previous two studies, it addresses the German software industry and includes firms of all sizes, including very small firms. It is further different because it uses primary data for the measurement of all concepts and includes a large set of strategy and business model characteristics. This study is particularly interesting because it analyzes strategic groups and business model configurations in concert on the same dataset, thus allowing for the most accurate comparison of results between the two concepts.

Section 7 summarizes the findings to the research questions, outlines the implications for researchers as well as practitioners, and acknowledges the limitations. Section 8 concludes this thesis.

1.3 Contributions

This thesis contributes to multiple research streams including software business, business model, strategic group, and configurational research. The findings further provide guidance to practitioners, such as decision makers in software firms, investors, analysts, and policy makers.

Software business research benefits from this first detailed configurational analysis of the software industry. Observed configurations describe existing strategies and business models, which allow for conclusions why certain competitive approaches prevail while others do not. Their characteristics are indicators of success factors in the software industry, thus helping to explain and possibly predict how a software firm will attempt to modify its competitive approach in order to become more successful. On the other hand, results on conflicting dimensions of firm success can help explaining why firms deliberately choose different strategies. Given a relevant classification of software firms, future studies can use the classification to analyze additional effects beyond the scope of this thesis.

Business model research benefits from the comparison with configurational and strategic group research in particular. The advanced theoretical and methodological body of knowledge from strategic group research is transferred and shown to be applicable for the analysis of business models. The application to software firms in this thesis advances business model research in multiple ways. First, it clearly separates the configurational from the conventional approach. Second, it expands the theoretical foundation of the concept with established theories from strategic group research. Third, it expands the business model methodology with rigorous methods which are well established in configurational analysis. Finally, the most comprehensive quantitative analysis of business model configurations in the software industry is provided by evaluating a large array of business model characteristics on different datasets.

Strategic group research benefits from detailed analyzes of the software industry, comparisons with previous findings on other industries, and inclusion of the rarely used success dimensions risk and risk-adjusted performance. Configurational research in general benefits as two particular organizational configurations are analyzed and compared.

Practitioners will find the demarcation of the software industry into clearly distinguishable configurations particularly useful. Decision makers in startups and established firms alike can use the prototypical configurations as templates to design their own strategies and business models. The decision making can be guided by the empirical results regarding the effects on firm success. When transitioning from one strategy or business model to another, decision makers are provided with an overview which key characteristics need to be modified and which interrelationships exist. Novel, potentially disruptive strategies and business models can be identified if they do not conform to any prevalent configuration.

Investors, analysts, and stakeholders in general are given theoretically and empirically founded classifications of software firms. These classifications can be used to compare firms, estimate their potential, and valuate them. Thus, stakeholders are given a scheme to guide their decisions which firms to be involved in, while differentiating between the potential performance and risk associated with the firms.

Section 2

Research background

This section reviews extant literature on the core concepts, theories, methods, and empirical results relevant to this thesis. Section 2.1 outlines the economic properties of software products and markets that indicate the forces working within the software industry and make it a unique subject. Section 2.2 provides an overview of extant research on the determinants of firm success, dimensions of success, and measurement issues. Section 2.3 reviews the literature on strategic groups with a particular focus on the methodology. Section 2.4 reviews the literature on business model conceptualizations, extant configurational research in that area, and compares it to strategic group research to highlight potential for improvement. Finally, Section 2.5 concludes this section by relating the findings to the research questions discussed in Section 1.1 and stating testable hypotheses that can be derived from extant research and tested in the subsequent studies.

2.1 Software industry

The focus of this thesis is on organizations in the software industry, i.e., software firms. A *software firm* is any organization whose core competence includes the creation of software products. This includes software creation in its narrow sense, such as the development of standard or custom software, as well as related service activities (Buxmann et al., 2012, p. 8-9). In a similar vain, if not said otherwise, the term *software product* is used to denote standard software, custom software, or related services. For instance, a prepackaged antivirus software, an internally developed online shop, or customizing activities on an SAP installation are all referred to as software products. The set of all software firms is defined as the *software industry*.

From a practitioner's point of view, the relevance of the software industry can be seen from its size, growth, and importance to other industries. Lovelock et al. (2015) estimate worldwide IT spending to US$ 3,710 billion in 2014, growing by 9.1% since 2010. Though actual figures can very depending on the firms included in the sample, it becomes clear that the software industry is of remarkable size and considerable growth. Also, by looking at individual firms, the most valuable and fastest growing firms are software firms, such as Apple, Google, and Microsoft (PricewaterhouseCoopers, 2014, p. 14). Beyond these figures, the importance of software spreads across many other industries, such as retail and logistics, that digitize their traditional business with IT and software (Veit et al., 2014, p. 48-49).

For researchers, the software industry provides a unique context of study because "software is not like other businesses" (Cusumano, 2004, p. 1). The *economic properties of the software industry*

#	Economic property
	Software markets
1	Exponential performance improvement per unit cost of hardware with time
2	Intangibility
3	Indestructibility
4	Reproducibility
5	Transmutability
6	Portability by information systems
7	High economies of scale
8	High economies of scope
9	New pricing models
10	Value assertion requires consumption
11	Utility-dependent value
12	Opportunities for differentiation strategies
13	Presence of network effects
14	Presence of lock-in effects
15	Security and privacy exposure
	Software products
16	Development with information systems
17	Dependency on information systems
18	High complexity
19	High need for good software architecture
20	Possibility of standardization
21	Iterative development
22	Mechanisms for rights management
23	Secondary role of performance
24	Preferability of software over hardware implementations
25	Customer involvement in product development
26	Support of users during information processing
27	Customer involvement in value creation

Table 2.1: Economic properties of the software industry.

make it different from others contexts. Whereas many of the individual properties can be found in other industries, the total set of properties is unique and leads to software-specific phenomena. For instance, open source software (OSS) and cloud computing are specific to software. Generic concepts cannot capture the software-specific characteristics of these phenomena. Consequently, industry-specific research is necessary in order to fully understand software firms.

The remainder of this section develops a comprehensive set of economic properties that make the software industry unique. A literature review is conducted based on multiple sources with a particular focus on properties of software products, software firms, and software markets. Table 2.1 summarizes the economic properties obtained from the following sources: Buxmann et al. (2012), Schief (2013), Messerschmitt and Szyperski (2003), Engelhardt (2008), Stelzer (2004), Klosterberg (2010). Following Hess et al. (2012, p. 5), the obtained properties are broadly classified as attributes of software markets and software products. A property is attributed to software products if it has a direct impact on how products are made and a property is attributed to software markets if it impacts market forces, players, or prices.

2.1.1 Economic properties of software markets

The first economic principle of *exponential performance improvement per unit cost of hardware with time* is known as Moore's Law. Moore (1965) predicted a doubling of components in integrated circuits every year. This prognosis of exponential growth has been observed in multiple areas such as computing power, storage, and networks (Messerschmitt and Szyperski, 2003, p. 30). After decades of observation Moore's Law has been taken further stating that "the performance per unit cost of material information technologies increases exponentially with time" (Messerschmitt and Szyperski, 2003, p. 29). As a result, massive computing power, storage, and network capacity are available at low cost. They are thus affordable to a broad range of organizations and individuals. The law's end seems inevitable in the future, as stated by Moore himself: "We won't have the rate of progress that we've had over the last few decades. I think that's inevitable with any technology; it eventually saturates out. I guess I see Moore's law dying here in the next decade or so, but that's not surprising." (Courtland, 2015). However, we haven't seen the end of it yet and thus dramatic improvements in performance per cost unit for material can be further expected. Applied to software, Moore's Law states that any amount of money buys faster hardware tomorrow than today. The difference in performance is exponential which allows software running on the hardware to store exponentially more information, execute faster, and communicate faster.

Software does not have a physical manifestation. It is a purely logical set of instructions and thus immaterial, resulting in *intangibility* (Messerschmitt and Szyperski, 2003, p. 21). Fundamental characteristics of intangible products such as information in general are indestructibility, reproducibility, and transmutability (Choi et al., 1997, p. 69-74). Software is *indestructible* because its usage does not result in a loss of quality. It is *reproducible*, because copies can be made without a loss of quality. And software is *transmutable*, because cost-effective modifications can be made to create new variants of a product.

Intangibility is a prerequisite for the *portability by information systems*. Software can be developed and distributed globally over the internet, which results in low distribution costs when compared to material goods (Klosterberg, 2010, p. 261). This facilitates global competition while leaving little advantage to operations in home markets (Buxmann et al., 2012, p. 3).

Investments in setting up a development infrastructure are comparably low compared to the infrastructure required for the production of material goods. Nevertheless, the costs to create an initial version of a software are high (Klosterberg, 2010, p. 259). However, for every copy sold, the replication costs are low, thus resulting in *high economies of scale*. Additionally, vendors can realize *high economies of scope* by reusing software components across multiple products (Engelhardt, 2008, p. 12). Reusage of components is facilitated by the transmutability and standardization possibilities of software.

With software, *new pricing models* become feasible (Messerschmitt and Szyperski, 2003, p. 327-338). For instance, software can monitor objective measures of its usage and enforce payments using on the collected data. This allows for usage-based pricing upon such measures as number of users or number of transactions. In contrast, unit costs provide little guidance in software pricing due to the high economies of scale. A comprehensive overview of pricing parameters is provided by Lehmann and Buxmann (2009).

The purchase of software is a decision under uncertainty, where *value assertion requires consumption* (Engelhardt, 2008, p. 16). This is a general property of information and goods with that property are referred to as experience goods (Messerschmitt and Szyperski, 2003, p. 20).

In response, decision-making to purchase software is influenced by objective as well as subjective factors. On the subjective side, value is influenced by personal, situational, and external factors (Buxmann et al., 2012, p. 14-15), e.g., personality, culture, and recommendations. A related property is the *utility-dependent value* of software, which states that the subjective value of a software to the consumer is the principal factor defining the overall value (Messerschmitt and Szyperski, 2003, p. 344-346). Consequently, this opens up *opportunities for differentiation stategies* when providers work on serving individual customer needs better than competitors do (Messerschmitt and Szyperski, 2003, p. 326).

A major property of software markets is the *presence of network effects* (Buxmann et al., 2012, 22-37). With network effects, the value of a good depends on the size of the network, which goes beyond the value of the good in isolation (Katz and Shapiro, 1994, p. 93). The size of the network is an externality in economic sense and can be of direct and indirect nature. An example for a direct network effect in software markets are social networks, where the value of using a social network increases with the number of people using the network. An example for an indirect network effect in software markets is the enterprise resource planning software (ERP software) SAP, because with increasing network size the number of service provides increases as well, thus increasing the value of an SAP installation. The presence of network effects leads to an emphasis of a broad user base, because it increases the overall value of a software, thus attracting more users. More users generate additional revenue that generates above-average profits due to the high economies of scale.

Another major property of software markets is the *presence of lock-in effects* (Messerschmitt and Szyperski, 2003, p. 317-320). In general, software acquisition can be associated with two types of costs. There are the direct costs of acquisition for the product and there are the switching costs (e.g., for user training and redesigning business processes). The switching costs are often considerable, thus posing high change barriers and resulting in a lock-in for the customer.

Finally, certain factors of software contribute to additional *security and privacy exposure*. Given the complexity of software, it is likely for software to contain errors, some of which may be exploitable by unauthorized attacks (Engelhardt, 2008, p. 14). Since software is often accessible from remote information systems, attacks can be potentially run on a global scale. Software can further collect and expose its users' private data on purpose, such as for targeted advertising, thus compromising the privacy of its users (Messerschmitt and Szyperski, 2003, p. 59-60).

2.1.2 Economic properties of software products

It takes information systems to create software (Stelzer, 2004, p. 11). In particular, *development with information systems* means that software, such as integrated development environments or version control systems, is used to create new software. The better the software used for development, the better will be the result. Consequently, this defines a *dependency on information systems* (Stelzer, 2004, p. 11). However, the dependency is taken further to running software on information systems, as software requires a certain environment to run. This includes certain hardware as well as software, such as an operating system. Further, additional information systems are often used to allow for an interaction between software over networks such as the internet.

The process of software development is characterized by a *high complexity* (Engelhardt, 2008, p. 14). It requires an understanding of the software program code as well as the information systems used for development and for execution of the software. A particular consequence of high complexity is the *high need for good software architecture* (Stelzer, 2004, p. 14-15). With software architecture, software engineers abstract from details by creating modules which can be developed independently. The individual modules are made to interact with each other via module interfaces. As a result, in a team, no individual needs to be familiar with every part of the program code as long as there is an expert for each module and there is a person, typically the software architect, who knows the big picture. Further, good architecture enhances software flexibility, as modifications can be restricted to few modules without changes to other modules, thus lowering the overall effort required.

The modularity of software gives rise to the *possibility of standardization*. Through standardization, modules and their interfaces can be precisely defined. This makes modules exchangeable, thus enabling software providers to purchase parts of a software rather than developing everything themselves (Messerschmitt and Szyperski, 2003, p. 232-235). While standards allow for an easier integration of software, custom software can be made according to specific needs. Consequently, organizations must consider the optimal degree of standardization (Buxmann et al., 2012, p. 37-47).

Software is often created through *iterative development* which helps to collect early feedback from its users (Messerschmitt and Szyperski, 2003, p. 48-49). Iterative development uses early and frequent releases of prototypes to collect feedback from users. This is an important enabler of agile development methods, which help to deal with unknown and incomplete requirements (Messerschmitt and Szyperski, 2003, S. 81-82).

Because software is easily replicated without loss of quality, it poses particular challenges to the protection of intellectual property (Messerschmitt and Szyperski, 2003, p. 267-268). At an extreme, “once a copy is available on the Internet, intellectual property rights are practically unenforceable” (Buxmann et al., 2012, p. 3). Software providers thus pro-actively equip their products with *mechanisms for rights management* to hinder uncontrolled replication with techniques such as conditional access control and copy protection (Messerschmitt and Szyperski, 2003, p. 276).

Moore’s Law leads to a *secondary role of performance* in software development. Software developers can, to a certain degree, neglect performance in favor of enhacing other characteristics of the software at hand (Messerschmitt and Szyperski, 2003, p. 31). This is because improvements in hardware performance can outweigh insufficient software performance, thus freeing developers to focus on problems which cannot be solved with higher-performance hardware, such as usability and maintainability. Another consequence of Moore’s Law is the *preferability of software over hardware implementations*. This is because software implementations will benefit from future performance improvements in hardware it is running on (Messerschmitt and Szyperski, 2003, p. 32-33). Hardware implementations wouldn’t benefit from such improvements, but would require modifications themselves in order to keep up with competing hardware products.

As a consequence of utility-dependent value, there is considerable *customer involvement in the development process* to ensure customer-oriented design of the software and thus maximize its value to the customer (Messerschmitt and Szyperski, 2003, p. 48-49). A major concern is the *support of users during information processing*, e.g., by highlighting important information and making traversal between information pieces fast and user-friendly (Messerschmitt and Szyperski, 2003, p. 20). In addition to collecting feedback from customers during product development, there is often further *customer involvement in value creation* (Stelzer, 2004, p. 13). For instance, users can be involved in creating contents such as uploading pictures to a photography website or writing reviews on a website for product recommendations.

2.2 Determinants of firm performance and risk

Research on determinants of firm success, such as performance and risk, has a long tradition and remains of continuous interest today. The question what determines firm success or failure encompasses other research questions such as how firms behave and why they differ (Mehra, 1996; Porter, 1991). For instance, while researchers and practitioners are interested how certain concepts impact performance, by turning the logic around, the existence of significant effects is a prerequisite for a truly useful concept (Thomas and Venkatraman, 1988, p. 540-541).

The following Section 2.2.1 elaborates factors determining firm success and their theoretical background. Section 2.2.2 introduces performance as a multi-dimensional construct and risk as an interrelated concept to be taken into account when evaluating firm success. Section 2.2.3 addresses issues associated with the measurement of these complex concepts.

2.2.1 Determinants

A *determinant* of firm success is a factor or concept that has a theoretical and empirical impact on firm performance or risk. Theories explaining the origins of differences in firm performance provide insights into the nature of determinants and guidance to their operationalization. Two main perspectives have been developed in the past decades, including the market- and the resource-based view. The market-based view, being the traditional perspective, has been challenged in the past decades by the resource-based view (Spanos and Lioukas, 2001, p. 907-912). Though seemingly conflicting, both perspectives are complementary to the analysis of determinants of firm performance.

Research in the tradition of the *market-based view*, rooted in early works by Mason (1939) and Bain (1956), assumes that managerial choices regarding firm strategy are constrained by industry structure. Therefore, it is assumed that factors related to firm environment and industry in particular determine firm performance. This is sometimes referred to as the structure-conduct-performance paradigm, where structure refers to industry forces and conduct to firm strategy (Spanos and Lioukas, 2001, p. 908). The paradigm is that structure determines conduct and therefore ultimately performance.

The market-based view has been advanced further by Porter (1980, 1985, 1991) who recognized the role of firm strategy in influencing performance (Spanos and Lioukas, 2001, p. 908-909). In particular, Porter emphasized the importance of firm activities, thus his extended perspective can be referred to as the *activity-based view*. While industry structure remains the main determinant of performance, it is further influenced by a firm's *value system* or *value chain*, which refers to the configuration of value activities and their interrelations (Porter, 1991). *Value activities* are defined as "the physically and technologically distinct activities a firm performs . . . by which a firm creates a product valuable to its buyers" (Porter, 1985, p. 38). Whereas *value* can be defined as "the perceived worth in monetary units of the set of economic, technical, service and social benefits received by the customer firm in exchange for the price paid for a product offering, taking into consideration the available suppliers' offerings and prices" (Anderson et al., 1993, p. 5). For Porter (1991), competitive strategy is how a firm configures its value chain to respond to the five forces driving industry competition: competitors, potential entrants, buyers, suppliers, and substitutes. As a result of the value chain configuration, a firm can gain competitive advantage from cost advantage or differentiation. Thus, Porter's framework acknowledges that firm performance is determined by industry and firm effects. Both are interrelated and while industry structure affects the overall performance, a firm's competitive strategy affects its relative performance towards competitors.

In contrast to the market-based view, the *resource-based view*, rooted in the work of Penrose (1959), emphasizes the importance of firm-specific resources as determinants of firm performance (Wernerfelt, 1984; Barney, 1991). It assumes that firms base their strategies upon different resources, including: (1) physical capital resources such as IT infrastructure; (2) human capital resources such as training and experience of the employees; (3) organizational capital resources such as internal processes. These resources are heterogeneous across firms and may not be perfectly mobile, resulting in persistent performance differences across firms within an industry. In order for superior performance to substain, respective resources must be valuable, rare, non-imitable and non-substitutable (Barney, 1991).

The resource-based view builds upon the activity-based view by Porter (1991), but emphasizes the importance of firm resources (Barney, 1991, p. 105). It positions resources as the starting point of competitive strategy, where activity configuration follows and is dependent on resources. In addition to the performance impact stemming from the competitive positioning as a result of resources, resources may have a direct impact on performance by virtue of efficiency (Collis, 1994). Thus, from the perspective of the resource-based view, firm resources are the main origin of competitive advantage and, therefore, firm performance.

Though there are apparently conflicting assumptions between the market- and the resource-based view, for instance regarding the mobility of resources, there are important similarities between them as well. Both perspectives maintain that an attractive strategic position is crucial to gain competitive advantage and establish above-average performance (Spanos and Lioukas, 2001, p. 911). While the market-based view emphasizes external determinants, namely threats and opportunities, the resource-based view emphasizes internal determinants, namely strengths and weaknesses (Barney, 1991). This consensus of two seemingly opposing perspectives states that firm performance is determined by both, the industry the firm competes in and its resources. Consequently, both views are complementary in explaining firm performance (Grover and Saeed, 2004, p. 25).

Armed with these complementary views, it becomes possible to define the term *strategy* (also referred to as competitive strategy) as a bundle of resources available to a firm and the firm's positioning towards the forces driving industry competition. The combined framework of both views as outlined in Figure 2.1 relates strategy to the "strengths-weaknesses-opportunities-threats" analysis (SWOT analysis). It suggests that "firms obtain sustained competitive advantages by implementing strategies that exploit their internal strengths, through responding to environmental opportunities, while neutralizing external threats and avoiding internal weaknesses" (Barney, 1991, p. 99).

The measurement of strategy and the arising competitive advantage is complex, such that many studies empirically link resources directly to performance, thus eliminating the need to measure competitive advantage itself (Crook et al., 2008; Ketchen et al., 2007). The logic goes that if there is empirical evidence of resources affecting performance, then it must be because resources allow for a strategy that creates competitive advantage, as shown in Figure 2.2. As pointed out by Ketchen et al. (2007, p. 962), this is a short-link to reality, because "customers do not mail checks to a company just because the company possesses certain resources". Nevertheless, extant empirical studies often use firm resources as a proxy for concepts such as strategy and competitive advantage, even if these resources are not directly linked to performance. However,

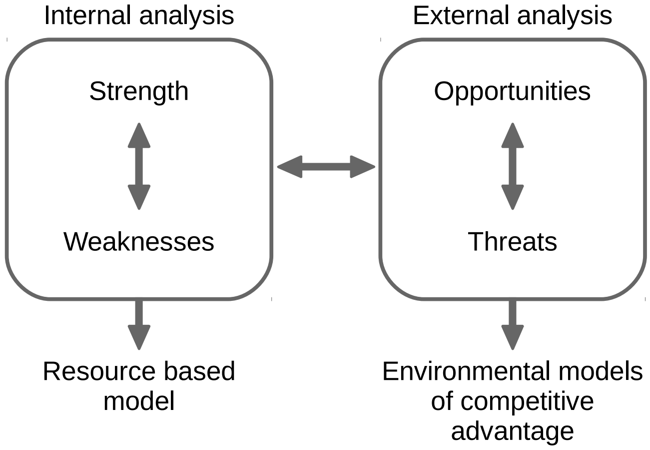

Figure 2.1: Strategy definition from the perspective of the SWOT analysis (see Barney, 1991, p. 100).

operationalization of such determinants remains difficult, for example, as Mehra (1996, p. 310) points out, industry-specific expertise is required to operationalize resource-based determinants. Arguably, such expertise requires deep knowledge of the industry and relevant data may be difficult to obtain.

In addition to the complexity of finding the right measures, the number of potential determinants in question is excessive. Capon et al. (1990) performed a meta-analysis and found more than 200 variables being used as determinants of performance. 25 years have passed since the publication of the study and it can be assumed that a follow-up study would report higher numbers today. Given that there are many determinants of performance, Capon et al. (1990, p. 1157) call for holistic studies which include many determinants and integrate different fields, rather than narrowly defined research.

Another implication from the market- and the resource-based view is that performance is affected at different levels: firm strategy, strategic group, and industry. While the industry level determines the average performance, idiosyncratic strategic decisions determine the relative performance by creating competitive advantage at the firm level. Industry and firm factors further both establish persistent strategic groups in an industry that impact firm performance as well.

The three levels are interdependent parts of an overall system in terms of systems theory (Scott, 1998). The parts are interdependent, because the overarching industry and strategic groups shape firm strategies, while firm strategies reshape those external factors (Short et al., 2007). For example, an industry imposes certain cost structures upon firms competing in that industry, thus impacting their pricing. However, a firm may decide to engage in price slashing, thus forcing its

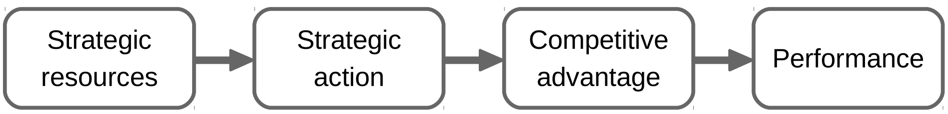

Figure 2.2: Conceptual link from resources to performance (see Ketchen et al., 2007, p. 962).

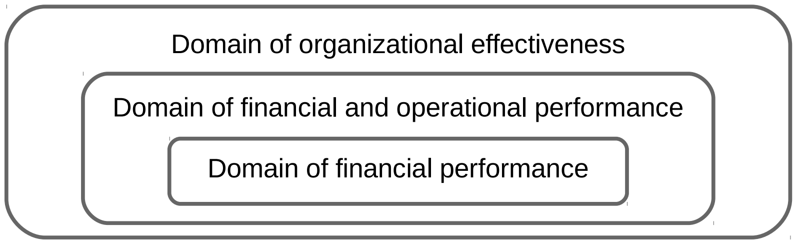

Figure 2.3: Scopes of the performance concept (Venkatraman and Ramanujam, 1986, p. 803).

main competitors to reduce their prices as well. This may change the firm's strategic group in terms of strategy and performance. As multiple firms engage in price slashing, overall profitability of the industry may be reduced, possibly leading to improved cost structures aiming to recover the profit margins.

2.2.2 Performance and risk

This thesis focuses on two aspects of firm success, namely performance and risk. The concept of firm performance has received considerably more research attention than firm risk. Though some studies use the term performance to denote both, economic returns and the associated risk, this thesis clearly differentiates between the terms performance and risk and uses the term *firm success* as a broader term encompassing both concepts.

Firm *performance* can be broadly defined as the "net flow of resources through a firm" (Deephouse, 1999, p. 148). It is a complex construct with conflicting dimensions, such as long-term growth versus short-term profitability. Thus, multi-dimensional operationalizations should be used to capture its complexity (Short et al., 2008; Slater and Olson, 2000). Furthermore, the scope of the concept must be clear, because the definition above refers to resources in general. Venkatraman and Ramanujam (1986, p. 802-804) define three different scopes, from a narrow focus on financial performance to a broad focus on organizational effectiveness in general as illustrated in Figure 2.3.

Venkatraman and Ramanujam (1986, p. 802) define the most narrow scope *financial performance* as "the fulfilment of the economic goals of the firm", reflected in measures such as return on assets (ROA), price / earnings ratio (PER), and sales growth. Zott and Amit (2007) differentiate two categories of financial performance measures: accounting-based and market-based measures. Accounting-based measures such as ROA are historical measures of past, realized, short-term performance. In turn, market-based measures such as PER reflect the expectations of capital markets regarding future, long-term cash flows to shareholders. The two types of measures thus capture different dimensions of performance. Empirical studies showing varying results between these measures confirm the multi-dimensional nature of the performance concept (Keats and Hitt, 1988; Hoskisson et al., 1993). Furthermore, market-based measures help to reduce shortcomings related to the usage of accounting-based measures. For instance, as Fisher and McGowan (1983) point out, the accounting rate of return does not equal the economic rate of return due to factors such as depreciation schedules. In recognition of the shortcomings in accounting-based performance measures, multiple studies include market-based measures in strategic management (e.g., Amit and Wernerfelt, 1990; Hoskisson et al., 1993), strategic group (e.g., Short et al., 2007; DeSarbo et al., 2009), and business model research (e.g., Zott et al., 2011).

Operational, *non-financial performance* relates to "success factors that might *lead* to financial performance" (Venkatraman and Ramanujam, 1986, p. 802, emphasis in original). This is reflected in such measures as market share, value added, and product quality. An even broader perspective is provided by *organizational effectiveness* with factors which cannot be linked to performance directly or indirectly such as employee turnover and contribution to society (Venkatraman and Ramanujam, 1986, p. 802). Organizational effectiveness goes well beyond the scope of most studies.

Another aspect of firm success, other than firm performance above, is firm risk (Venkatraman and Ramanujam, 1986; Capon et al., 1990; Houthoofd and Heene, 1997). In strategic management, *risk* is often defined as the "unpredictability of business outcome variables such as revenues, costs, profit, market share, and so forth" (Bromiley et al., 2001, p. 261). Following the definition of financial performance above, the term *financial risk* can be defined as the unpredictability of the fulfilment of the economic goals of the firm. A common measure for unpredictability is the variability of factors representing financial performance, such as variability of ROA or variability of PER over time.

The two aspects of firm success, performance and risk, are interrelated. For instance, steady profits are associated with less risk than volatile profits. Firms prefer steady over volatile profits, because steady profits allow for better planning. In turn, better planning allows for lower risk buffers, e.g., firms with lower risk should have lower capital and liquidity requirements. Therefore, measures of financial and non-financial performance can be adjusted for risk in order to judge the overall, core performance of a firm (Cool and Schendel, 1987; Houthoofd and Heene, 1997).

The empirical findings on the relationship between performance and risk yielded conflicting results. The capital asset pricing model (CAPM) by Sharpe (1964) predicts a positive relationship between risk and capital market returns and received considerable empirical support (Nickel and Rodriguez, 2002). This finding seems intuitive, as it suggests that decision makers and investors will not accept higher risk unless they gain higher performance. However, Bowman (1980) observed a negative relationship between performance and risk using accounting-based measures. The latter observation was so unexpected and counter to the CAPM that Bowman's finding became known as the *Bowman's paradox* and inspired scholars to explain the sources of this phenomenon (e.g., see Nickel and Rodriguez, 2002 for a comprehensive review of the literature). The significance of Bowman's paradox lies in its implication that firms do not face a tradeoff between performance and risk. This raises the question why any investor would be interested in investing in a firm that is less successful in all dimensions of success.

Widely accepted explanations of Bowman's paradox root in the prospect theory and the behavioral theory (Andersen et al., 2007; Bromiley et al., 2001). Prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1986) holds that expected performance outcomes affect risk-taking. When prospects are positive, individuals tend to be risk-averse and when prospects are negative, individuals tend to be risk-taking. Thus, prospect theory suggests that well performing firms keep their risks low, while low performing firms increase their risks. Behavioral theory provides a consistent view, suggesting that risk-behavior is driven by the level of performance in comparison to given aspiration levels (March and Shapira, 1987; Bromiley, 1991). Consequently, decision-makers will seek to take higher risks to increase performance when their relative performance is low (Bazerman, 1984; Hartman and Nelson, 1996).

While Ruefli (1990) questions empirical evidence supporting Bowman's paradox as a product of misspecifications and spurious effects, the empirical evidence is considerable (e.g., Bowman, 1980;

Fiegenbaum and Thomas, 1986; Cool and Schendel, 1988; Veliyath and Ferris, 1997; Andersen et al., 2007). Further studies in this research stream find support for the prospect and the behavioral theory in particular (e.g., Fiegenbaum, 1990; Sinha, 1994; Gooding et al., 1996; Andersen et al., 2007). Bowman's paradox is thus another empirical finding supporting the view that accounting- and market-based measures capture different dimensions of the performance concept.

2.2.3 Measurement

As has been pointed out thus far, this thesis conceptualizes firm performance via the dimensions financial performance, non-financial performance, and risk. Other than the conceptualization of performance, its measurement is another issue that has been of particular interest to scholars (Venkatraman and Ramanujam, 1986, 1987). Empirical measurement approaches can be differentiated with regard to the data source and the assessment mode of measures. Data source refers to the usage of primary or secondary data and assessment mode refers to objective or subjective estimation of measures. Along these demarcation lines, a four-cell classification is defined as illustrated in Figure 2.4.

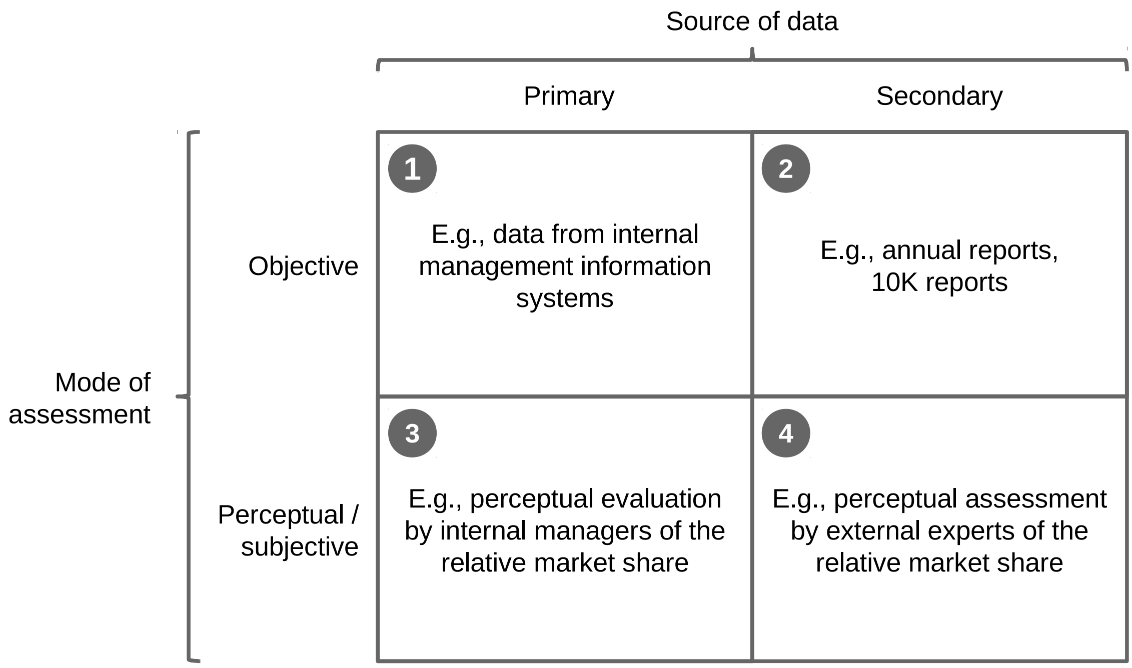

Figure 2.4: Classification scheme for measures (see Venkatraman and Ramanujam, 1987, p. 110).

As pointed out by Venkatraman and Ramanujam (1987, p. 110-111), no single approach is superior to the others. Each approach is associated with its own advantages and disadvantages. For instance, while objective measures seem preferrable to subjective measures because they are not prone to any rating bias, Covin (1991, p. 448) points out different cases when it might be more appropriate to choose subjective measures. First, firms might be reluctant to provide certain data to the public. Second, even when respondents are willing to provide the data, if this data isn't publicly available, then there is no way to validate the responses. Third, when dealing with small firms, such data may be difficult to interpret. Finally, the absolute magnitude of the measures may vary from industry to industry and thus corrupt comparability. Similarly, while secondary data allows for replication, it might not be available or insufficiently accurate. On the other hand, primary data can be collected according to the particular needs but challenges replication and may be biased. Venkatraman and Ramanujam (1987, p. 111) recommend method triangulation by combining the extremes, namely objective secondary data and subjective primary data, to test hypotheses and be more confident of the empirical results.

2.3 Strategic groups

Strategic group research is a field with over 40 years of history in strategic management research (Hunt, 1972). The basic idea of strategic groups is at the heart of configurational research: to view an industry "as composed of clusters or groups of firms, where each group consists of firms following similar strategies in terms of the key decision variables" (Porter, 1979, p. 215). Consequently, a *strategic group* is defined as a "group of firms in an industry following the same or a similar strategy along the strategic dimensions" (Porter, 1980, p. 129). The focus on a particular industry is a major aspect of this definition, because it separates strategic groups from generic strategies (see Section 1).

Considerable criticism has been expressed by scholars regarding strategic group research. Hatten and Hatten (1987, p. 329) argued that strategic groups are merely "an analytical convenience" with insufficient theoretical foundation. Barney and Hoskisson (1990) emphasized the lack of theoretical and empirical rigour. McGee and Thomas (1986) as well as Thomas and Venkatraman (1988) questioned the relevance of the concept altogether because the empirical evidence confirming a link between strategic groups and performance was lacking. Overall, several scholars "not only pointed out seeming limitations of configurational research but also called into question the merit of conducting such research" (Short et al., 2008, p. 1054).

In response to the early criticism, scholars have addressed the expressed weaknesses. They developed the theory further, designed a common methodology, and collected empirical evidence. Though previous findings have been mixed, a meta-analysis (Ketchen et al., 1997) as well as recent studies (Murthi et al., 2013; Schimmer and Brauer, 2012; DeSarbo et al., 2009) found empirical evidence supporting the presence of performance differences across a broad range of industries.

Other than performance, there is much less empirical evidence regarding risk and risk-adjusted performance. As outlined in Table 2.5, there are only six studies providing empirical results on risk-related dimensions of success and the most recent study dates back to the year 2006. Moreover, as Table 2.2 indicates, the overall results have been mixed. In the case of risk-adjusted performance, Cool and Schendel (1987) find no differences across groups, Houthoofd and Heene (1997) find differences for the larger strategic scope groups, but not for strategic groups, and Pandian et al. (2006) find differences in most, but not all periods. Clearly, this stream of research requires further empirical evidence.

Even less strategic group research exists for the software industry. The only available study by Short et al. (2007) analyzes prepackaged software firms. Consistently to the results on other industries, the authors find significant performance differences across strategic groups. However, Short et al. (2007) analyze twelve industries in their study and do not provide any interpretations specific to the software industry. Another study by Grover and Saeed (2004) focuses on internet-based businesses. Arguably, these can be related to the software industry. The authors identify four distant groups. The best performing group consists of large firms that succeed in scaling their

Reference	Results on	
	Firm risk	Risk-adj. performance
Cool and Schendel (1987)	No differences across strategic groups	No differences across strategic groups
Cool and Schendel (1988)	Differences across strategic groups in 9 out of 18 time periods	
Houthoofd and Heene (1997)		Differences across strategic scope groups, no differences across strategic groups
Veliyath and Ferris (1997)	Differences across strategic groups	
McNamara et al. (2003)	No differences across strategic groups	
Pandian et al. (2006)		Differences across strategic groups in 3 out of 5 time periods

Table 2.2: Empirical results of industry-specific success studies in strategic group research regarding risk and risk-adjusted performance.

operations. The worst performing group is comprised of mid-sized firms in the sample that partly succeed in scaling but have low margins and high liabilities. The two average performing groups consist of the smallest firms with low scale effects.

In what follows, Section 2.3.1 lays out the theoretical foundations of strategic groups and Section 2.3.2 introduces the methodology to derive and evaluate strategic groups. This provides the foundation to address the white spots in extant strategic research, being the impact on risk and analyzes of the software industry, and to analyze software business models with configurational methods.

2.3.1 Theory

The main theoretical approach to strategic groups, *mobility barriers*, proposes the existence of intra-industry entry barriers, which delineate an industry into groups and lead to differences across them (McGee and Thomas, 1986; Murthi et al., 2013). Mobility barriers come into existence as a result of strategic decisions. Strategic decisions are long-term decisions that cannot be easily reversed or imitated without considerable costs and risk. Consequently, the strategic position of a firm is persistent over time. Firms following similar strategies thus form persistent strategic groups. The groups are delineated by their key strategic decisions which function as intra-industry entry barriers. The barriers prevent most desirable strategic groups from additional competition. Because firms cannot easily switch into more desirable groups to improve their performance, the competition level in each group remains stable over time. Stable competition levels, in turn, help to preserve extant performance differences across groups.

Originally, the theory of mobility barriers is rooted in the perspective of the market-based view, because intra-industry entry barriers are external factors affecting firm decisions. However, mobility barriers can be explained from the perspective of the resource-based view as well (Mehra, 1996; Leask and Parnell, 2005). Because resources such as patents are heterogeneously distributed, they delineate an industry into groups of firms having similar resources. They further function as mobility barriers because resources may not be perfectly mobile, thus firms cannot easily acquire more favorable resources, which results in sustainable performance differences across groups. It is thus apparent that the market- and the resource-based view are both consistent with the theory of mobility barriers. As reflected in several empirical studies (e.g., Cool and Schendel, 1987; Fiegenbaum, 1990; Mehra, 1996), both views are complementary in explaining strategic group characteristics and performance.

The second theoretical argument using cognitive models (Porac et al., 1989) suggests that managers mentally partition an industry into groups of firms. A manager uses these groups to define the strategic position of his firm, suppliers, and competitors. He further forms beliefs regarding the success factors of each group (Porac et al., 1989, p. 399). The sum of a manager's beliefs and assumptions forms his *cognitive model* of an industry. The cognitive model is then used to derive actions that have an impact on firm success. Thus, from this perspective, strategic groups can be viewed as a tool which is actively used by decision makers to make sense of their competitive environment (Reger and Huff, 1993, p. 115).

The theory of cognitive models links strategic groups to the action-based view of the firm, because perceived groups directly impact firm actions and actions, in turn, affect firm success. However, external market factors in the tradition of the market-based view have an impact on the mental models of decision makers and thus impact firm success as well (Porac et al., 1989, p. 399). The mechanisms behind cognitive models have been specified further using group identities (Peteraf and Shanley, 1997). The theory of strategic group identity suggests that groups recognized by their members have a strong identity that impacts their actions, whereas groups with weak identities are transient artefacts.

The application of both theories, mobility barriers and cognitive models, in extant research is summarized in Table 2.3. It indicates that cognitive models are a more recent development in the field with fewer references overall. Mobility barriers have clearly received more attention in the reviewed studies. Notably, mobility barriers chronologically precede cognitive models but remain of continuous interest today. It appears that both theories provide complementary approaches to strategic groups and are thus used in parallel by scholars today.

Theory	**References**
Cognitive models	7: Cheng and Chang (2009); Deephouse (1999); DeSarbo et al. (2009); Nath and Gruca (1997); Neill and Rose (2006); Reger and Huff (1993); Schimmer and Brauer (2012)
Mobility barriers	24: Cool and Schendel (1987, 1988); Desarbo and Grewal (2008); Ebbes et al. (2010); Ferguson et al. (2000); Grover and Saeed (2004); Houthoofd (2009); Houthoofd and Heene (1997); Ketchen et al. (1993); Lawless et al. (1989); Leask and Parker (2007); Mas-Ruiz and Ruiz-Moreno (2011); McNamara et al. (2003); Mehra (1996); Murthi et al. (2013); Nair and Kotha (2001); Nath and Gruca (1997); Pandian et al. (2006); Porter (1979); Short et al. (2007); Veliyath and Ferris (1997); Wiggins and Ruefli (1995); Zúñiga-Vicente et al. (2004)

Table 2.3: Overview of applied theories in strategic group research. Nath and Gruca (1997) use both theoretical foundations, the reference is thus counted twice.

2.3.2 Methodology

The evolution of strategic group analysis has resulted in a multitude of approaches and methods to derive and evaluate strategic groups. This section summarizes the overall methodology in seven steps as shown in Figure 2.5. Each step of the analysis addresses a major activity in the overall analysis. The methodology has been extracted from strategic group studies which apply the industry-specific strategic group concept and provide original empirical results on performance or risk effects of strategic groups. A structured overview of all relevant studies is given in Table A.1 (p. 116).

Step one in strategic group analysis positions the research with regard to the theoretical argument and the empirical approach to derive groups. The theoretical argument chooses between the available theories, sharpens the concept at hand, and derives testable hypotheses. The empirical approach relates to fundamental decisions that precede the selection of particular methods to derive groups. In general, the empirical approach can be either inductive or deductive. The choice of the empirical approach is strongly linked to the theoretical argument.

Two general empirical approaches are used to derive and compare groups: the inductive and the deductive approach (Ketchen et al., 1993). The *inductive approach* numerically derives groups from data using statistical techniques such as clustering, resulting in a taxonomy of strategic groups. This approach maximizes internal validity of the groups by deriving groups fitted to the data. The number of groups, their main characteristics, and performance are not specified before the groups are formed. The inductive approach is therefore particularly useful for exploratory research. On the downside, this sacrifices generalizability and additional care must be taken to refute the possibility that the groups are more than mere statistical artefacts.

The *deductive approach* uses theoretical argument to propose a typology of strategic groups. It clearly states the number of groups, their main characteristics, and performance outcome based on theoretical argument. Being rooted in theory, the deductive approach is particularly useful for confirmatory research related to the underlying theory. In comparison to the inductive approach, Ketchen et al. (1997) found that deductively derived groups show a weaker impact on performance than inductively derived groups.

Both, the inductive and deductive approach are important to strategic group research and their application depends on the research question at hand, such as exploring an industry or confirming the predictions of a typology. However, given the definition of strategic groups as an industry-specific concept, a deductive approach requires the application of an industry-specific theory to derive groups. None of the industry-specific performance studies in strategic group research does that, such that all studies presented in Table 2.3 use the inductive approach - though some of them additionally use the deductive approach to derive generic strategies.

Step two in strategic group analysis defines the data being used for empirical testing. Because strategic groups are an industry-specific concept, the data needs to be limited to particular industries. Further limitations include the selection of a time period and a sampling frame for the firms to be included in the sample.

In extant research, most studies analyze a single industry. Only three studies include multiple industries (Short et al., 2007; Veliyath and Ferris, 1997; Wiggins and Ruefli, 1995). Table 2.4 provides an overview of industries being the subject of at least two strategic group studies. It shows that the banking and the pharmaceutical industry are of particular interest to strategic group research. The culprit for the popularity of both industries appears to be grounded in the

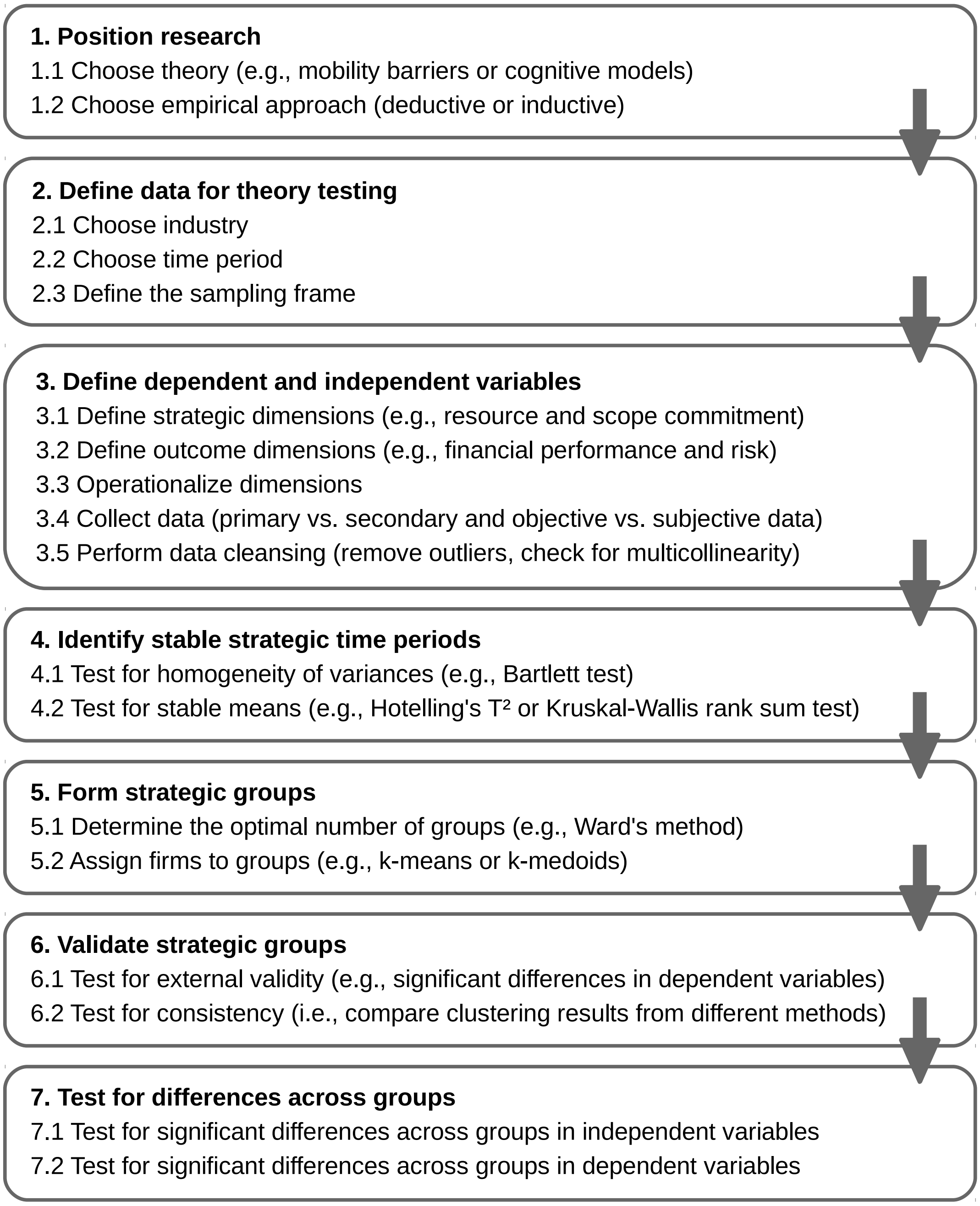

Figure 2.5: Strategic group analysis methodology.

Industry	References
Banking	9: Deephouse (1999); DeSarbo et al. (2009); Desarbo and Grewal (2008); Ebbes et al. (2010); Mas-Ruiz and Ruiz-Moreno (2011); McNamara et al. (2003); Mehra (1996); Reger and Huff (1993); Zúñiga-Vicente et al. (2004)
Pharmaceutical	8: Cool and Schendel (1987, 1988); Leask and Parker (2006); Pandian et al. (2006); Short et al. (2007); Veliyath and Ferris (1997); Wiggins and Ruefli (1995); Leask and Parker (2007)
Airline	2: Murthi et al. (2013); Veliyath and Ferris (1997)
Hospital	2: Ketchen et al. (1993); Nath and Gruca (1997)
Insurance	2: Ferguson et al. (2000); Schimmer and Brauer (2012)

Table 2.4: Overview of industries included in multiple strategic group studies.

availability of detailed data as well as in the industry dynamics (e.g. McNamara et al., 2003; Pandian et al., 2006).

Step three in strategic group analysis provides the data for subsequent analyzes. Variables are defined, operationalized, relevant data is collected, and cleansed. The independent variables represent strategic dimensions that are used to form strategic groups. The dependent variables are the outcome variables, such as firm success. The theoretical foundations guiding the selection of strategic and success dimensions have been discussed in Section 2.2.

Following the theoretical arguments from the perspectives of the market- and resource-based views, inductive studies address both views by defining strategic variables that cover the two dimensions competitive scope and resource commitment (Cool and Schendel, 1987). *Scope commitment* defines the markets in which a firm operates and the degree of firm operations in these markets. *Resource commitment* defines how a firm uses its resources and the degree of resource usage. Though these two dimensions still allow for a broad selection of relevant variables, they do define a minimum of dimensions to be covered. As a consequence, a wide range of variables has been used to represent strategic dimensions. For instance, Ketchen et al. (1993) extracted more than 80 variables from extant research. Ketchen et al. (1997) differentiated between studies defining a broad or narrow array of variables and found that broad definitions decrease classification errors and have a stronger performance effect. Mehra (1996) provides empirical evidence that variables derived from the perspective of the resource-based view are more appropriate to detect differences across groups.

Looking at dimensions of firm success, most studies in strategic group research address financial firm performance, as can be seen from Table A.1 (p. 116). The most popular variables are ROA and return on equity (ROE). Notably, both are accounting-based measures. Other popular accounting-based measures are return on sales (ROS) and profit margin. The most popular market-based measure is Tobin's q, followed by dividend yield and market-to-book ratio. The most popular non-financial variable is market share.

Regarding risk, few studies include risk or risk-adjusted performance as relevant success dimensions. Table 2.5 summarizes the studies including measures of risk or risk-adjusted performance. The summary shows that most studies use the variance of a performance measure as a measure of risk. Another measure is beta, which can be readily obtained from some secondary sources. How-

Reference	Risk measures	Risk-adjusted performance measures
Cool and Schendel (1987)	Variance of each performance measure, e.g., ROS	Each performance measure divided by its variance
Cool and Schendel (1988)	ROS variance	ROS divided by its variance
Houthoofd and Heene (1997)	Not included	ROA divided by its variance
Veliyath and Ferris (1997)	Beta of daily market returns; Variance of daily market returns	Not included
McNamara et al. (2003)	Variances of ROA, ROE, and OPM	Not included
Pandian et al. (2006)	Not included	CBRET

Table 2.5: Overview of risk and risk-adjusted performance measures used by industry-specific success studies in strategic group research.

ever, beta is only applicable to measures rooted in capital markets, whereas variance is applicable to any measure.

Nearly all reviewed industry-specific performance studies in strategic group research use objective data obtained from secondary sources, such as Compustat. No study uses subjective data from secondary sources. Few studies rely on objective data from primary sources (Nath and Gruca, 1997; Reger and Huff, 1993) and subjective data from primary sources (Houthoofd, 2009; Neill and Rose, 2006).

Data cleansing involves the removal of outliers as well as tests for multicollinearity, because highly correlated variables have a negative impact on clustering procedures (Ketchen and Shook, 1996), thus leading to difficulties in step 6. The strategic dimension, which is measured by these correlated variables, is overrepresented and has a higher impact on the formation of strategic groups than other dimensions. This corrupts the theoretical foundation of the strategic variable and therefore the strategic groups become flawed. Common tests include (1) the Spearman's rank correlation coefficient which measures the dependence between two variables. (2) The variance inflation factor (VIF) is calculated for every variable (Fox and Monette, 1992) and has been applied in some strategic group studies (Schimmer and Brauer, 2012; Nair and Kotha, 2001). A factor greater than 10 indicates correlated variables (Zott and Amit, 2008). (3) The condition number of eigenvalues is computed for the entire data matrix of strategic variables (Belsley et al., 1980). If the condition number exceeds 30, multicollinearity exists in the matrix (Deephouse, 1999).

Overall, strategic group research defines high data requirements to derive meaningful groups, often with the consequence of a comparably low sample size. For instance, Short et al. (2007) obtain less than 100 firms for 10 out of 12 industries.

Step four in strategic group analysis identifies *stable strategic time periods*. As has been argued by Fiegenbaum et al. (1990, p. 134), meaningful strategic groups require "time periods of strategic homogeneity with regard to competitive strategy behavior". In unstable time periods strategic variables are volatile and thus do not allow for a reliable measurement of underlying strategies. Stable strategic time periods must fulfil two requirements in all strategic measures (Fiegenbaum et al., 1990, p. 136-137): homogeneity of variances and stable means over the time period. For tests of homogeneity of variances the Bartlett test has been suggested by Fiegenbaum et al. (1990). Nair and Kotha (2001) recommend the Fligner-Killeen test as a non-parametric alternative to the parametric Bartlett test. For tests of stable means Fiegenbaum et al. (1990) suggest the parametric Hotelling's T^2 test. An alternative is the non-parametric Kruskal-Wallis rank sum test.

Step five in strategic group analysis forms strategic groups by applying clustering algorithms to the strategic variables. A general limitation of clustering is that it will always produce clusters. Following the recommendations of methodologists to increase clustering validity (Punj and Stewart, 1983; Ketchen and Shook, 1996), most studies have used a two-stage clustering procedure to form objective strategic groups. Within the first stage, hierarchical agglomerative clustering (HAC), i.e., Ward's method with Euclidean distance, is performed to determine the optimal number of groups and their centroids. The optimal number of groups can be selected based on the Mojena criterion (Mojena, 1977) or based on the modified Mojena criterion (Milligan and Cooper, 1985). Within the second stage, the cluster means are used as starting points for k-means clustering. This iterative, non-hierarchical clustering procedure finds the best groupings given the number of groups for each strategic period. Figure 2.6 illustrates the steps of the two-stage clustering procedure.

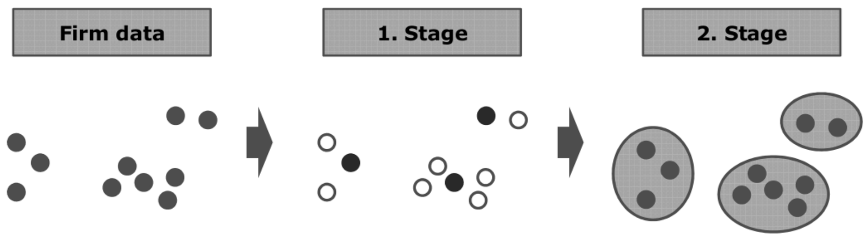

Figure 2.6: Two-stage clustering procedure.

Prior to the actual clustering of firms to strategic groups, standardization of strategic variables needs to be considered because variables with larger value ranges have a higher impact on the formation of strategic groups than variables with smaller ranges (Ketchen and Shook, 1996). Standardization ensures that each variable has the same impact in clustering. The disadvantage of standardization is the possibility of destroying meaningful differences between variables. Because no clear guidance can be given on whether to standardize or not, it has been proposed to compare the results between them and assess the validity of both when inconsistencies arise (Ketchen and Shook, 1996).

Step six in strategic group analysis validates the strategic groups obtained from clustering procedures. Validation is required in order to ensure the meaningfulness of the clustering solution (Punj and Stewart, 1983). Some studies have attempted to validate clustering solutions with tests for differences across groups in strategic variables. However, as pointed out by Anderberg (1973, p. 15): "The trouble with making such tests is that they are hardly relevant. The whole focus of the clustering criterion and algorithm is to give a set of clusters that are well differentiated from each other." It has been thus suggested that validity must be tested on external variables which have not been used to form the clustering solution (Ketchen and Shook, 1996). For instance, differences in dependent success variables across groups could be regarded as indicators of meaningful clustering solutions. Additionally, the consistency of a clustering solution is a necessary, but not sufficient condition of validity. In general, consistency is estimated by comparing the clustering solution to other solutions which are obtained with different methods or samples (Ketchen and Shook, 1996). For instance, average proportion of non-overlap (APN) is a consistency measure that indicates how many firms are assigned to a different cluster if the respective variable is removed. Though no guidance can be provided as to how high the figure should be, it can be argued that variables without any influence are not indicative and variables with too much influence question the stability of the clusters. As a guideline, the influence of each variable should be disproportionate to the number of variables overall.

Step seven in strategic group analysis analyzes the group characteristics with regard to their strategies and performance impact. For that, tests identifying significant differences across strategic groups are performed on strategic and performance variables. For tests across clusters the Kruskal-Wallis rank sum test (e.g., used by Cool and Schendel, 1988; Houthoofd and Heene, 1997) provides a non-parametric alternative to the parametric analysis of variances (e.g., used by Veliyath and Ferris, 1997; Ebbes et al., 2010). While these methods detect overall differences, a *post-hoc analysis* aims at identifying individual groups which are significantly different. The post-hoc analysis compares all pair-wise combinations of strategic groups and identifies significantly different pairs. For these pair-wise comparisons, Ketchen et al. (1993) and Ferguson et al. (2000) have used the Bonferroni test, which adjusts the p-values, because multiple comparisons suffer from family-wise error rate (Bettis, 2012). Significant differences across clusters indicate group characteristics and can be used to interpret strategies applied by firms in the cluster and their performance impact.

The performance impact is often measured using a *lagged structure*, because some strategic decisions may require a certain time period before having an impact on performance (Deephouse, 1999; Short et al., 2007). For instance, if strategy and performance are measured based on a five-year period, a time lag of two years means that each strategic period and the corresponding performance period overlap in three years. Figure 2.7 illustrates the example.

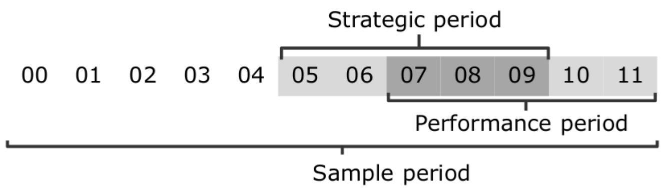

Figure 2.7: Exemplary lagged structure.

2.4 Business model configurations

Driven partly by the shift from traditional to electronic business, the term *business model* emerged in the mid-1990s. It has become a concept of growing research interest to scholars in strategic management and information systems (IS) research. Conceptual and empirical work continues to accumulate and the rising importance of the business model field is reflected in the growing number of publications and comprehensive literature reviews in recent years (Burkhart et al., 2011; Lambert and Davidson, 2012; Zott et al., 2011). The reviews find that various aspects of business models are addressed by scholars such as theoretical foundations, conceptualizations, classifications, and performance implications.

The following sections identify how the business model concept can contribute to our under-

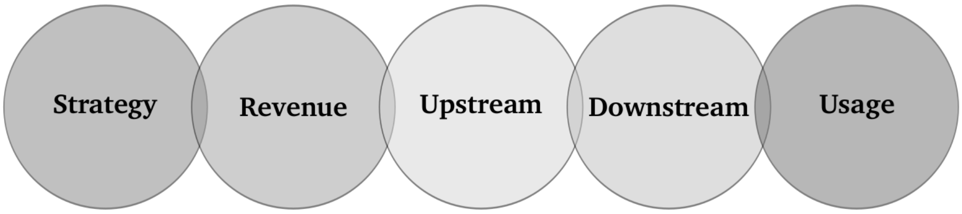

Figure 2.8: The five dimensions of the software business model framework.

standing of the software industry and how business model research can benefit from the configurational analysis in this thesis. For that, Section 2.4.1 introduces the business model concept. Section 2.4.2 reviews empirical configurational business model studies. Finally, Section 2.4.3 compares business model research with strategic group research in particular and configurational research in general.

2.4.1 Concept

Despite the growing body of conceptual and empirical work, the definition of the concept itself remains unclear (Zott et al., 2011; Al-Debei and Avison, 2010). The complexity of the concept is illustrated by an often acknowledged definition by Al-Debei and Avison (2010, p. 325) who define a business model as an "abstract representation of an organization, be it conceptual, textual, and/or graphical, of all core interrelated architectural, co-operational, and financial arrangements designed and developed by an organization presently and in the future, as well all core products and/or services the organization offers, or will offer, based on these arrangements that are needed to achieve its strategic goals and objectives". However, as pointed out by Krumeich et al. (2012), business models are often defined by their components. For instance, Zott and Amit (2010) define a business model as an activity system with two sets of parameters: design elements (content, structure, and governance) and design themes (novelty, lock-in, complementarities, and efficiency).

Generally, business models are regarded as multi-faceted and thus multiple dimensions are used to capture the concept (e.g. Al-Debei and Avison, 2010; Krumeich et al., 2012; Schief and Buxmann, 2012; Wirtz et al., 2015). While multiple scholars in this research stream suggest different sets of components, this thesis uses the Software Business Model Framework (SBMF) by Schief and Buxmann (2012), because it combines a broad set of business model components specific to the software industry. The SBMF is a morphological field (Ritchey, 1998) with 25 business model components. The components are classified in the five dimensions illustrated in Figure 2.8. Each dimension consists of five business model components. For each component, there are discrete choice options, some of which may be mutually exclusive. The SBMF is illustrated in Figure 2.9.

The first dimension *strategy* characterizes the strategic decisions of a software firm. The component *value proposition* describes how the firm generates its competitive advantage. For instance, a unique image or design can be related to Porter's differentiation strategy, such as is the case for Apple products. A lower price can be related to the cost leadership strategy, such as is the case for some webspace providers. The component *investment horizon* outlines the financial commitment and objectives. For instance, an income model results in investments up to the point that the business generates sufficient and stable incomes to the principles. This is presumably particularly

Strategy												
Unique Selling Proposition	Image	Quality	Features	Innovation Leadership	Intimate Customer Relationship	Cost Leadership / Efficiency	Network Leverage	One Stop Shop				
Investment Horizon	Subsistence Model	Income Model	Growth Model	Speculative Model	Social Model	Cross Finance						
Value Chain Activity	Research	Develop-ment	Production	Marketing	Imple-mentation	Education	Operation	Maintenance	Support	Replacement		
Degree of Vertical Integration	Low	Medium	High									
# of Cooperation Partners	None	One	Few	Many								
Revenue												
Sales Volume	Low	Medium	High									
Revenue Source	Direct	Advertising	Commission									
Pricing Assessment Base	Usage Based	Hybrid Combination	Usage Independent									
Structure of Payment Flows	Upfront	Hybrid Combination	Recurring Payments									
Revenue Distribution Model	Low	Medium	High									
Upstream												
Software Stack Layer	Application Software	Systems Software	Hardware Control / Embedded Software	(Web) Content								
Platform	Desktop Computers/ Notebooks	Servers	Mobile	Cloud Computing	Embedded Systems	Social Media	Game Consoles					
License Model	Proprietary: Sell Usage Rights	Proprietary: Sell all Rights to Customers	Open Source: Copyleft Licenses (e.g. GPL)	Open Source: Permissive Licenses (e.g. BSD)								
Degree of Standardization	Individual Production	Batch Production	Bulk Production									
Key Cost Driver	Research & Development	Marketing & Sales	Services	Third Party Software Licenses	Hardware	Subcontracting						
Downstream												
Localization	Local (Germany)	EMEA (Europe, Middle East, Africa)	AMERICAS (North-, Central-, and South America)	APJ (Asia, Pacific, Japan)								
Target Customer	Small Organizations	Medium Organizations	Large Organizations	Private Individuals								
Target Industry	All	None	Manu-facturing	ICT	Finance & Insurance	Wholesale & Retail	Pharma & Chemicals	Transport & Storage	Services -Health etc	Construct. & Utilities	Public Sector	Others
Target Users	Business - Broad Workforce	Business - Dedicated Specialists	Business - Managers	Consumers	Software Developer							
Channel	Sales Agents	Events	Telesales	Online Shop	Retail Stores							
Usage												
Implementation Effort	Low	Medium	High									
Operating Model	On Premise	Both: On Premise & On Demand	On Demand									
Maintenance Model	Daily	Weekly	Monthly	Quarterly	Biyearly	Yearly						
Support Model	Standard Support	Few Support Options	Customer Specific Support									
Replacement Strategy	One Release	Few Releases	Many Releases									

Figure 2.9: The software business model framework (Schief and Buxmann, 2012).

relevant for small firms, who prefer steady incomes rather than risky growth. Another choice option is the growth model, in which case growth is valued more than profits in an attempt to maximize firm value. Amazon is an example of a company following the growth model, because it reinvests its entire profits and additional capital, thus even generating losses while growing further. The component *value chain* outlines the main value activities of a firm. For instance, firms offering pre-packaged software will regard development as one of their main activities, while operations may not be that important. On the other hand, firms providing cloud computing solutions will regard operations as one of their main activities, while development may not be as important. The component *degree of vertical integration* defines how many activities are carried out internally by the firm vs. the activities carried out externally (e.g., via outsourcing). For instance, a firm that mostly relies on suppliers will be characterized as having a low degree of vertical integration. The component *number of cooperation partners* outlines how many partners are involved in the firm's value creating activities. For instance, a consulting firm providing customization services for a specific enterprise software may have the provider of the enterprise software as the only cooperation partner.

The second dimension *revenue* characterizes the financial flows of a firm and its pricing model. The component *sales volume* indicates the volume of products sold, e.g., as the number of product installations. For instance, an international standard software firm such as Kaspersky has a high sales volume. The component *revenue source* indicates where revenues come from. For instance, many news websites on the internet derive revenues from advertising and directly from the reader. The component *pricing assessment base* indicates how prices are defined. For instance, cloud infrastructure offerings on the internet such as Dropbox employ a usage-based pricing scheme. The component *payment flow structure* indicates the point in time when payments are made. For instance, cloud offerings often use recurring payments on a regular basis. The component *revenue distribution model* indicates what part of the revenues is shared with other stakeholders. For instance, firms distributing mobile applications over a marketplace such as Google Play Store often share a high percentage of their revenues with Google.

The third dimension *upstream* characterizes the products and their development. The component *software stack layer* indicates what type of software the firm's products relate to. For instance, in case of a firm offering an antivirus software would be classified as a system software firm. The component *platform* indicates the technical platform that the product employs. For instance, smartphone applications use the mobile platform. The component *license model* indicates the legal arrangements regarding the software code. For instance, open source projects often use copyleft licenses such as GNU General Public License (GPL) that requires everyone making modifications to the software to publish the modified software code with the same license. The component *degree of standardization* indicates whether the product is highly standardized or highly individual. For instance, mobile applications are usually the same for all customers and are thus referred to as bulk production. The component *key cost driver* indicates the dominating cost drivers associated with the product. For instance, a software reseller will have third party software licenses as his key cost driver.

The fourth dimension *downstream* characterizes the target markets and the distribution of products. The component *localization* indicates the geographic markets targeted by the firm. For instance, a global firm like SAP targets North, Central-, and South America (AMERICAS), Asia, Pacific, and Japan (APJ) as well as Europe, Middle East, and Africa (EMEA). The component *target customer* indicates the type of target customers by the size of targeted organization. For instance, large ERP software firms like SAP target large organizations whereas a mobile game developer targets private individuals. The component *target industry* indicates the industries for which the product is made. For instance, a fraud management software in telecommunications targets the information and communication technology (ICT) industry. The component *target user* indicates the type of end-users of the product. For instance, a computer-aided design software (CAD software) firm targets dedicated specialists rather than a broad workforce. The component *channel* indicates the types of sales channels used by the firm to target customers. For instance, mobile applications are sold through an online shop such as Google Play Store.

The fifth dimension *usage* characterizes the services that come along with the usage of the product. The component *implementation effort* indicates the required installation and configuration effort to run the product. For instance, an antivirus software for private consumers should require only low efforts, whereas an ERP software will require high efforts. The component *operating model* indicates how the product is deployed for usage. For instance, cloud software is available on-demand, whereas software meant for installation and execution at customer side is available on-premise. The component *maintenance model* indicates how frequently a firm releases updates of its product. For instance, Ubuntu Linux is released biyearly. The component *support model* indicates the type of support contract offered by the firm. For instance, Microsoft offers a hybrid combination of standard support to the general public and customer-specific support to partners for Microsoft Windows. The component *replacement strategy* indicates the number of different product releases available at a time. For instance, Oracle Websphere Application Server is only available in few releases at a time.

The software business model framework suggested by Schief and Buxmann (2012) is used in this thesis for multiple reasons. It was designed specifically for the software industry and thus uses many software-specific components, such as the software value chain, software stack layer, platform, target user, and operating model. This framework further defines a broad range of components to capture many different aspects of a software firm's activities. Notably, the framework is well rooted in state-of-the-art literature on business models and reflects the components of other studies. Finally, this framework is well suitable for quantitative research, because it defines choice options for each component, thus allowing for quantitative data collection and evaluation.

2.4.2 Configurational results

Another stream in business model research builds upon business model components to derive business model configurations. This includes deductive studies that conceptually derive classes from theory or expert knowledge (e.g., Malone et al., 2006; Kontio et al., 2005) as well as inductive studies that empirically derive classes with statistical techniques (e.g., Morris et al., 2013; Valtakoski and Rönkkö, 2010). This "exploratory and largely descriptive research is required before explanatory and predictive research can be undertaken" (Lambert and Davidson, 2012, p. 7). It is thus a prerequisite for hypothesis and theory testing, where business model classes can be used as dependent or independent variables (e.g., in statistical analyzes of determinants of firm performance).

Business model configurations in the software industry have been of interest to multiple scholars. In order to build upon previous findings, this thesis provides a detailed literature review of peer-reviewed studies that fulfil the following criteria: (1) the study specifically addresses the software industry, (2) it uses discrete configurations (e.g., a taxonomy, typology, or classification in general), and (3) the study clearly focuses on the business model concept as a unit of analysis.

Overall, six relevant studies could be identified. The references and structured content summaries are given in Table 2.6. The studies are analyzed based on different criteria. The first criteria *classification approach* differentiates between inductive and deductive approaches taken by the respective authors. This fundamental discriminator has an impact on the methods, the results, and their applicability. The statistical technique is included for inductive studies. The second criteria *data source* characterizes the source of the empirical data used in the study. The third criteria *sample* outlines the sampling frame, sample size, and the time period of the data. The fourth criteria *business model components* summarizes the components used to conceptualize business models. The fifth criteria *configurations* outlines the business model classes used or found in the study. The sixth criteria *success measures* names the measures being used to evaluate success effects of business models. Finally, the seventh criteria *main results* summarizes the main findings presented by the study.

Bonaccorsi et al. (2006) analyze the business models of OSS firms. Their main focus is on firms that combine OSS and proprietary software. For the 138 OSS firms from Italy, five components are used to operationalize the firms' business models: share of OSS sales to total sales; share of OSS products to products overall; types of solutions offered (only OSS, mainly OSS, indifferently proprietary and OSS); strategic importance of OSS; intensity of GPL usage. The data is collected through a survey. Upon the measures, an HAC analysis is performed. As a result, two business model configurations are derived: more OSS oriented and less OSS oriented business models. The configurations are used to analyze the determinants of OSS adoption. They find that switching costs, size, experience with proprietary software, and network externalities negatively impact openness.

Engelhardt (2004) analyzes the business models used by 105 German software and IT-services firms after their initial public offering (IPO) at the Neuer Markt segment. Two business model configurations are deductively derived. Though the method is not clearly stated, it appears to be mostly based on the author's reasoning. Entrepreneurial business models are radically innovative, aggressively oriented towards growth, and being employed by young firms. Traditional business models are employed by mature firms, with an orientation towards defensive growth and incremental innovation to reduce business risk. Accordingly, Engelhardt (2004) includes performance measures as components of business models. However, the author does not use these configurations to study their effects. Instead, he uses predefined industry groupings (e.g., business software) as independent variables in regression analysis. He then associates the groupings with the business model configurations to draw conclusions. Unfortunately, the source of the groupings and the secondary data in general is not stated. The main result of the study is that standardization appears to be disadvantageous for performance, as opposed to the assumptions of entrepreneurial business models.

Reference	Classification approach	Data source	Sample	Business model components	Configurations	Success measures	Main results
Bonaccorsi et al. (2006)	Inductive using HAC (average linkage method)	Primary (survey)	146 Italian OSS firms (year 2003)	5: share of OSS sales; share of OSS products; solution types; strategic importance of OSS; intensity of GPL usage	2: more OSS oriented; less OSS oriented	None	Switching costs, size, proprietary experience, and network externalities negatively impact openness.
Engelhardt (2004)	Deductive	Secondary (not specified)	105 German software and IT-services IPOs (years 1997-2001)	7: innovation; standardization; profitability; growth; age; finance; corporate governance	2: entrepreneurial; traditional	ROS; sales; sales growth; productivity growth	Least standardized business models perform the best.
Kontio et al. (2005)	Deductive	Primary (survey)	163 Finnish software firms (year 2004)	2: standardization; product-orientation	4: solution consultants; product integrators; product tailors; product licensors	None	Software firms align their business models and development processes.
Rajala and Westerlund (2007)	Deductive	Primary (interviews, survey) and secondary (public information)	6 Finnish software vendors (years 2002-2003)	2: involvement in customer relationships; standardization	4: software tailoring; applied formats; resource provisioning; standard offerings	None	Configurations differ in their emphasis on internally and externally obtained resources.

Table 2.6 – continued on next page

Table 2.6 – continued from previous page

Reference	**Classification approach**	**Data source**	**Sample**	**Business model components**	**Configurations**	**Success measures**	**Main results**
Schief et al. (2012)	Inductive using HAC (average linkage method)	Secondary (annual public reports)	77 world-largest public software firms (year 2010)	6: company focus; vertical integration; product-orientation; product focus; target customer; license model; localization; pricing model	7, 13, 34: not named	ROA, CfMoS, Tobin's q	Some business models overly represented (e.g., focus on business customers). Unique business models vary strongly in performance while common business models achieve moderate performance.
Valtakoski and Rönkkö (2010)	Inductive using two-stage clustering (K-mediods and several HAC methods)	Primary (survey)	612 Finnish software firms (year 2009)	4: offering; activities; value network; and revenue logic (measured by sales shares of various sources, e.g., content and ads)	8: software product; deployment project; development service; ASP and SaaS; not software; content and ads; software consulting; hardware	Sales; sales growth; personnel, willingness to grow, profitability, productivity	Business models differ in size, willingness to grow, sales growth and profitability.

Table 2.6: Overview of studies on business model configurations in the software industry.

Kontio et al. (2005) analyze the business models employed by 163 Finnish software firms. They deductively derive two business model dimensions from own expertise. The first dimension reflects the standardization of the offering, i.e., how well the offering could be duplicated without customer-specific work. The second dimensions reflects the product-orientation of the offering, i.e., the share of product sales to total sales. As a result, four business model configurations are derived. Solution consultants mainly provide services by tailoring products. Product integrators mainly provide services such as user training for highly standardized products. Product licensors mainly sell highly standardized products. Product tailors mainly sell products and provide some customer-specific adjustments. The main finding of the study is that software firms align their business models and development processes. For instance, product-oriented firms use more milestones and business models with less standardization are more inclined to rely on OSS.

Rajala and Westerlund (2007) present case studies on six Finnish software vendors. They collect primary and secondary data on the firms and deductively derive four business models based on two business model components. The components include the degree of involvement in customer relationships and the standardization of the offering. The first configuration software tailoring refers to business models with low standardization and deep customer involvement, such as individual software development. Applied formats refers to business models with high standardization and deep customer involvement, such as ERP software providers. Resource provisioning is characterized by low standardization and low customer involvement. Standard offerings refer to business models with high standardization and low customer involvement, such as standard antivirus software. The main finding of the study is that business model configurations differ in their usage of internal and external resources. For instance, firms following the standard offerings business model possess the internal resources for product development, but acquire external resources for product distribution.

Schief et al. (2012) analyze the business models of the 77 largest public software firms worldwide. They collect secondary data from annual reports and inductively derive business models configurations. The components are based on the software business model framework by Schief and Buxmann (2012) and include company focus (share of software sales to total sales), vertical integration, product-orientation, product focus (sales of instrastructure software relative to application software), target customer (private or business), license model (importance of proprietary relative to OSS), geographic focus, and pricing model. Using HAC, different optimal clustering solutions are identified with 7, 13, and 34 configurations. The authors do not assign names to the clusters. The configurations are used as independent variables for descriptive performance statistics of the clusters. The performance measures include ROA, cash flow margin on sales (CfMoS), and Tobin's q. The main result of the study is that there are multiple possible configurations that are not pursued by the firms in the sample. For instance, nearly all firms use a proprietary license, focus on business customers, and receive direct payments. Further, the study finds that there are unique business models followed by companies such as Google. With respect to performance, unique business models seem to have the highest and the lowest performance, while more common business models achieve a rather average performance. The authors also find that vertical integration seems to have a positive effect on performance.

Valtakoski and Rönkkö (2010) analyze the business models of 612 software firms from Finland. They collect primary data through a survey in the year 2009 and inductively derive business model configurations. The components are derived from the literature, encompassing offering, activities, value network, and revenue logic. The authors operationalize the business model concept by measuring the revenue shares of different sources, e.g., license and advertising revenues. A two-stage clustering is performed to form configurations. The resulting taxonomy includes eight configurations: software product, deployment project, development service, application service provider (ASP) and SaaS, not software, content and ads, software consulting, and hardware. Using several performance measures, the authors find differences across configurations in size, willingness to grow, sales growth, and profitability. For instance, they find that ASP and SaaS firms report the highest willingness to grow. However, the highest actual growth is measured for the traditional software product, deployment project, and development service firms.

Several conclusions can be drawn from this literature overview. The inductive and the deductive approach are both used by researchers in this stream and are thus considered necessary and valuable. Though the sample of six studies is small, there appears to be a shift from the deductive to the inductive approach over the years. The deductive studies were published within the period 2004-2007, whereas the inductive studies were published from 2006-2012. It appears that the inductive approach has gained more interest in recent years than the inductive approach. The inductive studies all use HAC to form configurations and Valtakoski and Rönkkö (2010) even apply the two-stage clustering method. Consequently, it can be said that HAC is well accepted by scholars in the field. Four of the studies rely on primary data for empirical evaluations, three of them use surveys to collect the data. This might well reflect the difficulty to obtain the data to measure business models. Regarding geographic focus, five studies use data from a single country. Three of the studies use data on software firms from Finland. Only Schief et al. (2012) use data from firms worldwide. Authors probably limit their geographic focus because of the difficulty to obtain relevant data. Looking at business model components, it can be seen that deductive studies use much less components to derive configurations. The reason is probably that deductive studies need to limit their scope because the dimensional space becomes too large with each component (Lambert, 2015). Even though Engelhardt (2004) names seven components to characterize business models, he uses much less characteristics to derive the groups for empirical analysis. Looking at the use of concept components across all studies, some components are used in multiple studies. These include standardization, product-focus (i.e., importance of products vs. services), license type (i.e., importance of proprietary vs. OSS), company focus (i.e., software or non-software), and pricing model (i.e., importance of advertising). Interestingly, only three studies analyze performance implications and report performance differences. Only Engelhardt (2004) mentions business risk as a characteristic of business models, but does not provide an empirical measurement of risk.

2.4.3 Comparison with strategic groups

There are conceptual and theoretical links between business model configurations and strategic groups. As discussed in Section 1, business model configurations can be seen as a broader concept than strategic groups, because the former can be defined on organizational features in general, thus including competitive strategy, whereas strategic groups are limited to competitive strategy. This is reflected in the SBMF, which includes strategic as well as non-strategic components. This conceptual link is theoretically justified, as both, business models and competitive strategy, are rooted in the perspectives of the market- and resource-based views. More specifically, the value system of the activity-based view is used as the theoretical underpinning of both concepts. As a result, conceptualizations of business models will generally comprise more components and require more data for empirical evaluation than conceptualizations of competitive strategy.

A difference between business model and strategic group research is that the former does not provide theoretical explanations for the existence and effects of configurations, whereas the latter specifically addresses the group level and provides two theories for it: mobility barriers and cognitive models. However, it appears that both theories can be transferred to business model configurations. Since both theories apply to the specific concept of strategic groups, they should be relevant to the broader concept of business model configurations as well. Moreover, the theories are not specific to competitive strategy, but can be generalized. I.e., the theory of mobility barriers can be used to explain that some business model decisions cannot be changed easily, thus leading to persistent differences across configurations. For instance, a firm focusing on application software cannot become an infrastructure firm in a short period of time, because the development of an infrastructure software arguably cannot take place in a short timespan. Moreover, because of interdependencies between different characteristics, a component might be difficult to change due to interdependencies with other components that are hard to change. Similarly, it can be argued that practitioners form cognitive models such as on-demand vs. on-premise and derive their actions from these models.

Methodologically, researchers in both directions use the deductive as well as the inductive approach to derive configurations and test for performance effects. Whereas in strategic group research the two-stage clustering is a widely used method to form configurations, only one out of three inductive business model studies uses the same method. Because different methods can yield varying results, business model research should consolidate its methods. It appears that business model research could generally use two-stage clustering. The multi-faceted nature of business models may impose difficulties in combination with k-means, which does not allow for missing values. Business model data may well contain missing data as it may be difficult to obtain data on each component for all firms. As a result, business model researchers may need to use other clustering procedures than k-means. However, it seems that all extant studies could have applied the same or similar two-stage clusterings and delivered more comparable results by means of similar methods.

The validation of derived configurations is of major importance in strategic group research. In business model research, only Valtakoski and Rönkkö (2010) explicitly validate the resulting clusters, though validation is a mathematical procedure and can certainly be applied to any configurational clusters independently of the concept at hand. Similarly, strategic group methodology considers stable periods, but none of the business model studies mentions or tests stable periods. In general, it appears that no study on business models in the software industry nearly grasps the full methodological foundation provided by extant configurational research.

Looking at the effects being studied in both research streams, it becomes apparent that scholars in both streams are interested in performance implications. Business models are used to study performance implications as well as other effects unrelated to success. Risk hasn't been subject of configurational business model research thus far, but has been of some interest to scholars in strategic group research.

2.5 Conclusion and hypotheses

The previous sections introduced the main concepts, theories, and methods for the analysis of strategic groups and business model configurations in the software industry. They also provided an overview of available empirical results including success studies. Overall, it could be established that strategies and business models are related concepts that can be approached with configurational analysis. This final section relates the findings from the reviews to the research questions presented in Section 1.1. It further suggests testable hypotheses relevant to the research questions, as far as these can be derived from extant research. The hypotheses are tested in the three quantitative studies presented in Section 4, 5, and 6. These studies will also provide exploratory results well beyond the defined hypotheses.

With respect to Research Question 1, which relates to the potential contribution of configurational research for the analysis of the software industry, the literature review described how strategic groups and business model configurations can be inductively derived from firm data. These configurations indicate which strategies and business models prevail in the software industry as well as the key characteristics that make them different. By relating the configurations to success measures, it can be evaluated which strategies and business models perform better than others.

The review further indicated two white spots in extant research, particularly related to software firms: (1) missing detailed results on strategic groups in the software industry and (2) insufficient application of configurational analysis to software firm business models. Both white spots can be addressed with the well-developed theoretical and methodological foundation from strategic group research, offering the potential to improve our understanding of the software industry. Most importantly, state of the art theories and methods, e.g., two-stage clustering, can be transferred from strategic group to business model research, thus addressing the criticism that accompanies business model research to this day.

Turning to Research Question 2, which relates to value creating activities of software firms, the literature review outlined the value chain, rooted in the activity-based view, as the theoretical foundation to strategies and business models. Value creating activities are an inherent part of both concepts. Moreover, mental models of decision makers lead to activities that may change a firm's competitive advantage and, thus, ultimately impact its success. This emphasizes the importance of the value chain concept to this thesis. Section 3 addresses the second research question in detail to methodically derive value activities of software firms.

With respect to Research Question 3 and Research Question 4, which relate to prevailing strategies and business models as well as their success effects, the literature review showed varying classifications suggested by deductive and inductive studies. As this thesis follows the inductive approach and applies different conceptualizations, no hypotheses are derived here regarding which strategies and business models may be found. Of the following hypotheses, the first four relate to aspects of differences across configurations. As discussed, the presense of such differences indicates the existence of configurations and is a prerequisite for deeper analyzes in this thesis.

The first hypothesis addresses the existence of differences across configurations in their key characteristics. The theory of mobility barriers suggests that firms are delineated into persistent groups by targeted markets and available resources. Though originally applied to strategic groups, it has been discussed above that similar arguments can be applied to business models as well. Hypothesis 1 suggests that firm characteristics will vary across configurations, such that each configuration will be unique with respect to particular characteristics. Upon confirming this hypothesis, characteristics of different configurations can be analyzed and compared further to derive meaningful interpretations of each configuration. Ideally, industry experts will be able to relate the obtained configurations to abstract groupings used by these experts.

Hypothesis 1 *Configurations differ in structural characteristics.*

The theory of mobility barriers further explains how success differences come into existence and persist across configurations. It is thus relevant to Research Question 4, which relates to the success effects of configurations. As discussed above, firm success is a multi-dimensional construct. The following hypotheses relate to different dimensions of firm success. Hypothesis 2 states that some configurations will have higher financial performance than others, Hypothesis 3 states that some configurations will display higher risk than others, and Hypothesis 4 states that performance differences will still be present after adjustment for risk. Upon confirming these hypotheses, further conclusions can be drawn by looking into the relationship between particular configuration characteristics and success effects.

Hypothesis 2 *Firm performance varies across configurations.*

Hypothesis 3 *Firm risk varies across configurations.*

Hypothesis 4 *Risk-adjusted firm performance varies across configurations.*

The final hypothesis addresses possible interdependencies and tradeoffs between different success dimensions. According to Bowman's paradox, there is a negative relationship between accounting-based firm risk and performance. That is, firms with high accounting-based performance show low accounting-based risk and vice versa. Hypothesis 5 suggests that this effect might be present at group level as well.

Hypothesis 5 *Accounting-based firm performance and risk are negatively related across configurations.*

Section 3

Software industry research framework[1]

As discussed in the previous section, the activity-based view is the main theoretical foundation to strategies and business models. It suggests that firms compete by combining activities into value chains. This section provides a software-specific value chain framework for application in this thesis. It reviews extant research on value chains specific to software firms and to the software industry. The results are used to methodically develop a detailed software value chain. The subsequent sections use the developed software value chain to conceptualize and measure software firm strategies and business models. These measurable concepts then serve as the foundation to the quantitative studies in this thesis.

3.1 Value chain concept

The concept of the value chain was initially introduced by Porter (1985) as a tool for developing and sustaining competitive advantage of a firm. Porter (1985) defines a *value chain* as a set of activities, which can be separated in primary and secondary activities. By disaggregating a firm into its various activities, the value chain allows for a better understanding of costs behavior and sources of differentiation. Even though the value chain was developed for single firms and business units, it can be applied to entire industries as well (Barnes, 2002; Li and Whalley, 2002; Messerschmitt and Szyperski, 2003; Stanoevska-Slabeva et al., 2007). An obvious way to do so is by applying the analysis to an abstract typical firm representing the industry. Such a value chain would contain typical activities performed within an industry. This broader view is important since competition has been shifting from the firm level to the network level (Parolini, 1999; Spekman et al., 1994). Furthermore, looking beyond the value chain of a particular firm can be an important source of innovation (Pil and Holweg, 2006).

The key benefit of the value chain concept is its simplicity and high-level view of a firm or industry providing a simple model of the activities performed. Value chain analysis is widely used in research and praxis (Stabell and Fjeldstad, 1998). However, there are various limitations, which should be taken into account when performing analyzes based on value chains. Most of those limitations result from the fact that no clear-but rules can be stated on such high-level concepts as industry and activity frontiers. Depending on the point of view the value chain might include or

[1] Parts of this section were published in Pussep et al. (2011), Pussep et al. (2012), and Schief (2013).

exclude another upstream or downstream activity. The granularity is up to individual judgment as well, since activities can be enriched or condensed by aggregation or disaggregation. The final decision should depend on the goal of analysis and the target group (Porter, 1985). An important limitation is that the ordering of activities as shown by the value chain is no indication of the actual chronological order: "ordering of activities should broadly follow the process flow, but ordering is judgmental as well" (Porter, 1985, p. 48). Thus, activities shown might be executed in many ways, including parallel execution or exclusion of some activities. According to these limitations value chains of the same firm might be depicted in various forms. Analyzes performed on different value chains might lead to varying results, and, consequently, become incomparable.

3.1.1 Generic requirements on value chains

Explicitly outlining the requirements which a value chain should satisfy in order to enable industry and firm analysis can arguably lead to more comparable results among researchers. However, there is little literature dealing with value chain requirements. Therefore, it is suggested to adopt principles from related fields such as software development processes. The proposed requirements can be separated in two groups. The first group is concerned with the range of the value chain, stating where the value chain starts and where it ends. The second group deals with the isolation and separation of activities within the value chain.

In order to identify the boundaries of a value chain, it is suggested to focus on two requirements. First, the value chain should contain all important activities of a generic firm competing in the corresponding industry. Secondly, the value chain should range from the very beginnings of the corresponding product (such as research and product idea) up to its operation and final replacement.

The second group of requirements has been addressed by Porter (1985, p. 45) who suggested that "the basic principle is that activities should be isolated and separated that (1) have different economics, (2) have a high potential impact of differentiation, or (3) represent a significant or growing proportion of cost". Additional requirements can be identified when applying basic principles from software engineering. Software systems are often decomposed in components, with components themselves hierarchically composed of finer-grained components (Messerschmitt and Szyperski, 2003). Components should be separated such that coupling (interdependence between components) is low and cohesion (binding of the elements in a component) is high, resulting in higher software quality (Dhama, 1995). However, those generic principles do not provide a single level of granularity: "defining granularity is quite complex since it cannot draw on theoretical groundings ... granularity can hardly be measured in terms of absolute numbers, because of the subjectivity of the related concepts that may determine the granularity in question" (Haesen et al., 2008, p. 376). Therefore, there is no single definition of a component (Cusumano, 2004). It is suggested that the view of hierarchical decomposition also applies to value chains, where coarse-grained activities comprise finer-grained activities.

To my knowledge, the application of software engineering principles to value chain analysis has not been proposed before. In the context of value chains the principle of high cohesion states that comprising sub-activities within an activity should be highly related to each other. The principle of low coupling states that dependencies between different activities should be low, such that they have as little impact on each other as possible. Notably, linkages between activities are inherited in hierarchical decomposition: aggregating highly linked sub-activities into distinct activities will result in high coupling between those activities as well. Finally, the appropriate granularity depends on the respective research purpose. While a single activity might envelop the entire value chain, activities can also be defined on a very fine-grained level, resulting in a large number of activities.

3.1.2 Review of software value chains in previous works

The generic value chain as proposed by Porter (1985) consists of the five primary activities (1) inbound logistics, (2) operations, (3) outbound logistics, (4) marketing and sales, and (5) service. Primary activities are supported by firm-wide support activities. The visual size of the value chain is determined by the value generated. A margin is added to the visual representation. Evaluating the original concept, it appears that little is left of the generic value chain. All work reviewed did not take supporting activities or the margin into account and no mapping to the generic primary activities has been done (Barnes, 2002; Li and Whalley, 2002; Messerschmitt and Szyperski, 2003; Stanoevska-Slabeva et al., 2007). Concerning the proposed requirements the generic value chain does not take into account software-specific activities. Furthermore, activities such as outbound logistics are of little importance in context of software due to its intangibility and low distribution costs.

Messerschmitt and Szyperski (2003) propose two software-specific value chains. They term them requirements and supply value chain. The requirements value chain encompasses the activities (1) analysis and design, (2) implementation, (3) provisioning, and (4) operation. The supply value chain encompasses the activities (1) implementation, (2) provisioning, (3) operation, and (4) use. Another distinction made by the authors is the discrimination of application and infrastructure which is important because the use activity does not apply to infrastructure software since there are no users involved. The inclusion of the use activity seems questionable, since software is usually applied in a different industry, such that the value created should be attributed to the industry where the usage takes place. Also the distinction between the two value chains seems questionable when looking at the industry as a whole. For instance, the value chains overlap in three activities. Finally, the authors themselves note that "other standard business functions like sales [are not] specifically discussed" (Messerschmitt and Szyperski, 2003, p. 123).

Software engineering processes focus on software development, "which refers to the range of activities ... surrounding the creation of working software programs" (Messerschmitt and Szyperski, 2003, p. 70). A broader view is taken by lifecycle processes which "include all activities of a product or a system during its entire life, from the business idea for its development, through its usage and its completion of use" (Crnkovic et al., 2005, p. 321). Both works refer to the waterfall model. As initially proposed by Royce (1970), the process of the waterfall model consists of seven phases: (1) system requirements, (2) software requirements, (3) analysis, (4) program design, (5) coding, (6) testing, and (7) operations. Extended models for instance include phases such as conceptualization and upgrade (Messerschmitt and Szyperski, 2003). Those activities can be viewed as activities in the software value chain. However, since the focus is rather technical, the waterfall model misses downstream activities such as marketing and operations. Furthermore, the detailed technical point of view blows up the number of upstream activities which should be reduced in the unified software value chain. Other related concepts are the software product development framework (Xu and Brinkkemper, 2007) and the software supply network (Jansen et al., 2007).

3.2 Unified software value chain

Following the requirements on a comprehensive value chain and the review of previous works, this section introduces the unified software value chain. A first empirical proof of concept is provided based on expert interviews with five software firms.

3.2.1 Conceptualization

Works proposing unified value chains are usually based on literature reviews (Barnes, 2002; Li and Whalley, 2002). Stanoevska-Slabeva et al. (2007) outline their research approach in three steps: broad literature review, 18 case studies, and aggregation to a generic value network. Porter (1985, p. 45) just recommends: “starting with the generic chain, individual value activities are identified in the particular firm”. It becomes apparent that there is little guidance on how to propose a unified value chain. The following recursive approach is suggested here: (1) Definition of the industry and its frontiers. (2) Listing of value activities of proper granularity and naming entailed sub-activities. (3) Ordering of value activities in a visual representation of the value chain. A broad understanding of the value creation process is used here, such that the boundaries stretch from the business idea to the disposal and replacement of the product.

As has been shown in the previous section, all concepts of software value chains presented thus far do not fulfil major requirements such as listing software-specific activities or encompassing the entire industry. Therefore, those approaches are combined and a unified software value chain is proposed following the outlined approach. Research methods such as literature review, case studies, and expert interviews are applied throughout the entire process.

The second step is built on the literature review performed in the previous section. Whereas the initial waterfall model provides thourough insight into upstream activities, the extended waterfall modell and lifecycle process models allow for a better understanding of downstream activities. Furthermore, the Standard Industry Classification (SIC) code descriptions can be used for additional software activity definitions. Past works scrutinizing the software industry used SIC codes 7371, 7372, and 7373 (Léger et al., 2005; Léger and Quach, 2009). SIC code 7379 focuses on software services and code 8243 on establishments training users in the use of software, these are included as well. Based on this combination of sources the single activities are separated as outlined in Figure 3.1 and defined as shown in Table 3.1.

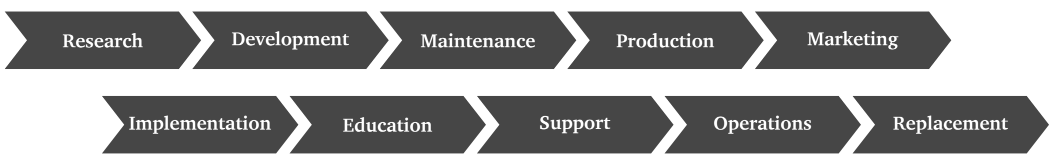

Figure 3.1: Overview of the unified software value chain.

#	Activity name	Subactivities	Detailed description
1	Research	(1) Development of a product vision, (2) fundamental research of algorithms, (3) decision upon major technologies and subsystems, and (4) proof of concept.	This activity comprises fundamental product research. A product vision is developed and fundamental algorithms are researched. Major technologies and subsystems are selected. A first proof of concept is provided through a prototype or analysis of algorithms, technologies and subsystems. The result is a product idea, algorithm or proof of concept. Unlike in the following activities, no code is created here which becomes part of the actual product.
2	Development	(1) Requirements engineering, (2) software design, (3) coding, (4) subsystem testing, (5) subsystem integration, (6) system testing, (7) user documentation, and (8) provisioning.	This activity comprises the actual software development process. Based on requirements, a software design is created. The entire system is decomposed into subsystems. Subsystems are programmed and tested separately, before they are integrated and tested as a combined system. The user documentation is created and the product is compiled to an executable and versioned product. The result is an executable version of the product.
3	Maintenance	(1) Requirements engineering, (2) software design, (3) coding, (4) subsystem testing, (5) subsystem integration, (6) system testing, (7) user documentation, and (8) provisioning.	Same as development, but the focus is on bugfixing and enhancing an existing product, whereas the activity development aims at the creation of a new product. Within maintenance, disruptive changes are not allowed. Instead, incremental changes are made by the product maker to an existing product in the marketplace.

Table 3.1 – continued on next page

Table 3.1 – continued from previous page

#	Activity name	Subactivities	Detailed description
4	Production	(1) Assembly, (2) printing, and (3) packaging.	Within assembly, software and respective documentation are bundled to one package. The assembled software package is printed to a physical medium and the documentation is printed on paper. In packaging the physical product artefacts are packaged in a physical package. The result is a product with all attributed artefacts, which is ready for shipment.
5	Marketing	(1) Launch, (2) price, (3) place, (4) promotion, (5) bundling, and (6) brand management.	Providing a means by which buyers can purchase the product and inducing them to do so, such as sales and promotion. The result is a readily marketed product in the marketplace, such that customers are aware of the product and the product is available for purchase or has been purchased already.
6	Implementation	(1) Installation, (2) configuration, (3) adjustment, and (4) business process reengineering.	The installation comprises the transmission of the packaged binaries to the customer's information system. Moreover, it ensures that the binaries can be executed without runtime errors. Configuration allows the setting of software parameters and software modifications according to the customer's needs. Finally, adaptations can be performed that modify or enhance the functionality of the software product and employ business process changes.
7	Education	(1) Training and (2) certification.	Training of users and third party firms. In addition, certifications attest users and third party firms a certain degree of seniority in the handling of a software product.
8	Support	(1) Primary support and (2) development support.	Support can be differentiated in primary and development support. While the first subactivity deals with the support of users, the second activity relies on deep technical knowledge and implies code reviews.

Table 3.1 – continued on next page

Table 3.1 – continued from previous page

#	Activity name	Subactivities	Detailed description
9	Operations	(1) Hosting, (2) monitoring, (3) backup, and (4) upgrade.	The operations activity ensures the execution and management of a product on an information system during actual usage by customers. By monitoring the system behavior can be analyzed and supervised. To minimize damages through data loss, regular data back-ups need to be planned, run, and administered. Finally, the information system needs to be upgraded to new releases during its lifecycle.
10	Replacement	(1) Alternatives, (2) migration, and (3) shut-down.	First, replacement deals with the decision if the product (once it becomes outdated and reaches the end of it's lifecycle) shall be replaced by an alternative system. If the decision for an alternative is made, data needs to be migrated from the legacy to the new system. Subsequently, the legacy system is shut-down. A seamless transition to the new system is the main target at this stage. After the irrevocable data destruction of confidential information, the shut-down activity is completed.

Table 3.1: Activity descriptions of the unified software value chain.

3.2.2 First empirical proof of concept

As a first proof of concept for the unified softwre value chain, expert interviews have been conducted with five software firms from Germany. The sample contained three big IT consulting firms producing individual software for their customers. Those were selected due to their great insight into the software industry, which extends well beyond the mere software development stages. The other two firms were one big and one small standard software producers, therefore reflecting different firm sizes in the sample. Each conducted semi-structured interview lasted about one hour. Three of them were performed via telephone and two face-to-face. The firms have been provided with a management summary of the main results.

The interviewees have been presented with the activity descriptions as shown above. They have been asked (1) whether they agree with the presentation, (2) which activities they perform, and (3) how those activities are performed: via market, firm hierarchy, or a mixture of both (Williamson, 1991), thus revealing their make-or-buy strategies for every activity. All interviewees agreed with the presentation of the unified software value chain, even though some of the activities had to

Firm	Activity									
	Research	*Development*	*Maintenance*	*Production*	*Marketing*	*Implementation*	*Education*	*Support*	*Operations*	*Replacement*
A	■	■	■	■	□	□	■	■		■
B	□	□	■	□	□	■	■	■	■	■
C	□	□	■	□	■	■	■	■	■	■
D	■	□	■	■	■	■	■	■		■
E	■	■	■	■	■			■	■	□

Table 3.2: Results of the expert interviews for each firm (A to E). Full squares denote activities carried out by firm hierarchy. Empty cells denote activities carried out by market and empty squares denote a mixture of market and firm hierarchy.

be explained in more detail by the interviewer. Also, all interviewees confirmed that their firms performed all of the activities. The results of the third question are presented in Table 3.2.

Interestingly, all five firms perform the activities maintenance and support by hierarchy. Even the IT consulting firms seem to maintain their products and provide support, which indicates that their services go well beyond sub-contracted development. Less surprising, the activity performed by market the most is operations, indicating that hosting and related services are sub-contracted more often. Overall, the results indicate a high degree of vertical integration, with the number of activities performed by hierarchy ranging from seven to eight activities.

This first empirical proof of concept demonstrated that industry experts agree with the unified software value chain. Each of them was further capable to apply the concept to depict an important strategic aspect of his firm.

3.3 Hierarchy of activities

While the previous section provided a holistic value chain for the software industry, this section is concerned with two further aspects of value chain conceptualization. First, Porter (1985, p. 45) stated that "activities should be isolated and separated that (1) have different economics, (2) have a high potential impact on differentiation, or (3) represent a significant or growing proportion of cost". However, none of the reviewed works used explicit methods to ensure the separation of activities based on these criteria. Secondly, the appropriate level of granularity is an unsolved issue as well, requiring an explicit treatment. Consequently, this section seeks to answer the following questions:

1. How to ensure the economic uniqueness of activities within a value chain?

2. How to provide an approapriate level of activity granularity within a value chain?

By answering these questions the concept of the unified software value chain is further enhanced. Furthermore, a methodical approach is developed which is applicable to any industry. For the development of the value chain methods, some fundamental guidelines are respected. First, the methods are built upon related work in the area of value chain concepts. Further, the perspective is enhanced by software industry-specific research that provides means and hence allows domain experts to analyse economic characteristics of software value activities. Finally, from a methodical point of view, well-known and proved methods are used. By compelling logic and reasoning, these methods need to be combined and arranged into a study setting that allows to be used as a method to distinct value chain activities. In what follows, the following four steps are performed:

1. Definition of activity attributes: Starting with the previously defined unified software value chain, economic properties of the software industry are used to derive economic attributes describing each activity.

2. Measurement of activity attributes: For each activity, a quantitative value is derived based on the Delphi study method.

3. Assertion of activity uniqueness: By pairwise comparison of activities, a minimal set of attributes is identified which allows a perfect separation of the activities.

4. Construction of a value chain hierarchy: Using clustering algorithms, activities are combined to higher levels of abstraction, resulting in a hierarchy of value chains which comprise ten to three activities.

3.3.1 Definition of activity attributes

Following the ultimate goal to use quantitative methods to answer the previously defined questions, activities need to be made measurable. For that, attributes are defined and value ranges are assigned to the attributes. Thus, activities are described by a vector, where its size equals to the number of attributes and the vector elements correspond to the attribute values of the respective activity. Using this representation of activities, quantitative analyzes can be performed.

The definition of attributes is crucial in order to obtain meaningful results. The following criteria should be met by attributes describing an activity:

- Relevance: Attributes must be relevant for the industry at hand in order to be able to describe it. Furthermore, attributes should be relevant for all types of created products (e.g., in case of software: individual and custom software).

- Completeness: Attributes should capture all facets of the activities with regard to some concept. The concept might be a theory providing a set of properties relevant to the indusry at hand. If those properties are chosen as attributes, then all of them should be used, such that the selection of attributes can be claimed to be complete with regard to the underlying theory.

- Determinability: Attributes should have a clearly defined range of values and the attributes should be determinable by an expert.

In case of the unified software value chain, the economic properties of the software industry (see Section 2.1) provide a theoretical foundation. However, it was not possible to use these properties directly, as they have not been defined as variables, e.g., often not having a clear value range. A further challenge was the mere number of properties. In response, an interlayer of measurable attributes was introduced. The attributes are shown in Table 3.3. The attributes are dichotomic,

Activity attribute		Values	
#	**Description**	**Value "0"**	**Value "1"**
A	Activity results are rather:	tangible / "can be touched"	intangible / immaterial
B	Decisions made during activity performance are rather:	strategic	technical / operational
C	Per customer (thus, for one product instance) the activity is rather performed:	multiple times	one time
D	Activity execution requires knowledge of the product source code:	yes	no
E	Activity execution requires deep IT understanding:	yes	no
F	End-users are involved in activity execution:	intensively	loosely
G	Activity can be performed once and its results can be reused multiple times (for multiple customers) by the producer:	yes	no
H	On first activity execution, the following costs prevail:	personnel	non-personnel (e.g., hardware)
I	In relation to the point of productive usage on customer side (the go live), the activity is chronologically performed:	before	after
J	Activity results are rather a change in:	human knowledge	information systems
K	The extent by which activity results contribute to the compatibility of the software:	high	low

Table 3.3: Activity attributes and their value ranges.

	Economic property	Most relevant activity attributes									
		A	B	C	D	E	F	G	H	J	K
1	Exponential performance improvement	□	□	□	□	□	□	□	□	■	□
2	Intangibility	■	□	□	□	□	□	□	■	□	□
3	Indestructibility	■	□	□	□	□	□	□	□	□	□
4	Reproducibility	■	□	■	□	□	□	■	□	□	□
5	Transmutability	■	□	□	□	□	□	□	□	□	□
6	Portability by information systems	■	□	□	□	□	□	□	□	□	□
7	High economies of scale	□	□	■	□	□	□	■	□	□	□
8	High economies of scope	□	□	■	□	□	□	■	□	□	□
9	New pricing models	□	■	□	□	□	□	□	□	□	□
10	Value assertion requires consumption	■	□	□	□	□	□	□	□	□	□
11	Utility-dependent value	■	□	□	□	□	□	□	□	□	□
12	Opportunities for differentiation	□	■	□	□	□	□	□	□	□	□
13	Presence of network effects	□	□	□	□	□	□	□	□	□	■
14	Presence of lock-in effects	□	□	□	□	□	□	□	□	□	■
15	Security and privacy exposure	□	□	□	□	□	■	□	□	□	□
16	Development with information systems	□	□	□	■	■	□	□	□	□	□
17	Dependency on information systems	□	□	□	□	□	□	□	□	□	□
18	High complexity	□	□	□	■	■	□	□	□	□	□
19	High need for good software architecture	□	□	□	■	■	□	□	□	□	□
20	Possibility of standardization	□	□	□	□	□	□	□	□	□	■
21	Iterative development	□	□	□	□	□	■	□	□	□	□
22	Mechanisms for rights management	□	□	■	■	□	□	□	□	□	□
23	Secondary role of performance	□	□	□	□	□	□	□	□	■	□
24	Preferability of a software implementation	□	□	□	□	□	□	□	□	■	□
25	Customer involvement in product development	□	□	□	□	□	■	□	□	□	□
26	User support during information processing	□	□	□	□	□	■	□	□	□	□
27	Customer involvement in value creation	□	□	□	□	□	■	□	□	□	□

Table 3.4: Mapping from economic properties of the software industry to relevant activity attributes. Full squares denore a relevant relationship, while blank squares denote a relationship which is not necessarily relevant or not relevant at all. Attribute descriptions can be found in Table 3.3.

thus allowing for a clear distinction between two extreme values. A Likert scale was not chosen, because it was desired that in unclear cases there would be explicitly no value assigned.

For the development of the attributes, three researchers designed questions with two extreme answering options. For instance, the question "Does the activity execution require deep IT understanding?" could be answered with "yes" or "no". Each question was then rephrased to a statement, e.g., "the activity execution requires deep IT understanding". The main goal was to find attributes which broadly cover all economic properties of the software industry while clearly covering the most important properties, such as Moore's Law and intangibility. Table 3.4 shows how each economic property is relevant to at least one of the attributes. For example, attribute C (multiple performance of an activity per customer) is related to economic property 4 (reproducibility), because an activity which is performed multiple times per customer is likely easier to replicate than an activity which needs to be adjusted on every execution. The only exception is attribute I, which isn't relevant to any economic property. This attribute was introducted nevertheless in order to reflect chronological order within the value chain (Porter, 1985, p. 48).

The introduced interlayer of attributes fulfills the previously stated requirements, because: (1)

the attributes reflect the economic properties of the software industry and are therefore relevant; (2) the attributes cover all underlying economic principles and are therefore complete; (3) all value ranges are binary, this improves the determinability of attributes because in many cases it is easier to evaluate based on two extremes than on a scale of gradually different values.

3.3.2 Measurement of activity attributes

Given the attributes to describe activities, they can be classified by assigning values to attributes. Though this could be done by a single researcher who is familiar with both, economic theories and the software industry, a Delphi study with twelve participants was used instead. This builds upon a broader understanding and abstracts from the views of a single researcher.

The Delphi study, being an iterative feedback technique among an expert panel, was developed at Rand Corporation in the 1950s (Dalkey and Helmer, 1963; Landeta, 2006). The main objective usually is to obtain a reliable consensus among a group of experts on a complex issue (Okoli and Pawlowski, 2004). Furthermore, the Delphi study allows follow-up interviews leading to a deeper understanding. Finally, it comprises a virtual panel of experts that can be contacted asynchronoously, thus allowing including experts from different locations. The structure of the applied Delphi study is derived from Okoli and Pawlowski (2004) and contains the following phases:

- Questionnaire design: The initial questionnaire was derived from the activity attributes by rephrasing them as questions. For each activity/attribute combination each participant was asked to choose betweend the two possible values and leave a comment justifying the judgement. The questionnaire includes a detailed description of all activities and their sub-activities as presented in Section 3.2.

- Pre-test: A pre-test was conducted with one participant to assure that activity descriptions and attribute values were understood correctly. This participant had similar level of domain knowledge to all the participants used in the following steps. He was excluded from further phases and his judgements were not included in the final results.

- Participant selection: In general, the number of participants involved should be within the frange of 10-18. The selected twelve participants were experts in the field of software industry and had a similar background and level of knowledge.

- Delphi round: The survey was rolled out to all participants in the format of a questionnaire and was returned within a given timeframe.

- Result analysis: Participants' answers and comments were analysed by the three moderators. For each activity/attribute combination a satisfactory level of agreement was reached if at least 80 % of all participants gave the same judgement. All comments were evaluated in order to identify misunderstandings of activites or attributes.

- Reiteration: Steps 4 and 5 were reiterated until the judgements reached a satisfactory degree of consenses and misunderstandings were resolved.

The resulting consensus values are shown in Table 3.5. It contains values where a minimum level agreement of 80 % could be reached after the final final round. All elements below this threshold are empty, indicating not available (NA) values. Since further analyzes work on available values

	Activity	Attribute value									
		A	B	C	E	F	G	H	I	J	K
1	Research	* 1	* 0	* 1			* 0	* 0	* 0	* 0	* 0
2	Development	1	* 1		* 0		* 0	* 0	* 0	* 1	* 0
3	Maintenance	* 1	* 1	* 0	* 0		* 0	* 0	* 1	* 1	
4	Production	* 0	* 1	* 1	* 1	* 1		* 1	* 0	* 1	* 1
5	Marketing		* 0	* 0	* 1		* 0	* 0	* 0	* 0	* 1
6	Implementation	1	* 1	* 1	* 0	0	* 1	* 0	* 0	* 1	
7	Education	* 1	* 1	* 0	* 1	* 0		* 0	* 0	* 0	* 1
8	Support	* 1	* 1	* 0		* 0		* 0	* 1	* 0	* 1
9	Operation	* 1	* 1	* 0	* 0	* 1		* 1	* 1	* 1	* 1
10	Replacement	* 1	* 1	* 1	* 0		* 1	* 0	1	* 1	

Table 3.5: The consensus matrix after the final Delphi round. Empty cells contain NA values and indicate elements where the consensus level among experts is below 80 %. Elements marked with an asterisk reach a consensus level of more than 90 %. The values 0 and 1 correspond to attribute value ranges defined in Table 3.3.

only, they are based on element values where the high 80 % level of consensus among experts could be reached.

The objective of this Delphi study was to find as many activity/attribute combinations with a satisfactory level of agreement as possible. In total, three Delphi rounds were carried out. During this process consensus could be reached on 84 % of all combinations. For others, participants' comments lead to the conclusion that no consensus could be reached. In most cases, this was due to the dependence of the judgement on product type or other assumptions which could not be pre-defined bacause of the requirement on the generic nature of the unified software value chain. For instance, for activity marketing and attribute A it was no possible to decide if the result was tangible or intangible. Whereas product placement can be tangible in case of a physically touchable product in a shop, in can be intangible as well if the product is marketed through the Internet (e.g., software as a service). Table 3.6 summarizes not decidable combinations and provides further descriptive details for each round.

In the second and third Delphi round, Kendall's W coefficient of concordance was used to provide participants with a qualitative assessment of consensus ranging from "very weak" to "very strong". This statistical method is often used in Delphi studies, particularly in the area of ranking-type Delphi studies (Schmidt, 1997). Further important changes include the deletion of attribute D as it was too redundant with attribute E. Finally, attribute K was changed after first round, because participants' comments indicated too broad scope of network effects.

	Round 1	Round 2	Round 3
Number of attributes	10	10	10
Number of elements at round start	110	100	100
Elements deemed as undecidable	20 (all elements of attribute D and $5A$, $2C$, $1E$, $9E$, $1F$, $2F$, $3F$, $8G$, $9G$, $10G$)	26 (6 in addition to round 1: $5F$, $6F$, $4G$, $3K$, $6K$, $7K$)	26 (no change compared to round 2)
Number of elements at round end	90	84	84
Number of participants questioned in this round	12	12	4
Total number of questions asked	1,320	376	6
Total Kendall's W	0.54	0.76	0.77
Number of NA values in the consolidated matrix	35	18	16
Number of decidable NA values (compared to the final result)	25	2	0
Number of updated judgements in this round	1,320	152	4

Table 3.6: Descriptive information after each Delphi round. Elements are identified through shortcuts such as $5A$, where 5 denotes the activity marketing and A the respective attribute.

3.3.3 Assertion of activity uniqueness

In order to assert the uniqueness of each activity, each of them must be shown to have own specifics when being compared to all other activities. An activity is defined as *unique* if there is no other activity which has exactly the same values in all attributes. Attributes, where at least one of the two activities has an NA value are not compared. This treatment of NA values is reasonable because an NA value is not necessary different from other values and is not necessarily equal to another NA value (Witten and Frank, 2005). This treatment of NA values makes the appearance of duplicates more likely when the share of NA values is high. It can be regulated by adjusting the threshold which is used to transform average votes to values in the consensus matrix. A low threshold results in fewer duplicates, but undermines the confidence in the values due to the low consensus levels. As a consequence, a high threshold should be chosen in order to ensure a sufficiently reliable proof of uniqueness within the value chain. For this study, a threshold of 80 % was used.

Based on this definition of uniqueness, each activity illustrated in the consensus matrix in Table 3.5 is unique. Furthermore, it appears that far less attributes are necessary in order to ensure the economic uniqueness of activities. Trying out all possible combinations of attributes, two minimal combinations could be derived: (B, G, H, I, J, K) and (B, C, G, H, I, J). Thus, there is no set of less than six activities which would lead to a unique software value chain. Also, attributes B, G, H, I and J appear to be more important than others, since they are contained in both sets, whereas attributes C and K are substitutes in this particular use case.

3.3.4 Construction of a value chain hierarchy

The appropriate level of granularity in a value chain may vary depending on current needs. For strategic questions a high-level view of few activities might be more suitable, whereas looking at processes requires a more detailed view. Starting with ten activities as defined in section 3, more coarse-grained levels of granularity can be provided by combining existing activities. Within this process two questions arise:

1. Which activities should be combined?
2. How to evaluate and choose between different levels of granularity?

It appears reasonable that the second questions cannot be answered for all cases, but some criteria can be provided to indicate which levels should not be used. With regard to the first question, the most similar activities should be combined, whereas the similarity between remaining activities should be low.

The combination of similar objects is called clustering. A generic bottom-up approach that starts with a single object per cluster and successively combines those to clusters is HAC (Chakrabarti, 2003). Two modifications have been made to the generic HAC algorithm:

1. It can deal with NA values.
2. Selected activities are actually merged to a new activity, resulting in a new set of activities.

The similarity measure between two activities is calculated as the number of all equal values, divided by all attributes including NA values (Finch, 2005). Due to the low value range of the similarity measure, an additional goodness criterion is used when there are multiple candidates for combination. For instance, this is the case on level 9 of the resulting hierarchy as shown in Figure 1. There the similarity value is 0.7 for the activities pairs (replacement, implementation) and (education, support). Therefore, the goodness criterion is used. It is defined as the average of absolute differences between all attribute values, whereas the attribute values are not 0 or 1, but the average consensus values obtained from the Delphi study. Thus, the goodness criterion differs from the similarity value in the usage of all attributes, as well as the attribute values.

The resulting hierarchy and describing statistics are shown in Figure 3.2. Starting from the bottom at level 10 with 10 activities, in each round two activities are successively combined. As a result, it can be seen that logically similar activities are combined. For instance, the implementation of a new system replaces an old system, such that similarities between these activities can be assumed. Education and support are both targeted at increasing user's capabilities with the software, thus there is compelling logic in the combination of those activities as well. The combination of operation and maintenance is reasonable, since software companies providing operation will usually take care of maintenance as well. The combination of R&D and marketing is likely to be caused by the fact that both activities differ from other activities by nature. The combination seems hard to justify from a logical point of view.

With regard to the appropriate number of activities, multiple criteria can be used. Their nature can be quantitative and qualitative. Four criteria are used: (1) similarity value of combined activities, (2) relative number of NA values in resulting consensus matrix, (3) uniqueness of activities, and (4) logical reasoning between combined activities. When a criterion significantly worsens between two levels, it is an indicator that a reasonable level of granularity has been reached before the combination between those two levels takes place. This "stepsize" rule was proposed by (Sokal and Sneath, 1963). A discussion of different rules can be found in (Milligan and Cooper, 1985).

The similarity measure decreases from level 4 to 3 by 0.2, whereas it only decreases by at most 0.1 in all previous steps. Therefore, it suggests that level 4 is preferable. The second criterion suggests level 6 or 4. However, level 5 and 6 contain duplicate activities, such that the third criterion discourages their usage. Finally, applying the fourth criterion, it appears that level 4 is unfortunate due to the combination of R&D and marketing. In conclusion, level 7 is recommended, resulting in a value chain of the activities (1) development, (2) replacement & implementation, (3) maintenance & operation, (4) education & support, (5) production, (6) marketing, (7) R&D. For more coarse-grained granularity level 4 should be used.

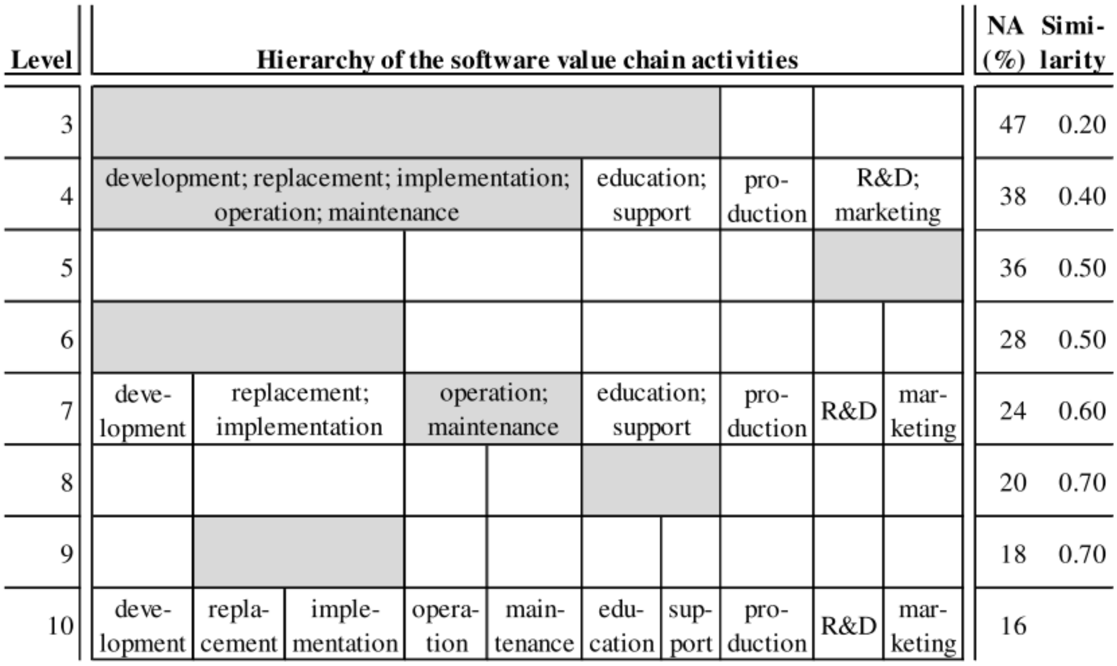

Figure 3.2: Resulting hierarchy of activities after each round. Starting with ten activities (at the bottom), in each round two activities are successively merged to a new activity. The resulting combined activity is highlighted in grey. Activity labels are not shown for non-desirable hierarchy levels.

3.4 Conclusion

The value chain is a widely applied concept and theoretical foundation of firm strategies as well as business models. This section methodically developed a value chain for software firm and software industry analysis. It therefore provides answers to Research Question 2 regarding value creating activities of software firms. The results suggest that software firms create competitive advantage by combining the following ten activities: (1) research, (2) development, (3) maintenance, (4) production, (5) marketing, (6) implementation, (7) education, (8) support, (9) operations, and (10) replacement. This representation of the software-specific value chain fulfills multiple desirable requirements as ensured by the developed and applied method: (1) the value chain contains all relevant activities of software firms, (2) it ranges from the very beginnings of the product to its end, (3) the activities are economically unique, and (4) they have an appropriate level of granularity. Further addressed requirements include low coupling and high cohesion. Building upon the detailed value chain representation with ten activities, two further, more coarse-grained, representations were derived which fulfil the requirements and can be used for more abstract analyzes.

The developed value chain concept is used in the subsequent empirical studies to measure activities of software firms as components of firms' strategies and business models. The empirical studies also provide insights into the applicability of the developed concept as a research framework for software firms and the software industry. The empirical results show how software firms configure their value chains and which activities have more significant implications than others.

There are limitations to be taken into account when applying the suggested framework. The granularity of activities and frontiers between them are up to individual judgement of the researcher, as only generic guidelines can by applied to such high-level concepts. The developed hierarchy of activities suggests different appropriate levels of granularity, however, researchers might choose different representations to tune the framework to their problem at hand. As pointed out by (Porter, 1985, p. 45), the final decision should depend on the goal of analysis and the target group. As this leaves room for individual judgement, it may lead to varying operationalizations. This, in turn, may handicap the comparability of results.

Section 4

Study 1: Business models of the top 120 public software firms

The previous sections introduced strategic groups and business model configurations as two related concepts under the umbrella of configurational research. Methods from strategic group analysis were discussed, combined, and extended in order to derive a common methodology for the analysis of configurations in general and business model configurations as well as strategic groups in particular (see Section 2.3.2). The methodology comprises multiple steps, most of which are often applied in strategic group research, but only few of these steps have found their way into studies on business models. This opens up the question which business model configurations and effects can be found in the software industry by following the suggested methodology.

This empirical study applies the configurational analysis methodology to analyze the business models of the 120 largest public firms in the US. The results are used for testing the hypotheses discussed in Section 2.5 and for explorative analyzes to address the research questions discussed in Section 1.1. The findings suggest that there are three prevalent business models with significant differences in performance and risk. The most successful business model exhibits a broader market scope than its alternatives, which may be explained with the exploitation of network effects.

The remainder of this section is organized as follows. Section 4.1 describes the data collection and application of the configurational analysis methodology. The results are presented in Section 4.2 and discussed in Section 4.3. Finally, Section 4.4 concludes the study.

4.1 Data and method[1]

This section describes the data collection and application of the methodology as shown in Figure 2.5 (p. 23). The contents are organized according to the methodology steps two to seven.

4.1.1 Sample

The sample firms were selected for the year 2010 from the Bureau van Dijk Orbis database based on relevant Standard Industry Classification (SIC) codes. Following previous studies (Lavie, 2007; Léger and Quach, 2009; Izci and Schiereck, 2010) on the software industry, a firm was included in

[1]Parts of this section were published in Schief et al. (2013).

the sample if its primary SIC code was one of the following: 7371 (computer programming services), 7372 (prepackaged software), 7373 (computer integrated systems design), or 7374 (computer processing and data preparation and processing services).

The selection was limited to public firms listed at stock exchanges in the US. These firms are required to publish 10-K and 20-F annual reports, which were used in this study to obtain information on the firms' business models via expert classification. Additional financial data could be obtained from the Thomson Financial Worldscope database. The selection was further limited to firms with no single stakeholder holding a direct stake over 50 percent. These firms are less exposed to external control, such that the firms' performance can be better attributed to their business model choices.

From the overall sample, the 120 firms with the highest revenues were selected. This limitation was necessary because the following expert classification was very time consuming and could only be performed on a limited number of firms.

4.1.2 Business model variables

This study uses the business model components of the SBMF (see Section 2.4.1) as indepedent variables, they are referred to as business model variables. Because the components in the SBMF are too numerous and their value ranges too detailed, the data collection had to be limited to eleven variables. The variables and details on their measurement are shown in Table 4.1. It shows that seven variables required expert classification of annual reports, while only four variables could be measured based on data from the Thomson Financial Worldscope database. For some variables no continuous measure could be provided. In these cases, a decision rule was formulated which defines a binary value space for the respective variable, e.g., the variable value chain is not measured in terms of the percentage of service revenues but is a binary variable, where the value 0 denotes firms mostly providing services while the value 1 denotes firms mostly providing products.

The variables used in this study are specific to the software industry in at least one of the following aspects: (1) The variable deals with an aspect that may not be relevant to other industries (e.g., license model). (2) The definition and terminology of options are highly industry-specific (e.g., software stack layer). (3) The assignment rules need to be specifically formulated for the industry (e.g., for the variable value chain, all potential software and associated products and services need to be considered and documented in the assignment rule). (4) Expert knowledge is required to conduct the classification as the nature of software needs to be explored (e.g., analyze from which target customer a solution can be used and examine the distribution of associated revenues).

#	Component (variable)	Values	Method	Measure or decision basis
1	Value chain	Binary: 0=service, 1=product	Expert judgement of annual report; 10-K item 1, 6, 7, and notes	Decision based on the percentage of firm revenues from products (e.g., licenses, physical products, hardware) and services (e.g., maintenance, support, consulting, training, SaaS)
2	Degree of vertical integration	Metric	Calculation on secondary data from Thomson Financial	Value added (sales minus costs of goods sold) divided by sales (Hutzschenreuter and Gröne, 2009, p. 282)
3	Sales volume	Metric	Calculation on secondary data from Thomson Financial	Natural log of firm revenues (Short et al., 2007, p. 155)
4	Revenue source	Binary: 0=direct, 1=third party	Expert judgement of annual report; 10-K item 1, 7, and revenue recognition	Decision based on the percentage of revenues from direct payments (incl. subscriptions) and third party payments (e.g., from advertising)
5	Payment flow structure	Metric	Calculation on secondary data from Thomson Financial	Sum of deferred short- and long-term revenues divided by total revenues, according to Dechow and Skinner (2000), deferred revenues can be related to subscription-based and hybrid revenue models
6	Software stack layer	Binary: 0=infrastructure, 1=application	Expert judgement of annual report; 10-K item 1, 6, 7, and notes	Decision based on the percentage of revenues from infrastructure software and application software, definitions are based on Forward and Lethbridge (2008): infrastructure=A.des, B, and C, rest is application software

Table 4.1 – continued from previous page

#	Component (variable)	Values	Method	Measure or decision basis
7	License model	Binary: 0=proprietary, 1=open source	Expert judgement of annual report; 10-K item 1, 1a, 7, and notes	Decision based on the percentage of revenues related to open source and proprietary products
8	Localization	Metric	Calculation on secondary data from Thomson Financial	Percentage of domestic sales divided by total sales
9	Target customer	Binary: 0=consumers, 1=businesses	Expert judgement of annual report; 10-K item 1, 7, and notes	Decision based on revenues obtained from businesses and consumers
10	Target industry	Binary: 0=few, 1=broad	Expert judgement of annual report; 10-K item 1	Decision based on the number of targeted industries few (less than 4) and many (at least 4)
11	Channel	Binary: 0=direct, 1=indirect	Expert judgement of annual report; 10-K item 1a and notes	Decision based on revenues obtained from direct and indirect sales, when no numbers available, words like “primary”, “substantial”, and “significant” were sometimes indicative

Table 4.1: Operationalization of business model components.

For the expert classification, the annual reports of all firms were obtained for the year 2010 from their websites. Three experts independently classified firms based on the annual reports. Each of the 120 firms was classified by two experts, such that each expert had to classify 80 firms. Given that seven variables had to be extracted from the reports, each expert had to perform 560 classifications. Following the individual classifications, the results were consolidated. For each firm and each variable, two classifications had to be consolidated. In total, there were 840 disagreements possible, but disagreements only occurred in 151 cases. Thus, the initial coding agreement was over 82 percent. Detailed measures of inter-rater agreement are given in Table 4.2. Overall, the rater agreement was considerably high, as indicated by values of Cohen's Kappa around 0.7 (Finch, 2005, p. 91). Final coding agreement was reached by discussing cases where disagreements occurred, resulting in a final agreement rate of 100 percent.

	Experts			**Total**
	1 and 2	**1 and 3**	**2 and 3**	
Total classifications	280	280	280	840
Consistent classifications	230	228	231	689
Consistent in percent	82.1	81.4	82.5	82.0
Cohen's Kappa	0.71	0.68	0.71	

Table 4.2: Inter-rater agreement of the expert classification.

The descriptive statistics of the variables representing business model components are shown in Table 4.3. The variable value chain indicates that 77 percent of the firms mainly generate their revenues with services rather than with products. The related activities are mainly performed inhouse, as is indicated by the variable degree of vertical integration, i.e., 61 percent of the revenue are due to the value added generated by the sample firms rather than by contractors. The variable sales volume is given as the natural logarithm of sales in United States dollars (USD) millions and shows a range from 132 million to over 28 billion. Regarding the variable revenue source, all firms in the sample obtain most revenues directly from their customers rather than from third parties. The variable payment flow structure shows that the average ratio of deferred to total revenues is 19 percent. This indicates a tendency towards up-front rather than recurring payments. 62 percent of the firms mainly generate their revenues with application software rather than infrastructure software. The variable license shows very little variation as ony one firm mainly uses open source licenses while all other firms mainly use proprietary licenses. 68 percent of the firm revenues are generated domestically rather than abroad, indicating a tendency towards the local markets. The variable target customer shows that 95 percent of the firms target business customers rather than private customers. 61 percent of the firms mostly serve a broad range of industries rather than having a narrow industry focus. Finally, 15 percent of the firms mainly use indirect sales channels rather than direct channels.

The following three business model variables could not be used as independent variables in the cluster analysis because they showed too little variance: (1) revenue source, (2) license, and (3) target customer. Further, in order to remove suspicious values and reduce the influence of outliers on cluster analysis, a data point of a metric variable was removed if it deviated from the mean by more than three standard deviations (Andersen et al., 2007, p. 417).

Variable	N	Mean	Median	Min	Max	SD
		Business model variables				
Value chain	110	0.77	1.00	0.00	1.00	0.42
Degree of vert. int.	118	0.61	0.63	0.10	0.98	0.24
Sales volume	116	6.44	6.26	4.89	10.26	1.13
Revenue source	117	0.00	0.00	0.00	0.00	0.00
Payment flow str.	114	0.19	0.13	0.00	0.86	0.19
Software stack layer	104	0.62	1.00	0.00	1.00	0.49
License	108	0.01	0.00	0.00	1.00	0.10
Localization	117	0.68	0.69	0.00	1.00	0.26
Target customer	115	0.95	1.00	0.00	1.00	0.22
Target industry	116	0.61	1.00	0.00	1.00	0.49
Channel	92	0.15	0.00	0.00	1.00	0.36
		Performance and risk variables				
OPM	115	12.07	10.62	-17.15	42.94	10.48
ROA	117	5.66	5.64	-13.74	26.61	6.83
Growth (sales)	116	11.47	10.00	-24.39	47.59	12.40
PBR	115	3.95	3.09	-1.03	16.85	3.07
PER	115	26.34	22.82	-243.17	440.13	59.91
OPM variance	117	17.15	5.12	0.05	395.38	41.35
ROA variance	114	27.39	5.37	0.01	325.91	55.53
Growth variance	117	249.05	79.89	2.02	1,765.91	363.63
PBR variance	114	2.07	0.33	0.00	42.71	5.91
PER variance	113	15,401.69	131.25	0.05	593,177.73	77,798.59
Risk-adj. OPM	116	8.19	1.56	-0.54	71.98	15.06
Risk-adj. ROA	117	6.64	0.67	-0.56	111.12	16.17
Risk-adj. growth	116	0.20	0.06	-0.51	2.82	0.40
Risk-adj. PBR	115	39.87	8.49	-5.84	1,058.33	112.52
Risk-adj. PER	113	4.11	0.13	-0.15	62.42	10.71

Table 4.3: Descriptive variable statistics. Values are shown after data cleansing of all metric variables by removing values which deviated from the mean by more than three standard deviations (Andersen et al., 2007, p. 417).

The variables showed only limited correlations, such that no variables had to be removed because of collinearity issues. Table 4.4 shows the correlations between all independent variables in the study. While all other correlations are well below 0.5, the correlation between the variables degree of vertical integration and payment flow structure is 0.62. However, the value is not too high and a subsequent analysis of the VIF did not indicate any problems with multicollinearity. The maximum VIF is 1.4 between the variables value chain and degree of vertical integration, thus well below the threshold of 10. The condition number is 2.2, thus well below 30.

4.1.3 Success variables

Multiple variables were used to measure financial firm performance: operating profit margin (OPM), ROA, sales growth, price / book ratio (PBR), and PER. The five variables measure different dimensions of firm performance. While OPM, ROA, and growth are accounting-based measures, the variables PBR and PER are market-based measures. Data on each variable was

	Degree of vert. int.	Sales volume	Payment flow str.	Software stack layer	Localization	Target industry	Channel
Value chain	-0.15	0.14	-0.04	0.36	0.16	-0.10	-0.44
Degree of vert. int.		0.10	0.62	-0.26	-0.29	0.15	0.25
Sales volume			0.07	0.04	-0.42	0.05	0.06
Payment flow str.				-0.17	-0.17	0.15	0.19
Software stack layer					0.25	-0.33	-0.29
Localization						-0.30	-0.22
Target industry							0.11

Table 4.4: Correlations between business model variables which were used as independent variables for cluster formation. For each value, the Pearson and the Spearman correlation coefficients were calculated. The higher value was included in the table.

obtained from the financial database for the years 2009 to 2011. The performance was calculated as the mean of these years to smooth out short-term trends and account for a certain time lag between firm action and performance impact (Mehra, 1996, p. 313). For each of the five variables, a risk measure was calculated as the variance of the time period at hand. Also, for each variable, a risk-adjusted measure was calculated by dividing the performance measure with the respective risk measure.

The descriptive statistics of the success variables are included in Table 4.3. The statistics indicate that the sample firms are very profitable, with an average OPM of 12.07 and ROA of 5.66. The sample firms are growing at 11.47 percent. The market values the firms at 3.95 times the firms' book value and 26.34 times the firms' earnings.

4.1.4 Analysis

Due to data restrictions, a test for stable strategic time periods, as suggested by the methodology, could not be performed. Such a test would have required collecting data on all business model variables for multiple periods, which was not possible given the labor-intensive classification procedure.

The business model configurations were derived inductively using the two-stage clustering procedure. Because the data is not normally distributed, non-parametric methods were used. Within the first stage the optimal number of configurations and their medoids were determined using the Mojena criterion. Within the second stage configurations were formed using the k-medoids method. Three configurations were obtained with 62, 32, and 24 firms. The descriptive statistics of the business model configurations are shown in Table 4.5.

The clustering results were validated by calculating the APN for each business model variable. The descriptive statistics are shown in Table 4.6. The results show values that for each variable

Variable	**Configuration**		
	BMC1	**BMC2**	**BMC3**
Number of firms	62	32	24
Business model variables			
Value chain	0.71	0.86	0.83
Degree of vertical integration	0.64	0.59	0.56
Sales volume	7.03	5.90	5.70
Revenue source	0.00	0.00	0.00
Payment flow structure	0.22	0.16	0.14
Software stack layer	0.33	0.89	1.00
License	0.02	0.00	0.00
Localization	0.61	0.78	0.75
Target customer	0.97	0.94	0.92
Target industry	0.78	0.00	1.00
Channel	0.25	0.03	0.11
Success variables			
OPM	14.48	11.73	6.13
ROA	6.38	5.51	3.97
Growth	9.64	13.06	14.08
PBR	3.91	4.06	3.89
PER	21.56	37.61	23.82
OPM variance	11.15	25.33	21.95
ROA variance	20.27	30.49	42.26
Growth variance	248.44	182.51	339.32
PBR variance	1.68	2.75	2.16
PER variance	10,837	14,342	29,504
Risk-adjusted OPM	9.97	8.79	2.90
Risk-adjusted ROA	8.32	6.83	2.11
Risk-adjusted growth	0.21	0.22	0.17
Risk-adjusted PBR	28.41	44.21	62.95
Risk-adjusted PER	4.05	4.90	3.26

Table 4.5: Descriptive business model configuration statistics.

30 to 40 percent of the firms would be assigned to a different cluster if the variable was removed. With eight business model variables, the presumably optimal APN value for each variable would be 0.13. Though the obtained values are much higher, they indicate that no single variable dominates the formation of configurations as all variables are about equally important.

Mean	Median	Min	Max	SD
0.36	0.37	0.30	0.40	0.58

Table 4.6: Descriptive statistics of the APN measure.

Tests for differences across configurations were carried out to identify idiosyncraties of configurations regarding business model and success variables. For each dependent and independent variables, the Kruskal-Wallis test was performed to test for significant variation across configurations. A post-hoc analysis was performed to identify significant differences between each pair of configurations. The results are shown in the following section.

4.2 Results

Table 4.7 shows the statistics of all variables that vary significantly across configurations. Four business model variables have significance levels below 0.1 percent and the variable channel has a significance level below 5 percent. Thus, there are structural differences across business model configurations, which confirms Hypothesis 1. The pair-wise comparisons of the posthoc analysis confirm these differences and additionally allow a more differentiated view, because they indicate individual configurations that are significantly different.

The configuration BMC1 is unique in all business model variables with the exception of the variable channel. The sales volume is the highest across all configurations, indicating that BMC1 consists of the largest firms. The low value in the variable software stack layer indicates a focus on infrastructure software. The value in the variable localization is low in comparison to the other configurations, indicating a stronger focus on international markets. The value of 0.78 for the variable target industry is much higher than for BMC2, indicating a focus on a broad set of industries. However, the value is lower than for BMC3. The variable channel is the highest across all configurations, indicating a stronger focus on indirect sales channels, though the differences towards BMC3 is not significant.

Configuration BMC2 has lower values in sales volume than BMC1, thus indicating that firms are smaller in comparison to BMC1. The value 0.89 in the variable software stack layer indicates a strong focus on application software as opposed to infrastructure software in BMC1. The variable localization indicates a strong focus on the domestic market when compared to BMC1. The variable target industry shows that firms in BMC2 have a narrow industry focus in comparison to all other configurations. Firms in BMC2 have further a strong focus on direct sales channels in comparison to BMC1.

Variable	Conf. mean			Conf. diff.	Pair-wise conf. diff. (p-values)		
	BMC1	BMC2	BMC3	(p-value)	BMC1/BMC2	BMC1/BMC3	BMC2/BMC3
Business model variables							
Sales volume	7.03	5.90	5.70	0.0000 ***	0.0000 ***	0.0000 ***	0.4179
Software stack layer	0.33	0.89	1.00	0.0000 ***	0.0000 ***	0.0000 ***	0.3472
Localization	0.61	0.78	0.75	0.0006 ***	0.0020 **	0.0285 *	0.8399
Target industry	0.78	0.00	1.00	0.0000 ***	0.0000 ***	0.0420 *	0.0000 ***
Channel	0.25	0.03	0.11	0.0364 *	0.0484 *	0.5997	1.0000
Success variables							
OPM	14.48	11.73	6.13	0.0063 **	0.9288	0.0049 **	0.1199
ROA	6.38	5.51	3.97	0.0401 *	0.7071	0.0452 *	0.4882
PER variance	10,837	14,342	29,504	0.0826 †	1.0000	0.0944 †	0.2034
Risk-adjusted OPM	9.97	8.79	2.90	0.0120 *	0.4380	0.0110 *	0.5038
Risk-adjusted ROA	8.32	6.83	2.11	0.0510 †	1.0000	0.0612 †	0.2890

Table 4.7: Main results of the study. Only significant variables are shown. Columns 2-4 show the mean variable values for each cluster. Column 5 shows the significance level obtained with the Kruskal-Wallis test across all clusters. Columns 6-8 show the pair-wise differences between clusters (post-hoc analysis with Bonferroni correction).
Significancies: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; †$p < 0.1$

Configuration BMC3 is very similar to BMC2 regarding the variables sales volume, software stack layer, localization, and channel. There are no significant differences between the two configurations in these variables. The only significant difference is in the variable target industry which indicates that firms in BMC3 have a broad industry scope in comparison to BMC2.

The results of the success variables representing performance and risk show significant differences across configurations. Performance differences are only present in accounting-based measures, whereas risk differences are only present in the market-based variable PER variance. However, the accounting-based measures show significant differences across clusters after adjustment for risk. Overall, these results confirm Hypothesis 2, 3, and 4. The pair-wise comparisons indicate that all performance and risk differences are found between the configurations BMC1 and BMC3. There are no significant differences between BMC1 and BMC2 as well as BMC2 and BMC3. BMC1 shows better performance and risk in all variables, indicating that firms in BMC1 are more successful than firms in BMC3. Though not significant, the descriptive results in the dependent variables are better in BMC1 than in BMC2 and the figures are better in BMC2 than in BMC3.

Finally, the results confirm Hypothesis 5, as BMC3 has the lowest values in all performance variables and the highest values in all risk variables, BMC2 is in the middle in all variables, and BMC1 exerts the highest performance and the lowest risk. Thus, the results confirm all hypotheses in the thesis.

4.3 Discussion

The application of the configurational analysis methodology exerted three dominant business model configurations in the sample. As eight business model variables were used as structural characteristics, a larger number of configurations was theoretically possible. This confirms the general assumption of configurational research, stating that the overall solution space is often reduced to few configurations in practice. In terms of software business models this means that out of the many theoretically possible configurations of software business model components, only three configurations dominate the software industry.

The first business model configuration BMC1 consists of the largest firms in the sample. Looking at the values of the other variables for this configurations, it can be seen how those enable the larger firm sizes in comparison to the other configurations. As the software stack layer indicates, firms in BMC1 are dominantly engaged in infrastructure products. Because infrastructure products are used as the foundation to run application products, they can be reused across multiple application products. Consequently, firms engaged in infrastructure have a larger market scope and can thus grow larger than firms engaged in application products. The component localization shows that firms in BMC1 derive more revenues from international markets than firms in the other two configurations. These firms thus have access to a larger geographic market, again allowing for a larger firm size. In a similar vein, firms in BMC1 target a broad range of industries in absolute terms, again having access to a larger market than firms targeting few industries, such as firms in BMC2. BMC1 further relies more on indirect channels than firms in configuration BMC2, which allows them to scale their sales operations better than through direct sales, which is necessary in order to serve the larger market scope.

The configurations BMC2 and BMC3 are very similar in all structural characteristics except target industry where BMC2 targets fewer industries while BMC3 has the maximal possible value on the scale used. Though these two configurations differ in just one variable, the difference appears meaningful, because both configurations offer application products which can be very industry-specific (e.g., financial software) or generic (e.g., office software), whereas infrastructure products will be more often generalizable across multiple industries (e.g., antivirus software). It can be said that BMC2 has the lowest market scope of all configurations, because it serves the lowest number of industries and no more geographic locations than BMC3. Meanwhile, BMC3 serves as many industries as BMC1, but less geographic locations. Both configurations, BMC2 and BMC3, consist of the smaller firms in the sample, though it should be kept in mind that those firms belong to the 120 largest US software firms. As argued above, the smaller firm size in BMC2 and BMC3 can be explained by the positive relationship between a firm's market scope and its size.

The results further allow for an interpretation why some configurations are not prevalent in the software industry. As mentioned in the previous paragraph, configurations providing infrastructure software have less need to limit themselves to few industries, whereas configurations providing application software can choose between few and many industries, depending on the application software at hand. This could explain why there are no configurations that provide infrastructure software but target few industries.

The obtained configurations further suggest that configurations providing infrastructure software target more international markets than configurations targeting domestic markets. A possible explanation is that application software is closer to the end-users than infrastructure software, because end-users will often use the application product directly (e.g., office software). Intimate knowledge of the end-user will give application software providers a competitive advantage. Arguably, domestic firms possess more intimate knowledge of the end-users due to cultural and linguistic closeness than firms from abroad. This could explain why there are no (1) configurations providing application software and targeting international markets to the same degree as configurations providing infrastructure software and (2) no configurations providing infrastructure software and targeting domestic markets to the same degree as configurations providing application software.

Looking at the absolute values of the business model variables, there is little usage of indirect channels and high targeting of domestic markets. The highest usage of indirect sales channels is in BMC1 with 25 percent of firms regarding indirect channels as their main sales channels. An explanation could be that software, as a digital good, can be distributed at low costs via direct channels, such that most software firms choose to omit intermediaries. As for the high share of revenues derived from domestic markets, the result is somewhat surprising, because the economic properties of software suggest that software can be distributed at low costs. However, it appears that geographic closeness is important, as all configurations derive more than 60 percent of their revenues from domestic markets.

The success variables indicate that BMC1 outperforms BMC3 in terms of performance and risk-adjusted performance. There is further slight evidence of outperformance in terms of risk. Though no significant differences are found between BMC1 and BMC2, the absolute values show that BMC1 is the more successful configuration. It appears that the more successful business model configurations in the software industry are the ones with the larger market scope. This seems plausible, because software has the economic property of high scalability, such that software firms with higher revenues can expect better performance in general. Larger market scope and size can be further associated with the exploitation of network effects.

The negative relationship between performance and risk confirms Bowman's paradox. However, the evidence is weak, as the only significant risk measure is PER variance. The significance of this risk measure is just below the lowest significance level, though a larger sample could make the differences more significant.

Looking at the performance variables, it is surprising to see that configurations differ in terms of accounting-based measures but not in terms of market-based measures. It can be assumed that market-based measures are influenced by more factors than accounting-based measures, because they will be affected by the same factors as accounting-based measures as well as market factors (e.g., rumors of an acquisition). For instance, increased product sales will affect accounting- and market-based measures, but a recommendation of a stock analyst will only affect market-based measures. Thus, analyzes of market-based measures have either to control for the additional influences or use larger samples in order to cancel out the influences of uncontrolled variables. With the given results, it cannot be ruled out that market-based differences exist, though the empirical evidence is practically non-existent in the given sample.

With the exception of one step, being the test for stable strategic periods, all steps of the configurational analysis methodology could be applied to the concept of business models of software firms in order to obtain business model configurations. However, high data requirements of the methodology and the business model concepts resulted in very labor-intensive collection of data. As a result, the data could be obtained, but for a limited sample and only one independent period. Moreover, the classification was only made possible by reducing most scales to binary measures. Though more fine-grained scales would have certainly led to more differentiated results, the binary scales were sufficient to demonstrate differences between configurations and gain insights into prevalent business model configurations in the software industry.

4.4 Conclusion

This study applied the configurational analysis methodology to business models. Three meaningful business model configurations could be derived. The results suggest that a positive relationship exists between the market scope of a configuration and its success in terms of performance and risk. The most successful configuration provides infrastructure software, targets international markets and a broad range of industries, makes more use of indirect sales channels, and consists of larger firms than alternative configurations.

This study contributes to configurational and business model research in multiple ways. It builds upon the largest number of independent variables used in configurational business model research thus far, which allows for deeper insights. Also, the methodology used follows much closer the state-of-the-arts methods from strategic group analysis. From a methodological point of view, the study confirms that the configurational analysis methodology can be applied to business models to provide meaningful business model configurations. Researchers and practitioners are provided with empirical evidence that business models targeting a broad market scope are more successful than alternatives targeting a narrow market scope. This suggests that the exploitation of network effects is a critical factor determining success in the software industry.

There are several limitations to this study. Due to data limitations, no test for stable strategic periods could be performed. The sample is further limited to the 120 largest public US firms. Further, the business model variables used to form configurations cover only about half of all SBMF components. Finally, no comparison with strategic groups was performed on this dataset.

Section 5

Study 2: Strategies in the prepackaged software industry

The first study in Section 4 applied the configurational analysis methodology to US software firms, providing insights on business model configurations and their effects on firm performance as well as risk. Though the methodology has its roots in strategic group analysis with many applications to firm strategies in extant research, there are white spots to be closed. The literature review in Section 2.3 showed that the software industry has not yet been subject of detailed configurational analysis and only few studies include risk as a dimension of success. Closing both white spots is relevant to this thesis in order to provide an analysis of the software industry and establish a comparability with the results obtained on business models of software firms.

This study applies the configurational analysis methodology to identify strategic groups in the software industry and analyze their impact on firm success, i.e., performance and risk. The sample comprises public firms worldwide that mainly provide prepackaged software. The large and focused sample allows for an application of the most extensive application of the methodology. The results reveal significant differences across groups in performance and risk. The factors affecting strategic group success are further consistent with those found on business model configurations in Study 1. There is strong evidence that firm size and broader scope have a positive impact on performance and risk, which can be explained with the economic properties of the software industry, such as network effects as well as high economies of scale and scope.

The remainder of this section is organized as follows. Section 5.1 describes the data collection and application of the configurational analysis methodology to firm strategies. The results are presented in Section 5.2. The implications and limitations are discussed in Section 5.3. Finally, Section 5.4 concludes the study.

5.1 Data and method

This section describes the selection of the sample firms, dependent and independent variables, identification of stable strategic periods, the application of clustering methods to derive strategic groups, and methods for hypotheses testing as well as exploratory analysis. The contents are organized according to the steps two to seven of the configurational analysis methodology as shown in Figure 2.5 (p. 23).

5.1.1 Sample

The sample was drawn from the Thomson ONE Banker database. Firms were selected which had at least 70 percent revenues in segments with the SIC code 7372 (prepackaged software). Diversified firms are thus excluded from the sample to avoid problems with statistical noise and difficulties in deriving conclusions from a heterogeneous sample (Short et al., 2007). Sample firms were further required to be public, because more data was available on these firms. Both, active and inactive firms were selected, thus avoiding survivorship bias. The sample was not limited to any particular geographical region. The final sample includes 758 prepackaged software firms in the timespan 2000 to 2011. The sample is larger and covers a longer period than in most studies reviewed in this thesis, which allows for an application of the most extensive configurational analysis methodology.

5.1.2 Strategic variables

This study uses the common inductive approach to conceptualize competitive strategy with strategic variables that can be partitioned in two dimensions: scope and resource commitment. The strategic variables function as independent variables in subsequent quantitative analyzes.

Two characteristics of scope commitment were included: (1) *Geographic scope* is measured as the percentage of international sales divided by total sales (Cool and Schendel, 1987). Firms with high domestic sales are more focused on local markets and are less likely to operate globally. (2) *Competitive scope* is measured as the number of SIC codes where a firm generates revenues. Whereas firms were selected which generate the major share of their revenues with prepackaged software, the total number of codes is an indicator of firm diversification and risk-aversion. This measure is known for its weaknesses as a proxy for diversification, but it has been found appropriate when only a “gross type of diversification” control is required (Hoskisson et al., 1993).

Four characteristics of resource commitment were included: (1) *Physical resources* were measured as the percentage of capital expenditures divided by total sales (Short et al., 2007). Firms with high capital expenditures rely on own technology, plant, and equipment. (2) *Financial resources* were measured as current assets divided by current liabilities (Short et al., 2007). High ratios indicate high liquidity, ability to pay bills, and ability to engage in mergers and acquisitions. (3) *Organizational resources* are measured as the natural logarithm of total sales (Cool and Schendel, 1988). High values indicate large organizational resources and can influence market power and economies of scale as well as scope. This measure can be viewed as a characteristic of scope commitment as well (Ferguson et al., 2000). (4) *Research and development (R&D) resources* are measured as the percentage of R&D expenditures divided by total sales (Short et al., 2007). High R&D expenditures indicate the pursuit of opportunities in new markets and technologies.

For each firm the strategic variables were averaged over five-year periods. This reduces volatility by smoothing out short-term trends (Mehra, 1996), as strategic positions should be long-term decisions. In each period, averaged values were only used for analyzes if data was available in all five years, otherwise the value was regarded as not available. Other cleansing procedures included the removal of obviously erroneous and logically impossible values, e.g., negative R&D to sales ratios. Further, for each firm and for each variable, values were removed which deviated from the mean by more than three standard deviations (Andersen et al., 2007).

Firm strategies can be altered and, therefore, group membership can change over time. However, if strategies are too volatile in a period of time, one cannot determine meaningful strategic groups and no meaningful conclusions can be drawn from such time periods. Thus, as pointed out in Section 2.3.2, stable strategic time periods need to be identified. Stable periods must fulfil two requirements in all strategic variables: homogeneity of variances and stable means over the time period.

For tests of homogeneity of variances the Fligner-Killeen test was applied as a non-parametric alternative to the parametric Bartlett test, because the data was clearly not normally distributed. For tests of stable means the Kruskal-Wallis rank sum test was applied. The results were confirmed with the parametric Hotelling's T^2 test. The test results for stable strategic time periods are shown in Table 5.1. The results are insignificant in all tests for the period 2004–2008, thus indicating it to be a stable strategic time period. In all other periods, at least one test was significant, indicating instability in at least one strategic variable. As a result, the one stable strategic period 2004–2008 was used for further analyzes of strategic groups. Though unstable periods are not reliable indicators, it should be mentioned that the results obtained from instable periods did not challenge the results obtained from the only stable period in this study.

Strategic period	**Kruskal-Wallis test**			**Hotelling's T^2 test**		**Fligner-Killeen test**		
	Statistic		**p value**	**Statistic**	**p value**	**Statistic**		**p value**
2000-2004	0.07	***	0.0000	2.11	0.9097	0.27	***	0.0000
2001-2005	0.53	***	0.0000	3.46	0.7493	0.41	***	0.0000
2002-2006	0.47	*	0.0245	2.02	0.9179	0.39	***	0.0001
2003-2007	0.05		0.1574	1.59	0.9530	0.57	*	0.0324
2004-2008	0.81		0.1703	1.46	0.9618	0.45		0.3215
2005-2009	0.76	**	0.0052	2.61	0.8559	0.69		0.1221

Table 5.1: Study 2: Stable strategic time periods test results.
Significancies: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; †$p < 0.1$

The strategic variables were tested for collinearity and multicollinearity, because highly correlated variables have a negative impact in clustering procedures. The correlations are shown in Table 5.2. The maximal absolut correlation is -0.34 between the variables organizational resources and R&D resources. This value is considerably low, thus not indicating any problems with collinearity. The maximum VIF is 1.3 between the variables financial resources and organizational resources, thus well below the maximally acceptable value of 10. The condition number is 1.8, thus well below the threshold of 30. In summary, no indications of collinearity or multicollinearity could be found in the data, such that all variables could be used for further analyzes.

Table 5.3 shows the descriptive statistics of the strategic variables for the sample firms in the stable period 2004-2008. The sample size is 133. The numbers show that the sample firms generate 34.81 percent of their revenues from international markets, thus emphasizing the importance of domestic markets to software firms, which is consistent with the data from Study 1. The variable competitive scope indicates that most firms focus on prepackaged software and less than half of the sample firms engage in a different field as well. The physical resources indicate that less than four percent of firm revenues are used for capital expenditures, thus emphasizing the low importance of physical resources to software firms. The variable financial resources shows that on average the sample firms hold more than two times their current liabilities as current assets, thus indicating high liquidity. The variable organizational resources indicates that the sample firms are considerably smaller than the sample firms in Study 1. Finally, the variable R&D resources indicates that the sample firms spend more than 17 percent of their revenues on R&D.

	Competitive scope	Physical resources	Financial resources	Org. resources	R&D resources
Geographic scope	-0.02	-0.11	0.08	0.33	0.00
Competitive scope		-0.01	-0.04	0.32	-0.24
Physical resources			0.17	0.16	0.07
Financial resources				0.03	0.21
Organizational resources					-0.34

Table 5.2: Correlations between strategic variables which were used as independent variables for cluster formation. For each value, the Pearson and the Spearman correlation coefficients were calculated. The higher value is shown in the table.

5.1.3 Success variables

Performance and risk were captured with two accounting-based measures: ROA and OPM. Both measures are commonly used in strategic group research. Moreover, OPM was used in the most recent study which tested for risk differences across strategic groups (McNamara et al., 2003).

For each period, the success variables were calculated as the means over the five-year time period. Risk was calculated as the standard deviation of the performance measures per period (Cool and Schendel, 1987, 1988). An additional measure of risk was tested, calculated as the sum of absolute differences between periods (Houthoofd and Heene, 1997). However, the results proved nearly the same and are thus not presented in this thesis. Risk-adjusted performance was calculated by dividing each performance measure by the respective risk measure.

The same data cleansing procedures have been applied to the success variables as to the strategic variables. The descriptive statistics of the success variables are included in Table 5.3.

5.1.4 Analysis

Prior to clustering, the standardization of all strategic variables was performed. The decision in favor of standardization was made, because it led to more distinctive differences across strategic groups. Also the cluster validation with APN measures indicated that the importance of variables was more equally distributed after standardization.

Following the methodology, a two-stage clustering procedure was performed. Within the first stage, HAC, i.e., Ward's method with Euclidean distance, was used to determine the optimal number of groups with the Mojena criterion. Within the second stage, k-means clustering was performed. The resulting clusters correspond to the strategic groups. The descriptive statistics of the strategic groups are given in Table 5.5.

The validity of the obtained strategic groups was estimated by calculating the APN values as shown in Table 5.4. Most values are close to 20 percent. With six strategic variables, this is close to the optimal value. The importance of financial resources is well below the other variables,

Variable	N	Mean	Median	Min	Max	SD
			Strategic variables			
Geographic scope	133	34.81	34.10	0.00	100.00	26.93
Competitive scope	133	1.47	1.00	1.00	5.00	0.78
Physical resources	133	3.22	2.69	0.57	10.75	2.14
Financial resources	133	2.50	1.98	0.47	10.69	1.78
Organizational resources	133	4.59	4.20	0.66	11.33	2.32
R&D resources	133	17.41	16.55	0.00	70.22	11.17
			Success variables			
OPM	121	-0.73	5.67	-150.90	38.51	28.54
ROA	122	-1.66	4.06	-98.75	24.76	20.40
OPM variance	121	16.89	6.13	0.58	335.83	42.96
ROA variance	122	13.39	7.11	0.38	95.81	18.24
Risk-adj. OPM	121	2.63	1.35	-2.79	28.65	5.24
Risk-adj. ROA	122	1.64	0.77	-2.41	26.64	3.65

Table 5.3: Descriptive variable statistics for the strategic period 2004-2008. Values are shown after data cleansing of all metric variables by removing values which deviated from the mean by more than three standard deviations (Andersen et al., 2007, p. 417).

indicating that this variable is a less meaningful discriminator for the sample at hand. The variable organizational resources shows higher APN values than the other variables, but is not very dominant. Overall, the APN values indicate a reasonable group validity, which is comparable to the results of the other studies in this thesis.

Variable	APN
Geographic scope	0.22
Competitive scope	0.20
Physical resources	0.21
Financial resources	0.07
Organizational resources	0.29
R&D resources	0.18

Table 5.4: Cluster stability given as the average proportion of non-overlap (APN).

The Kruskal-Wallis rank-sum test was used to test for differences across groups. As many previous studies in strategic group research have applied ANOVA as a parametric method, this test was also used. However, it led to the same conclusion as the non-parametric test, such that the results are not shown in this thesis. Also, all variables in this study clearly failed the Shapiro-Wilk normality test and are thus probably not normally distributed. Upon the test for differences across all groups, a post-hoc analysis was conducted with pairwise Mann-Whitney U tests to test for differences between each pair of groups.

Following common practice, a lagged structure was used to account for a certain time lag between strategic decisions and their impact. Two years of data overlap was chosen between strategic and evaluation periods, such that for the strategic period 2004-2008 the success period 2007-2010 was evaluated.

5.2 Results

Table 5.5 shows the results of the significance tests for differences across groups. With the exception of financial resources, all strategic variables are highly significant. Though the variable financial resources is not significant, it should be noted that its p value is not far from being significant at the level of 10 percent and might well become significant with a larger sample size. The results confirm Hypothesis 1, which states that there are structural differences across groups.

The success variables all vary significantly across clusters, thus Hypothesis 2, Hypothesis 3, and Hypothesis 4 are confirmed. There is also a clear relationship between higher performance and lower risk. For instance, group SG5 has the highest performance and the lowest risk values, whereas group SG4 has the lowest performance and the highest risk values. Ranking the groups by performance provides a perfect inverse ranking by the risk measure OPM variance and an almost perfect inverse ranking by ROA variance. Overall, it can be concluded that better performing groups exhibit lower risk, such that Hypothesis 5 can be confirmed.

The results of the posthoc analysis are shown in Table 5.6 and Table 5.7. The tables indicate individual differences between pairs of strategic groups and thus allow for a relative comparison of groups. Overall, the results confirm the findings of the previous tests across all groups. With the exception of financial resources and ROA variance, there is at least one significantly different pair of groups for each variable.

The pair-wise differences indicate which strategic variables make the groups unique. SG1 has significantly less organizational resources than three other groups. SG2 has significantly broader geographic scope and less organizational resources than two other groups. SG3 has significantly more physical resources than three other groups. SG4 has significantly higher R&D expenditures and less organizational resources than any other group. Finally, SG5 has significantly more organizational resources than three other groups and broader competitive scope than any other group.

The pair-wise differences further show that there are no significant success differences between the groups SG1, SG2, and SG3. Also, SG5 is very similar to these groups in terms of success, but shows significantly better performance than SG1 in terms of OPM. SG4 is the most noticeable group as it is significantly less successful by means of all success variables with the exception of ROA variance.

Variable	Group mean					Group diff. (p-value)	Share pair-wise group diff.
	SG1	**SG2**	**SG3**	**SG4**	**SG5**		
Number of firms	40	50	21	11	11		
Strategic variables							
Geographic scope	15.07	54.46	20.44	36.06	43.49	0.0000 ***	0.30
Competitive scope	1.31	1.19	1.50	1.25	3.45	0.0000 ***	0.40
Physical resources	2.25	2.72	7.12	2.31	2.48	0.0000 ***	0.40
Financial resources	2.03	2.31	2.93	4.42	2.28	0.1279	0.00
Organizational resources	2.99	5.70	4.62	1.99	7.95	0.0000 ***	0.70
R&D resources	16.48	16.24	12.65	43.06	9.52	0.0000 ***	0.50
Success variables							
OPM	-2.37	3.64	-4.10	-25.69	18.57	0.0002 ***	0.50
ROA	-0.74	1.49	-1.56	-26.84	8.29	0.0039 **	0.40
OPM variance	12.47	16.26	23.82	28.49	6.75	0.0053 **	0.40
ROA variance	18.08	8.64	9.57	33.08	5.65	0.0108 *	0.00
Risk-adj. OPM	1.54	3.05	1.92	-0.69	9.31	0.0002 ***	0.40
Risk-adj. ROA	0.95	1.82	2.38	-0.73	4.21	0.0025 **	0.30

Table 5.5: Differences across strategic groups.

Significancies: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$ †$p < 0.1$

Variable	Pair-wise group diff. (p-values)			
	SG1/SG2	SG1/SG3	SG1/SG4	SG1/SG5
		Strategic variables		
Geographic scope	0.0000 ***	1.0000	0.2674	0.0266 *
Competitive scope	1.0000	0.8884	1.0000	0.0000 ***
Physical resources	1.0000	0.0000 ***	1.0000	1.0000
Financial resources	1.0000	1.0000	0.3915	1.0000
Organizational resources	0.0000 ***	0.0013 **	0.1981	0.0000 ***
R&D resources	1.0000	0.5875	0.0000 ***	0.0998 †
		Success variables		
OPM	0.3965	1.0000	0.0539 †	0.0360 *
ROA	1.0000	1.0000	0.0505 †	1.0000
OPM variance	1.0000	1.0000	0.0858 †	0.5999
ROA variance	0.1916	0.9478	1.0000	0.2834
Risk-adj. OPM	0.4373	1.0000	0.0296 *	0.1091
Risk-adj. ROA	1.0000	1.0000	0.1017	0.3294

Table 5.6: Pair-wise differences with strategic group SG1.

Significancies: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$ †$p < 0.1$

Variable	Pair-wise group diff. (p-values)					
	SG2/SG3	**SG2/SG4**	**SG2/SG5**	**SG3/SG4**	**SG3/SG5**	**SG4/SG5**
			Strategic variables			
Geographic scope	0.0000 ***	0.2751	1.0000	1.0000	0.2266	1.0000
Competitive scope	0.3165	1.0000	0.0000 ***	1.0000	0.0001 ***	0.0004 ***
Physical resources	0.0000 ***	1.0000	1.0000	0.0000 ***	0.0000 ***	1.0000
Financial resources	1.0000	0.7026	1.0000	1.0000	1.0000	1.0000
Organizational resources	0.2296	0.0000 ***	0.1952	0.0004 ***	0.0198 *	0.0002 ***
R&D resources	0.9757	0.0000 ***	0.1163	0.0001 ***	1.0000	0.0000 ***
			Success variables			
OPM	1.0000	0.0017 **	0.3946	0.0557 †	0.3889	0.0055 **
ROA	1.0000	0.0053 **	0.9342	0.0144 *	1.0000	0.0110 *
OPM variance	1.0000	0.0131 *	0.8742	0.0484 *	0.8714	0.0152 *
ROA variance	1.0000	0.1038	1.0000	0.3394	1.0000	0.1009
Risk-adj. OPM	1.0000	0.0002 ***	0.6583	0.0557 †	0.5345	0.0017 **
Risk-adj. ROA	1.0000	0.0044 **	0.8919	0.0730 †	1.0000	0.0110 *

Table 5.7: Pair-wise differences between strategic groups SG2–SG5.

Significancies: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$ †$p < 0.1$

5.3 Discussion

The main objective of this study was to confirm findings from previous strategic group studies in the setting of the software industry. While a rich body of research exists for a multitude of industries, this is the first detailed stategic group analysis of the software industry. The extensive dataset allowed for the application of the full methodology, providing the most thorough demonstration of the configurational analysis of all studies in this thesis. With the application of the methodology, all hypotheses derived in Section 2.5 could be confirmed. This section discusses the implications for researchers and practitioners as well as the limitations.

As hypothesized, the results suggest that only a limited number of groups dominates the software industry. Though only five distinct groups are found in the dataset, their interpretation remains challenging, as most groups are not characterized by any single property that sets them apart from all other groups. There are only two groups that follow a *unique strategy*, in the sense that one group characteristic is significantly different from all other groups: SG4 has the highest R&D expenditures whereas SG5 has the most organizational resources and the broadest competitive scope of all groups. Consequently, the other three groups, that do not follow a unique strategy, become only distinguishable when looking at multiple characteristics. Such an interpretation is more difficult, as it does not allow for absolute statements as "SG4 has the highest R&D expenditures", but rather relative statements as "SG1 is smaller than any other group except SG4". With five groups and five characteristics there are 50 relationships to be taken into account. This implies that further simplifications need to be made in order to interpret the obtained groupings. For instance, in this study there are no success differences between the groups SG1, SG2, and SG3, such that interpreting differences between them is less relevant than differences towards the groups SG4 and SG5.

Looking at the groups SG5, SG4, and SG1–3, these can be clearly ranked according to their success. The group SG5 is the most successful group, as it has the highest performance and the lowest risk. The group SG4 is the least successful group. The groups SG1–3 are ranked between SG4 and SG5, thus displaying mediocre performance and risk. Interestingly, the groups SG4 and SG5 are small, each with 11 firms or about 4 percent of the total sample. This indicates that most firms conduct strategies that achieve mediocre success, while few firms manage to implement strategies that lead to extraordinary success. There are also few firms that follow unsuccessful strategies.

The reason why so few firms implement the most successful strategy of SG5 must be because it's difficult to implement that strategy. The group SG5 is remarkable as it comprises the largest firms in the sample that also have the broadest competitive and geographical scope. Indeed, this group is comprised of firms such as SAP AG and the Oracle Corporation, which are large firms that provide standard software as well as related services, thus engaging in multiple value chain activities and offering them on a broad range of markets. The success of these large firms is arguably based on network and scale effects discussed in Section 2.1. While this strategy seems to be easy to recognize, it is not easy to copy, as a smaller firm cannot easily grow to the desired size. Even if such a firm could obtain the required funding to make a product for a large market, it would still miss the customer base. And if the firm would rely an acquisitions to obtain the desired customer base, it would still have to overcome the difficult integration of the acquired resources to deal with the customer base. The implication from this is that practitioners should favor strategies that have a broad market and product scope. Meanwhile, researchers and policy makers should be concerned with the question how competition can be increased such that top performing firms with broad market and product scopes can be challenged by their competitors.

The reason why so few firms follow unsuccessful strategies seems to be easy to answer as well. Decision makers will change unsuccessful strategies or find their firms out of business. On the other hand, given the long evaluation period, the results suggest that firms in SG4 choose unsuccessful strategies and follow them for multiple years. Why do these firms choose to be unsuccessful? A possible explanation could be that they find themselves unable to change their strategies. The group characteristics offer an alternative explanation, however. As indicated by the high R&D expenditures of the firms in SG4, these firms could be following a temporary growth strategy, taking their time to bring out a new product to market. With the least organizational resources, these are presumably the youngest firms in the sample and are consequently striving to challenge their established rivals. Decision makers in established firms should thus monitor firms in SG4 as they might challenge their position in the future.

If firms in SG4 want to switch to a more successful strategy, it appears that they could cut their R&D expenditures to copy the strategy of the more desirable group SG1. However, the variables used in this study are too abstract to allow for such a conclusion, as it is impossible to say why these firms have such high R&D expenditures. Therefore, it is impossible to say if cutting these expenditures is a desirable or even feasible solution. The selection of the variables is the main shortcoming of the study. Because of the high data requirements, the data had to be collectable from secondary sources, which do not provide the detailed data required to accurately assess software firms. Here, the finer-grained concept of business models could allow for additional recommendations on how to adjust a given strategy or business model in order to become more successful. Starting out with the results from the strategic group analysis, further research could focus on SG4 and only investigate this group in detail, thus greatly reducing the effort required to collect the data. Due to the restriction of available variables from secondary data sources, an analysis of business model configurations is not possible on the given dataset. However, Study 3 in the following Section 6 addresses the comparative analysis of strategic groups and business model configurations.

5.4 Conclusion

This study applied the full configurational analysis to strategies of prepackaged software firms. The results show five distinct strategic groups. The most successful group contains the largest firms with the broadest scope in the sample. These findings are consistent with the results of Study 1 and could be linked to the exploitation of network effects as well as economies of scale and scope. The least successful group comprises small firms with high R&D expenditures. The other three groups, which contain more than 90 percent of the firms in the sample, show mediocre performance and risk.

This study contributes to configurational and strategic group research with its rigorous application of configurational analysis to software firms. The software industry hasn't been subject of detailed analysis in strategic group research thus far. The course of analysis shows the general applicability of configurational analysis to software firms. The results confirm that effects found across multiple industries are present in the software industry as well. As a major finding specific to software firms it has been found that the most successful strategies in the software industry are characterized by broad market and product scopes, thus exploiting network effects and high economies of scale as well as scope. This study further suggests how strategic group analysis and the analysis of business model configurations can be combined to provide recommendations for actions to practitioners. While strategic groups indicate which groups need to take action, it is the more detailed concept of business models that is likely to provide guidance on what actions are necessary to change the strategic positioning of a firm.

There are certain limitations to this study. Due to high data requirements, only secondary data could be used. As inherently software-specific strategic data could not be obtained from secondary sources that would fulfil the requirements, variables have been chosen which can be interpreted specific to software but are not as indicative as variables that measure software-specifics directly. As another consequence, it wasn't possible to perform an analysis of business model configurations on the data to test the recommended follow-up analysis of business models.

Section 6

Study 3: Strategies and business models in the German software industry

The first two studies, presented in Section 4 and Section 5, used the configurational analysis methodology for separate analyzes of business model configurations and strategic groups in the software industry on different datasets. The analyzes of the two different but similar concepts led to comparable results, i.e., positive impact of larger firm size and broader market scope on firm success. The upside of using different datasets is that the obtained results are more reliable, because they are less likely to be specific to a particular dataset. However, both studies used secondary data, whereas it is desirable to triangulate methods in order to raise confidence in the obtained results, e.g., by combining secondary objective with primary subjective data, as discussed in Section 2.2.3. Another shortcoming of the two datasets was that they didn't allow for a simultaneous analysis of business model configurations and strategic groups, because the collected data was concept-specific in each case. This might well have limited the comparability of the results between the two concepts. Further, in Study 1, business models were conceptualized with a small subset of the variables suggested by the SBMF, because of the difficulties associated with obtaining such data from secondary sources. Finally, for both studies, it cannot be rooted out that the results might be specific to US firms, because the first study analyzed US firms only and also the sample in the second study contained a large proportion of US firms. Overall, the first two studies have certain limitations that can be addressed by using a different dataset, which is the main reason for this third study.

This empirical study analyzes strategic groups and business model configurations in the German software industry. It differs from the first two studies in four major aspects, each of which addresses the previously mentioned shortcomings of the previous studies. The four aspects are: industry focus, data source and assessment mode, number of variables, as well as comparative analyzes of strategic groups and business model configurations. (1) The *industry focus* of this study is on the German software industry, the sample comprises software firms in Germany of all sizes. (2) The data was collected in a large-scale survey, including objective as well as perceived measures on firm strategies, business models, and success. Thus, the *data source and assessment mode* differs in this study, as it includes primary perceived measures in addition to primary objective measures. (3)

This study uses a broad *number of variables* to measure strategies and business models, including variables commonly used in configurational research as well as variables specific to the software industry as defined by the SBMF. This broad selection of variables allows for a fine-grained analysis of configurations, their characteristics, and interdependencies between characteristics. (4) The data covers strategic and business model variables, such that both concepts can be evaluated on the same dataset and *comparative analyzes* between them can be performed.

The results confirm the main findings from previous studies. Foremost, there are significant differences across configurations in strategic and business model characteristics. Success differences are present in performance and risk-adjusted performance, but not risk. There is strong empirical evidence that a broader market scope is associated with higher performance and risk-adjusted performance. For strategic groups there is also evidence of a positive association between size and performance. This study further demonstrates the importance of value chain activities. The activities maintenance and support are major characteristics of strategic groups as well as business model configurations, while additional activities are major separators between different configurations. This finding is important because value activities are the theoretical foundation of both concepts and the results provide empirical support in favor of this common foundation and its significance.

The remainder of this section is organized as follows. Section 6.1 describes the data collection and application of configurational analysis. The results are presented in Section 6.2 for business model configurations and Section 6.3 for strategic groups. The implications and limitations of the findings are discussed in Section 6.4. Finally, Section 6.5 concludes the study.

6.1 Data and method

In order to obtain detailed data on strategies and business models as conceptualized in the SBMF, a large-scale survey was carried out among software firms in Germany in the year 2013. It was a follow-up of a similar survey that was executed in the year 2012. The experiences from the previous year provided lessons learned for the final survey by helping to improve the targeted sample, the questionnaire, and data analyzes. This section describes the data collection and cleansing procedures. It further provides descriptive statistics of the data. The subsequent Section 6.2 and Section 6.3 use the data for analyzes of business model configurations and strategic groups. Concept-specific methods are given in the respective sections.

6.1.1 Variables

The independent variables, representing strategy and business model characteristics, were derived from the SBMF (see Section 2.4). Each component of the SBMF is represented by at least one independent variable in this study, while most components are captured by multiple variables. In total, data on 43 independent variables could be collected. The full list of the variables, their descriptions and measurements is given in Appendix A.3. Though the survey questionnaire was designed to measure more variables, some of them didn't show sufficient variation and were not included in this study.

Six dependent variables were used to capture perceived and objective firm success, as listed in the following enumeration:

1. Perceived performance: This variable is a combination of five different items. Respondents

could rate their perceived performance (in terms of absolute sales, sales growth, absolute profits, profit growth, and profit margin) towards their main competitors on a Likert scale from 1 (much lower) to 7 (much higher).

2. Perceived risk: This variable is a combination of five different items. Respondents could rate their perceived risk (in terms of volatility in absolute sales, sales growth, absolute profits, profit growth, and profit margin) towards their main competitors on a Likert scale from 1 (much lower) to 7 (much higher).

3. Risk-adjusted perceived performance: Calculated as perceived performance divided by perceived risk.

4. OPM: Measured on an interval scale from 1 to 5 with intervals: $< 0\%$, $0 - 10\%$, $10 - 20\%$, $20 - 30\%$, and $> 30\%$.

5. Revenue growth: Calculated as growth in sales revenue from years 2011 to 2012.

6. Personnel growth: Calculated as growth in personnel from years 2011 to 2012.

No objective measure for firm risk was included, because it was expected that respondents would have been reluctant or unable to provide such detailed information. For each calculated measure, if a component required for the calculation was missing because no response was provided on that component, then the calculated measure was treated like a missing value.

6.1.2 Sample selection

The data was collected from April to July 2013 through a large-scale online survey of software firms in Germany. The sample firms were selected in March 2012 and updated in March 2013. For the initial selection in the year 2012, firms fulfilling the following criteria were selected from the Bureau van Dijk Orbis database:

1. The firm was active, such that it could be contacted in order to obtain a response.

2. The firm was located in Germany, such that it could be defined as a German firm.

3. The firm's primary three-digit SIC code was 737 (computer programming, data processing, and other computer related services) or the firm's primary Statistical Classification of Economic Activities in the European Community revision 2 (NACE) code was one of the following: 5821 (publishing of computer games), 5829 (other software publishing), 6201 (computer programming activities), 6202 (computer consultancy activities). These codes were chosen because of their close correspondence to the SIC codes discussed in Section 4.1.

The initial sample was further merged with a list of firms obtained from the Hoppenstedt database. These firms were required to be classified with the SIC code 737. However, the vast majority of the final sample originated from the Bureau van Dijk Orbis database.

The sample was updated in the year 2013 with slightly modified criteria. The sample was extended by firms which fulfilled the following criteria:

1. The firm was active.

2. The firm was located in Germany.

3. The primary NACE code was one of the following: 582 (software publishing), 62 (computer programming, consultancy and related activities), or 63 (information service activities).

4. The firm had at least two employees.

Notably, the selection of codes was modified and a restriction to the number of employees was added. Both changes were the result of the experiences with the initial sample, which was used in the first survey in the year 2012. The experiences showed that NACE codes were more available for most firms than SIC codes. Also, there were many one-man firms in the initial sample that were irrelevant to analyzes performed on the final sample.

The sample was further manually updated during the process of the surveys in years 2012 and 2013. Where no contact data could be obtained from the database, it was manually obtained from the website. Entries without a website or non-reachable website were removed from the sample. Whenever possible, general email addresses were replaced by specific email addresses of employees of the firm. Participating firms had further the option to opt-out from the survey or identify themselves as non-software firms. In both cases they were permanently removed from the sample and added to a blacklist.

The resulting list consisted of more than 33,000 firms. The list was further screened for firms that were actually one entity (e.g., corporations with subsidiaries or holding firms), mainly using the domain of the firm website or email address of the contact persons. After these steps, the list included 32,611 firms. Email addresses were available for 21,583 sample firms. During the course of the survey, several hundred firms were removed since they reported not being active or not being in the scope of the survey.

6.1.3 Questionnaire design

The survey was implemented following the tailored survey design method (Dillman et al., 2009). The complete questionnaire is included in Appendix A.2. The English version and a German translation were made available to the participants through an online link. The original questionnaire was designed in English and then translated to German using an adapted back-translate procedure (Brislin, 1970). For that, two researchers, who were not involved in the development of the questionnaire, were asked to translate the questionnaire. The first researcher translated the questionnaire from English to German. The second researcher used the translated German version and translated it back to English. These translations were reviewed by the two designers of the questionnaire in order to identify good translations and questions which could lead to misunderstandings as indicated by a mismatch between the original and the back-translated English version.

6.1.4 Survey process

The data collection process began by sending out the main survey package to all 21,583 sample firms on 2nd April 2013. The roll-out mail contained information about the survey and instructions on how to participate in the survey. The delivery status of the emails was recorded, and a second batch of emails was sent a few weeks later. In the second round, all non-functioning email addresses in the first round (i.e., emails that bounced from the receiving mail servers) were removed or substituted with new untried email addresses, if available. During the survey, some respondents reported that the respective firms had moved, was not operating independently anymore, or that the contact person changed the employer. Whenever possible, a new contact address was obtained and contacted.

Several approaches were taken to convince the informant of the importance of the survey. Many organizations closely linked to the software industry were asked to endorse the survey. The survey was conducted on behalf of the Software-Cluster research project, which is funded by the German Federal Ministry of Education and Research. In addition, a promise was made to provide firm-specific reports of the responses as a further incentive to respond.

The total number of responses was 427, including 219 complete and 208 partial responses. Whereas the response rate may seem comparably low, the reasons can easily be explained. Firms were mostly contacted through their general email addresses. About ten percent of the emails could not be delivered. Also, 291 respondents chose to opt-out from the survey. Though the representativeness of the data is limited, the sample comprises a wide range of firms, ranging from the largest software firms in Germany (including SAP AG and Software AG) to small firms with just a few employees.

6.1.5 Descriptive statistics

Table 6.2 (p. 89) shows the descriptive statistics for all independent variables after data cleansing. The results indicate the prevailing strategic and business model choices of software firms in terms of individual firm characteristics. While there is a tendency towards the differentiation strategy (mean value 5.48), the cost leadership strategy is chosen by firms as well (4.10). With respect to growth and profit orientation, both objectives seem to be about equally important with a slight tendency towards growth (4.53 and 4.01). There are two value chain activities that stand out, namely development and implementation (0.84 and 0.64). Most software firms carry out these activities and regard them as their main value creating activities, followed by maintenance and support (0.43 and 0.39). For all other activities, less than 20 percent of all firms regard them as their main value creating activities. Vertical integration shows a magnitude of 72 percent with a rather low standard deviation of 12 percent. The mean size of the firms in terms of revenues is 8,324,000 EUR, with a minimum value of 120,000 EUR and a maximum value of 195,000,000 EUR. The revenue source is mainly the end-customer (2.03), with a tendency towards usage-independent pricing (2.72). Recurring and single payments are about equally important (3.77). The variable software stack layer indicates that there is a tendency towards task-specific software, rather than integrating products (3.15). About equally important and not important are the cloud platform (3.60) and open source license (3.09). The products are about equally standardized and individualized (3.87). With respect to firm costs, the major share of costs are product development (3.21), rather than sales (1.50) and infrastructure (1.36). The variable domestic scope indicates that the domestic market is important, though the value of 0.41 cannot be interpreted in absolute terms. The target customers are mainly small and medium enterprises (SME) (5.14) and large customers (6.23), while private customers are rarely targeted (1.42). The variable verticalization cannot be interpreted in absolute terms, but given the value range (0.00 to 6.00) the average software firm can be found in the middle of the scale (2.70).The main distribution channels are agents (3.82), events (4.85), and telesales (4.07). Less important are online (1.62) and retail (1.15) channels. The implementation effort for the offered products is about centralized with a mean value of 3.86. There is a tendency towards on-premise rather than on-demand products (2.70). The mean value of releases per year is 4.93 with a high standard deviation of 10.42. Standardized and non-standardized support contracts are about equally important (4.25). Finally, on average, firms maintain 3.4 parallel releases of their main product.

Table 6.1 shows the descriptive statistics of the success variables after data cleansing. The perceived performance and risk variables show values close to the mid of the value range (3.95 and 3.81). Thus, the mean relative performance is perceived as average. This indicates a realistic perception over the entire sample, since on average there shouldn't be any relative over- or under-performance. Consequently, the risk-adjusted perceived performance value of 1.18 is close to the optimal value of 1.00. The value of 2.94 in the OPM variable indicates that the average OPM is somewhere between 0 to 10 and 10 to 20 percent, thus well positive. The revenue and personnel growth figures are 33 and 24 percent, thus considerably higher than in Study 1. However, the sample in this study contains many SME and the data shows that smaller firms show much higher growth rates than large firms. The difference in growth rates between the two studies may thus be a result of different sample compositions. Notably, the first four success variables show lower response rates than most other variables in the study. In case of the perceived measures, it stands to reason that respondents found it more difficult to estimate their success towards competitors rather than to give concrete figures such as revenues and personnel. In a similar vein, risk has a lower response rate because respondents had to estimate the relative volatility of their performance, thus adding complexity to provide an estimate. As a calculated measure of the previous two perceived measures, risk-adjusted perceived performance could be only calculated where both components, perceived performance and risk, were available. The comparably low response rate for OPM is somewhat surprising, though it seems reasonable that many firms didn't want to provide that intimate figure in a public survey.

#	Variable	N	Mean	Median	Min	Max	SD
1	Perceived performance	68	3.95	4.00	1.00	6.80	1.21
2	Perceived risk	54	3.81	4.00	1.00	6.60	1.22
3	Risk-adj. perceived perf.	53	1.18	1.00	0.23	4.20	0.66
4	OPM	66	2.94	3.00	1.00	5.00	1.11
5	Revenue growth	75	0.33	0.17	-0.85	5.25	0.80
6	Personnel growth	75	0.24	0.12	-0.40	5.00	0.60

Table 6.1: Descriptive statistics of the dependent variables.

#	Variable	N	Mean	Median	Min	Max	SD
1	Differentiation	75	5.48	5.57	2.71	7.00	0.93
2	Cost leadership	75	4.10	4.00	1.50	6.67	1.25
3	Growth orientation	75	4.53	4.50	1.00	7.00	1.35
4	Profit orientation	75	4.01	4.00	1.00	7.00	1.29
5	Research activity	74	0.15	0.00	0.00	1.00	0.36
6	Development activity	74	0.84	1.00	0.00	1.00	0.37
7	Maintenance activity	74	0.43	0.00	0.00	1.00	0.50
8	Production activity	74	0.03	0.00	0.00	1.00	0.16
9	Marketing activity	74	0.19	0.00	0.00	1.00	0.39
10	Implementation activity	74	0.64	1.00	0.00	1.00	0.48
11	Education activity	74	0.11	0.00	0.00	1.00	0.31
12	Support activity	74	0.39	0.00	0.00	1.00	0.49
13	Operations activity	74	0.16	0.00	0.00	1.00	0.37
14	Replacement activity	74	0.04	0.00	0.00	1.00	0.20
15	Vertical integration	69	0.72	0.75	0.33	0.83	0.12
16	Cooperation	75	4.81	5.00	1.00	7.00	1.62
17	Size (in thousands)	74	8,324	1,500	120	195,000	24,330
18	Revenue source	65	2.03	1.00	1.00	7.00	1.78
19	Usage pricing	67	2.72	2.00	1.00	7.00	2.23
20	Recurring payments	70	3.77	4.00	1.00	7.00	2.09
21	Revenue sharing	67	1.16	1.00	1.00	3.00	0.41
22	Software stack layer	71	3.15	2.00	1.00	7.00	2.32
23	Cloud platform	73	3.60	4.00	1.00	7.00	2.18
24	Open source license	67	3.09	3.00	1.00	7.00	1.92
25	Standardized product	71	3.87	4.00	1.00	7.00	2.03
26	Product costs	66	3.21	3.00	1.00	5.00	1.09
27	Sales costs	68	1.50	1.00	1.00	4.00	0.66
28	Infrastructure costs	69	1.36	1.00	1.00	5.00	0.73
29	Domestic scope	66	0.41	0.40	0.20	0.64	0.15
30	SME customers	73	5.14	6.00	1.00	7.00	2.10
31	Private customers	72	1.42	1.00	1.00	7.00	1.21
32	Large customers	73	6.23	7.00	1.00	7.00	1.45
33	Verticalization	74	2.70	2.50	0.00	6.00	1.92
34	Agents channel	73	3.82	5.00	1.00	7.00	2.52
35	Events channel	74	4.85	5.00	1.00	7.00	1.73
36	Telesales channel	74	4.07	5.00	1.00	7.00	2.20
37	Online channel	72	1.62	1.00	1.00	7.00	1.40
38	Retail channel	73	1.15	1.00	1.00	7.00	0.91
39	Implementation effort	71	3.86	4.00	1.00	7.00	2.02
40	On-demand operation	70	2.70	2.00	1.00	7.00	2.01
41	Releases per year	59	4.93	1.44	0.17	52.00	10.42
42	Standardized support	69	4.25	5.00	1.00	7.00	2.05
43	Parallel releases	58	3.40	3.00	1.00	16.00	2.40

Table 6.2: Descriptive statistics of the independent variables.

6.2 Results on business models

All 43 independent variables measure components of the SBMF and could be potentially used for the formation of business model configurations. However, 11 variables were removed after correlation and multicollinearity analysis: production activity, implementation activity, vertical integration, standardized product, SME customers, private customers, events channel, telesales channel, on-demand operation, releases per year, and parallel releases. The remaining independent variables have a maximal absolute correlation of 0.5 between the variables marketing activity and retail channel (the correlation matrix couldn't be included due to the high number of variables). The maximum VIF has a magnitude of 8.3 between the variables open source license and development activity, thus below the upper bound of 10. The condition number has a magnitude of 13.2, thus well below the upper bound of 30.

The given data was further cleansed from firms for which less than 90 percent of the data was available in the independent variables. This step was performed in order to ensure that the clustering procedure had sufficient data to reliably assign firms to clusters. After this step, the final sample included 63 firms.

Due to data restrictions, some steps suggested by the methodology to derive configurations could not be performed. These include tests for a stable period as well as the consideration of a time lag between business model choices and performance impacts. Both would have required data over multiple periods. However, because the survey was very labor-intensive to carry out and also time-consuming for the respondents, it was not possible to collect data for multiple periods beyond the year 2012.

The formation of business model configurations followed the same steps as in Study 1. Two-stage clustering procedure was performed, using non-parametric methods due to the non-normal distribution of the data. As a result of the two-stage clustering procedure, three business model configurations were obtained, labeled as BMC1–3. The configurations contain 32 firms in BMC1, 13 firms in BMC2, and 18 firms in BMC3. Table 6.3 shows the descriptive statistics of the APN measure to estimate the validity of the obtained configurations. The statistics indicate that on average, 14 percent of the firms would be assigned to a different configuration if a variable was removed. This value is higher than what could be expected given the overall number of 32 independent variables. However, the average influence can be considered moderate. The minimal and maximal values indicate large differences. While some variables do not have any impact (e.g., growth and profit orientation), other variables have very high impact above 40 percent, namely: maintenance, support, operations, and replacement activity as well as verticalization.

Mean	**Median**	**Min**	**Max**	**SD**
0.14	0.06	0.00	0.53	0.82

Table 6.3: Descriptive statistics of the APN measure for cluster validation.

Table 6.4 (p. 92) shows the mean values for variables with significant differences across configurations. Non-significant variables are not shown due to place restrictions. Overall, 14 business model and 3 success variables vary significantly across configurations. The pair-wise comparisons indicate that for each business model variable there is at least one pair of configurations with significant differences in that variable. This fact allows for deeper insights into differences across configurations. For success variables, there is only one significant pair-wise comparison for the variable OPM. No variable shows significant differences in all pair-wise comparisons, indicating that no single variable is absolutely unique for each configuration. However, some configurations have unique characteristics in comparison to all other configurations.

Configuration BMC1 is unique in its combination of the variables maintenance activity, support activity, and verticalization. The values in these variables are the highest among all configurations. Configuration BMC2 is unique regarding the variables marketing activity and operations activity. Both variables have the highest values among all configurations. Configuration BMC3 is unique in the variable research activity, which has the highest value among all configurations. The pair-wise comparisons in success variables are not very indicative, as there is only one significant pair-wise difference overall.

The significant differences in business model variables confirm Hypothesis 1. Interestingly, the most significant variables are related to value chain activities and thus the strategic dimension of business models. The significant differences in perceived performance and OPM confirm Hypothesis 2. There are no significant differences in perceived risk, thus Hypothesis 3 is not confirmed. However, risk-adjusted perceived performance varies significantly, thus, Hypothesis 4 is confirmed. As there are no significant differences in perceived risk, Hypothesis 5 cannot be confirmed.

6.3 Results on strategic groups

The formation of strategic groups was performed upon a selection of strategic variables. Out of the available 43 variables collected through the survey, two types of variables were selected as strategic variables. The first type of variables belongs to the strategic dimension of the SBMF (variables 1 to 15 in Table 6.2, p. 89). Variables of the second type are those commonly used in strategic group research, representing scope and resource commitment. These variables include size, domestic scope, and verticalization. Though verticalization is not a common variable, it arguably measures the scope commitment of a firm.

The variable support activity was removed due to multicollinearity, as the variable resulted in VIF values above 21 and thus considerably higher than the maximal value of 10. Among the remaining 17 variables, the maximal absolute correlation in the sample is 0.42 for the variables differentiation and marketing activity. The correlations are shown in Table 6.5 (p. 93). The maximum VIF has a magnitude of 2.42 between the variables profit orientation and maintenance activity, thus well below the upper bound of 10. The condition number has a magnitude of 4.5, thus well below the upper bound of 30.

#	Variable	Conf. mean			Conf. diff.		Pair-wise conf. diff. (p-values)					
		BMC1	BMC2	BMC3	(p value)		BMC1/BMC2		BMC1/BMC3		BMC2/BMC3	
	Business model variables											
1	Differentiation	5.21	5.56	6.01	0.0083	**	0.8928		0.0047	**	0.6193	
2	Research activity	0.00	0.00	0.56	0.0000	***	1.0000		0.0000	***	0.0029	**
3	Maintenance activity	0.69	0.00	0.22	0.0000	***	0.0001	***	0.0055	**	0.2367	
4	Marketing activity	0.03	0.69	0.17	0.0000	***	0.0000	***	0.2934		0.0114	*
5	Support activity	0.66	0.15	0.17	0.0004	***	0.0079	**	0.0031	**	1.0000	
6	Operations activity	0.03	0.46	0.06	0.0003	***	0.0012	**	1.0000		0.0282	*
7	Usage pricing	2.10	2.92	3.83	0.0337	*	0.6619		0.0315	*	0.9290	
8	Revenue sharing	1.28	1.25	1.00	0.0859	†	1.0000		0.0937	†	0.1079	
9	Cloud platform	2.72	5.15	4.00	0.0025	**	0.0029	**	0.1456		0.4636	
10	Open source license	2.48	3.73	3.85	0.0397	*	0.0695	†	0.1648		1.0000	
11	Domestic scope	0.45	0.38	0.36	0.0698	†	0.3924		0.1096		1.0000	
12	Verticalization	3.60	1.62	1.78	0.0007	***	0.0084	**	0.0042	**	1.0000	
13	Events channel	4.66	4.77	5.72	0.0635	†	1.0000		0.0524	†	0.5103	
14	Retail channel	1.00	1.85	1.00	0.0201	*	0.0544	†	1.0000		0.2001	
	Success variables											
1	Perceived performance	3.65	3.65	4.42	0.0694	†	1.0000		0.1079		0.1931	
2	Risk-adj. perceived perf.	0.96	1.02	1.47	0.0857	†	1.0000		0.1017		0.4169	
3	OPM	2.41	3.30	3.18	0.0276	*	0.1512		0.0572	†	1.0000	

Table 6.4: Significant business model and success variables.
Significancies: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; †$p < 0.1$

	Cost leadership	Growth orient.	Profit orient.	Research act.	Development act.	Maintenance act.	Production act.	Marketing act.	Impl. act.	Education act.	Operations act.	Replacement act.	Size	Vertical integr.	Domestic scope	Verticalization
Differentiation	0.32	0.27	0.30	0.35	-0.18	-0.13	0.26	0.42	-0.07	-0.07	-0.16	0.14	0.19	-0.35	-0.12	-0.24
Cost leadership		-0.19	0.41	-0.16	-0.10	-0.01	0.26	0.25	0.10	-0.08	-0.09	0.20	0.08	-0.33	0.17	-0.15
Growth orientation			-0.09	-0.04	-0.11	-0.11	-0.08	0.30	-0.14	0.09	-0.06	0.24	0.17	0.07	0.02	-0.31
Profit orientation				0.07	0.15	-0.15	0.06	0.21	-0.16	0.15	-0.04	0.24	0.12	-0.08	-0.02	-0.14
Research act.					-0.08	-0.20	-0.08	0.02	-0.07	-0.15	-0.19	-0.09	-0.02	-0.10	-0.27	-0.16
Development act.						-0.16	0.07	0.06	-0.17	-0.36	-0.22	-0.33	-0.09	0.15	-0.07	0.16
Maintenance act.							-0.16	-0.32	-0.23	-0.18	-0.24	0.15	-0.18	-0.03	0.29	0.24
Production act.								0.15	-0.04	-0.06	-0.08	-0.04	-0.10	-0.08	0.07	0.04
Marketing act.									-0.24	-0.16	-0.06	-0.10	0.18	-0.18	-0.07	-0.36
Implementation act.										0.15	0.04	-0.16	0.27	0.02	-0.31	0.03
Education act.											-0.04	-0.07	0.05	0.17	0.06	0.05
Operations act.												-0.09	0.02	-0.03	0.00	-0.16
Replacement act.													0.10	-0.14	0.11	-0.13
Size														-0.06	-0.25	0.11
Vertical integration															0.07	0.18
Domestic scope																0.14

Table 6.5: Correlations between the independent strategic variables. Highlighted in gray is the maximal absolute correlation.

Other than the selection of variables, the applied method resembles the steps performed in the analysis of business models (Section 6.2). Firms for which less than 90 percent of the required data was available were removed from the sample, resulting in a final sample size of 72 firms.

As a result of the two-stage clustering procedure, three strategic groups were obtained, labeled as SG1–3. The number of firms in the groups is 38 in SG1, 20 in SG2, and 14 in SG3. Table 6.6 shows the descriptive statistics of the APN measure to estimate the validity of the obtained configurations. The statistics indicate that on average, 14 percent of the firms would be assigned to a different configuration if a variable was removed. This value is somewhat higher but close to what could be expected given the overall number of 17 independent variables. The minimal and maximal values indicate large differences. While some variables do not have any impact (e.g., development activity and size), other variables have a high impact above 30 percent, namely: research, maintenance, implementation, education, and operations activity, as well as domestic scope.

Mean	**Median**	**Min**	**Max**	**SD**
0.14	0.03	0.00	0.37	0.57

Table 6.6: Descriptive statistics of the APN measure for cluster validation.

Table 6.7 shows the mean values for all significant variables across configurations. Overall, 6 strategic, 5 business model and 2 success variables vary significantly across groups. Though the business model variables were not used to form strategic groups, they were tested for differerences across configurations such that comparisons can be drawn with business model configurations in these variables as well. While most independent variables show significant differences in at least one pair-wise comparison, some variables single out particular groups as unique in comparison to all other groups:

- SG1's unique characteristics are the low importance of the maintenance activity and the moderately important support activity.
- SG2's unique characteristics are the low importance of the implementation activity, high importance of the support activity, small firm size, high domestic scope, and focus on SME customers.
- SG3's unique characteristic is its low focus on the support activity.

#	Variable	Group mean			Group diff.		Pair-wise group diff. (p values)					
		SG1	SG2	SG3	(p value)		SG1/SG2		SG1/SG3		SG2/SG3	
	Strategic and business model variables											
1	Maintenance activity	0.03	0.75	1.00	0.0000	***	0.0000	***	0.0000	***	0.1475	
2	Implementation activity	0.82	0.00	1.00	0.0000	***	0.0000	***	0.2718		0.0000	***
3	Support activity	0.34	0.80	0.00	0.0000	***	0.0031	**	0.0384	*	0.0000	***
4	Size	13,164	1,353	6,290	0.0432	*	0.0764	†	1.0000		0.0931	†
5	Cloud platform	4.41	2.45	3.31	0.0046	**	0.0038	**	0.3984		0.8456	
6	Domestic scope	0.34	0.53	0.41	0.0001	***	0.0001	***	0.1793		0.0724	†
7	SME customers	5.06	6.30	4.50	0.0255	*	0.0632	†	1.0000		0.0489	*
8	Private customers	1.17	1.90	1.43	0.0332	*	0.0440	*	1.0000		0.4738	
9	Verticalization	2.09	3.28	3.44	0.0398	*	0.1018		0.1535		1.0000	
10	Online channel	1.49	2.26	1.21	0.0820	†	0.1859		1.0000		0.2183	
11	Implementation effort	4.26	2.79	4.43	0.0285	*	0.0377	*	1.0000		0.1185	
	Success variables											
1	Perceived performance	4.10	3.49	3.99	0.0685	†	0.1002		1.0000		0.1645	
2	Risk-adj. perceived perf.	1.23	0.90	1.46	0.0945	†	0.2199		1.0000		0.1412	

Table 6.7: Significant strategic and success variables.

Significancies: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; †$p < 0.1$

The significant differences in the independent variables confirm Hypothesis 1. Interestingly, the most significant variables are related to value chain activities. The significant differences in perceived performance confirm Hypothesis 2. There are no significant differences in perceived risk, thus Hypothesis 3 is not confirmed. However, risk-adjusted perceived performance varies significantly, thus, Hypothesis 4 is confirmed. As there are no significant differences in perceived risk, Hypothesis 5 cannot be confirmed.

6.4 Discussion

This study builds upon a unique dataset that relaxes some of the limitations present in the first two studies. This study further differs in its simultaneous configurational analysis of strategies and business models, as presented in the previous two sections. This section discusses the implications of the findings for researchers and practitioners as well as the limitations of the study.

Configurational analysis of both concepts, strategies and business models, provided consistent results to the hypotheses suggested in Section 2.5. Hypotheses 1, 2, and 4 could be confirmed. The hypotheses state that configurations differ in structural characteristics, performance, and risk-adjusted performance. Overall, this confirms that differences exist between configurations. Therefore, configurations exist in the German software industry and further exploratory analyzes can potentially provide insights beyond the hypothesized effects. Hypotheses 3 and 5, both related to risk, were not confirmed. Though this could indicate the absence of hypothesized effects or an insufficient meaningfulness of the concepts, another explanation could be the difficulty associated with the measurement of perceived risk. An indicator supporting this explanation is the lower response rate obtained for the perceived risk variable as opposed to the performance variable. Whereas 68 responses were provided for performance, only 54 responses could be obtained for risk. Thus, a higher response rate could well lead to significant differences in risk as well.

The results of this study provide an overview of the strategic groups and business model configurations in the German software industry in the year 2012. Table 6.8 summarizes the significant characteristics of the obtained configurations. The absolute values of these characteristics indicate where the configurations are positioned in the solution space of all theoretically possible strategies and business models. Notably, despite the large number of variables, the results show just three configurations for strategies as well as business models. This confirms the general assumption in configurational research that the overall solution space can often be reduced to few major configurations. Though the absolute values provide an industry overview, it is the relative values that indicate the positioning towards competitors. For instance, 46 percent of firms in BMC2 regard operations as a main value creating activity. This absolute value could be interpreted such that operations is not a very important activity for BMC2. However, looking at the alternative configurations BMC1 and BMC3, it becomes apparent that BMC2 relies on the operations activity much more than the alternative configurations. Therefore, in order to understand the competitive positioning of a configuration, the relative values towards alternative configurations should be evaluated.

Firms in BMC1 and SG2 share a set of common characteristics. Firms in these configurations regard maintenance and support as their main value creating activities, they use cloud platforms less than other firms, and they have a higher focus on domestic markets. Both configurations show lower performance than alternative configurations in all success variables. SG2 is further characterized by low implementation activity and effort. SME and private customers are of higher

#	Variable	Strategic group			Business model conf.		
		SG1	SG2	SG3	BMC1	BMC2	BMC3
	Significant independent variables shared by SGs and BMCs						
1	Maintenance activity	0.03	0.75	1.00	0.69	0.00	0.22
2	Support activity	0.34	0.80	0.00	0.66	0.15	0.17
3	Cloud platform	4.41	2.45	3.31	2.72	5.15	4.00
4	Domestic scope	0.34	0.53	0.41	0.45	0.38	0.36
5	Verticalization	2.09	3.28	3.44	3.60	1.62	1.78
	Significant independent variables specific to SGs						
6	Implementation activity	0.82	0.00	1.00			
7	Size (in thousands)	13,164	1,353	6,290			
8	SME customers	5.06	6.30	4.50			
9	Private customers	1.17	1.90	1.43			
10	Online channel	1.49	2.26	1.21			
11	Implementation effort	4.26	2.79	4.43			
	Significant independent variables specific to BMCs						
12	Differentiation				5.21	5.56	6.01
13	Research activity				0.00	0.00	0.56
14	Marketing activity				0.03	0.69	0.17
15	Operations activity				0.03	0.46	0.06
16	Usage pricing				2.10	2.92	3.83
17	Revenue sharing				1.28	1.25	1.00
18	Open source license				2.48	3.73	3.85
19	Events channel				4.66	4.77	5.72
20	Retail channel				1.00	1.85	1.00
	Significant dependent variables shared by SGs and BMCs						
1	Perceived performance	4.10	3.49	3.99	3.65	3.65	4.42
2	Risk-adj. perceived perf.	1.23	0.90	1.46	0.96	1.02	1.47
	Significant dependent variables specific to BMCs						
3	OPM				2.41	3.30	3.18

Table 6.8: Strategic group and business model configuration characteristics.

importance than in alternative strategic groups. Also, firms in SG2 are considerably smaller in size. BMC1 is characterized by low research and marketing activities as well as the lowest usage of usage pricing. Overall, the strategy and business model followed by firms in SG2 and BMC1 appears to target narrower markets than alternative configurations, as indicated by the domestic scope. The value chain activities indicate that firms in SG2 and BMC1 focus on generating value via classical activities of software firms, such as maintenance and support, while neglecting to differentiate themselves through research and marketing. These firms might also have lower possibilities to differentiate, as their products require less implementation effort and do not permit to generate additional value via implementation activities. The lower performance of SG2 and BMC1 could be a result of narrower markets and lower differentiation, which lead to lower margins generated by these firms.

Whereas there are similarities between BMC1 and SG2, there are no such evident similarities between the other strategic groups and business model configurations. Looking at the strategic groups SG1 and SG3, the success variables do not indicate which group is more successful. The main differences between them are in the maintenance and support activities. While maintenance is of major importance to SG3, support is more important to SG1. However, the importance of support for SG1 is significantly lower than for the least successful group SG2. From these results, it appears that firms focusing on generating value via the maintenance activity may be well more successful, while the support activity does not lead to extraordinary performance. Interestingly, the variable verticalization, which is also an indicator of a narrow market scope, shows that a narrow scope in terms of targeted industries is not necessarily associated with lower performance, as SG2 and SG3 are both highly verticalized but SG3 performs better than SG2. Looking at the business models BMC2 and BMC3, the success variables favor different business models. The main structural differences between BMC2 and BMC3 are in their focus on generating value via the activities research, marketing, and operations. While BMC2 generates value by operating and marketing their products, BMC3 focuses more on research and pays less attention to marketing and operations.

These results confirm the view that strategy and business models are similar, partly overlapping, but different concepts. The question arises if one of the concepts provides more meaningful configurations or predicts performance better than the other. Looking at business models, the resulting configurations have significant differences in more independent as well as dependent variables. This suggests that separating firms based on business models provides more meaningful configurations than configurations formed upon strategies. However, the most significant variables are the strategic variables. The strategic variables further had the biggest impact on the formation of configurations, as indicated by the high APN values associated with strategic variables. In fact, the most significant and influential variables in this study are the value chain activities. This suggests that while the full stack of business model variables leads to more meaningful results, the usage of strategic variables approximates the results well. Moreover, it seems only logical that business models are more accurate, because they are evaluated on a larger set of variables, which should capture additional firm information. On the downside, the detailed data is difficult to obtain and may often not be readily available. Thus, the usage of the less detailed concept of strategic groups can be a reasonable compromise between sufficiently accurate results and lower costs for data collection.

The more significant results provided by business models suggest that in general, more independent variables lead to better results. Though this observation seems trivial at first, it should be noted that the applied method must allow for a large set of variables to be used. For instance, linear regression would require a much larger sample given the number of variables in this study, whereas two-stage clustering is robust against a growing number of variables. Nevertheless, looking at APN as an indicator of cluster validity, it appears that strategic groups show a higher validity than business models. In particular, the configurations are less impacted by individual variables. This could indicate that common strategic variables are more balanced than the components of the SBMF. Other than that, the interpretation of APN values remains difficult as literature does not provide reference values which would allow for a comparison of cluster validity with other studies.

This study adds a different perspective to the two previous studies by using primary data. Despite different data, the results are comparable across studies. This method triangulation ensures higher confidence in the obtained results. However, the results in success variables are less significant for primary data. A possible explanation is that variables are more difficult to measure with primary subjective data, because subjective assessments are less precise than objective data from secondary sources such as financial databases. On the other hand, only primary data allows for measurements of detailed firm characteristics. To get the best of both worlds, it could be worthwhile to combine primary subjective assessments of independent variables with secondary objective measurements of dependent success variables.

6.5 Conclusion

The results of this study confirm the main findings from the two previous studies. Foremost, there are significant persistent differences across strategic groups as well as business model configurations in their characteristics and performance. There is strong empirical evidence that a broader market scope is associated with higher performance. For strategic groups, there is also evidence of a positive association between size and performance. The effects are present after adjustment for risk, though no significant differences could be detected for risk itself.

This study further demonstrates the importance of value chain activities. The activities maintenance and support are major characteristics of strategic groups as well as business model configurations, while additional activities are major separators between different configurations. This finding is important because value activities are the theoretical foundation of both concepts and the results provide empirical support in favor of this common foundation and its significance.

The main contribution of this study is the simultaneous analysis of strategic groups and business model configurations. Results obtained for both concepts are consistent with the previous findings. One particular configuration could be even identified with strategies and business models, indicating that this configuration is particularly distinctive. Business models further proved to provide more significant results in configurational analysis than strategies, but impose much higher data requirements due to the large number of variables and their detailed nature. The findings of this study suggest that strategic groups lead to similar findings and should be preferred over business model configurations when detailed analysis of business model characteristics is not required.

The dataset in this study adds different perspectives to the results obtained from the data in the first two studies, but it has shortcomings on its own. It is limited to software firms in Germany and provides data for a single period, the year 2012. A survivorship bias is present, as only existing and active firms could be contacted to participate in the survey. Further, the response rate to the survey is too low to be representative, such that the main value of this study is in its additional perspective to the results of the previous studies.

Section 7

Discussion

The previous sections addressed particular research questions, presented findings, and discussed their specific implications and limitations. This section presents an integrated discussion of these contents. The answers to the research questions are summarized in Section 7.1. Section 7.2 discusses the implications for researchers and practitioners. The limitations are discussed in Section 7.3. A final conclusion is presented in Section 8.

7.1 Research objectives

Section 1.1 formulated four research questions, each contributing to the overall research goal from a different perspective. The following sections summarize the findings for each research question.

7.1.1 Research Question 1: How can configurational analysis of strategies and business models contribute to our understanding of the software industry?

The literature review on strategic groups and business model configurations in Section 2 extracted the potential of configurational analysis from extant research. This indicates what can be expected from configurational analysis as a theoretical and methodological framework in general. Sections 4–6 then presented configurational analyzes of software firms, which provide results on the software industry in particular.

In general, configurational analysis identifies groups of firms that share key characteristics of a concept. The analysis can be applied to different concepts to obtain the respective configurations, whose characteristics are determined by their member firms. This thesis analyzed (1) firm strategies to obtain strategic groups and (2) business models to obtain business model configurations. Strategic groups (business model configurations) can be viewed as abstract, prototypical strategies (business models). The abstraction from individual firms leads to a great reduction of the solution space. The quantitative studies in this thesis reduced the solution space to just 3–5 competitive approaches. Arguably, this number closer resembles the mental models of practitioners than the huge potential solution space spanned by frameworks such as the SBMF. Also researchers aim to reduce the overall number of available options. For instance, the review of business model literature in Section 2.4 showed that deductive studies use classifications with 2–4 abstract business models. Moreover, because configurational analysis maximizes differences across configurations and minimizes differences within them, it indicates commonly used competitive approaches as well as those that are very distinctive. I.e., a competitive approach that is sufficiently distinctive will be identified as an own entity even if only few firms apply it. As such, when applied to strategies (business models) of software firms, configurational analysis provides a taxonomy of abstract, commonly applied and distinctive strategies (business models) in the software industry along with key characteristics that delineate them. Further, configurational analysis provides an assignment of firms to the different competitive approaches.

The quantitative studies demonstrated the applicability of configurational analysis as outlined in Section 2.3.2 to software firm strategies and business models. Whenever individual methods couldn't be applied, it wasn't due to characteristics of the concepts at hand, but data limitations of a particular dataset. The obtained results proved to be comparable to those known from extant research on other industries, e.g., the existence of configurations in general. As for the specific results on the software industry, these could be explained with the economic properties of the software products and markets. For example, Study 1 confirmed that differences across configurations exist in the software industry, which has been established for a multitude of other industries in previous studies. Moreover, the results of the study may explain the differences between configurations from a software-specific point of view, because characteristics are interconnected with other key characteristics. E.g., size is positively affected by characteristics such as focus on infrastructure software and larger market scope. As such, configurational analysis describes the available competitive approaches and helps to explain how they create competitive advantage.

By comparing the success of firms applying particular competitive approaches, it can be established which strategies (business models) are more successful than others. The effects can be explained by examining the theoretical impact of strategy (business model) characteristics on firm success. For example, the empirical results indicate that strategies (business models) enabling larger network effects, which are particularly important for software firms, are more successful than competitive approaches which do not facilitate network effects. Ultimately, such relationships between determining characteristics and their impact on firm success can be used to establish how firms should behave and predict their future behavior. Configurational analysis, therefore, helps to describe how successful different competitive approaches are, explain why the effects occur, and possibly predict future success as firms adjust their strategies and business models.

7.1.2 Research Question 2: Which activities are performed and combined by software firms to create competitive advantage?

The literature review in Section 2 identified the value chain as a common theoretical foundation of strategies and business models. Section 3 then specifically addressed the research question which value creating activities are combined by software firms to create competitive advantage. The applicability of the suggested value chain was then confirmed in the third quantitative study presented in Section 6.

The results suggest that software firms create value by carrying out the following ten activities: (1) research, (2) development, (3) maintenance, (4) production, (5) marketing, (6) implementation, (7) education, (8) support, (9) operations, and (10) replacement. The empirical evaluation indicated six activities that differentiate software firm strategies and business models, thus creating competitive advantage towards competitors: (1) research, (2) maintenance, (3) marketing, (4) implementation, (5) support, and (6) operations.

7.1.3 Research Question 3: Which strategies and business models prevail in the software industry?

By applying configurational analysis to strategies and business models of software firms, the quantitative studies in this thesis provided an overview of extant strategies and business models in the software industry. Study 1 analyzed business models, Study 2 strategies, and Study 3 analyzed both concepts. Though different datasets and conceptualizations were used, some common themes across studies emerged that help to answer this research question.

As a common theme across the three studies, characteristics associated with the scope of product and market operations have been particularly significant. In each study, scope is not a single variable but a combination of different key characteristics. In Study 1, all five key characteristics are associated with broader scope, i.e., high sales volumes, infrastructure software, high international sales, many targeted industries, and indirect sales channels. In Study 2, broader scope is determined by high sales volumes, high international sales, and broad product scope. In Study 3, broader scope is associated with high sales volumes, high international sales, and many targeted industries. In conclusion, prevailing software firm strategies and business models can be classified as those facilitating broader or narrower scopes. The classification appears simple, however, multiple characteristics represent scope and their combined effects determine firm success.

7.1.4 Research Question 4: How successful are the various strategies and business models along different success dimensions?

The quantiative studies in this thesis evaluated differences across strategic groups and business model configurations in terms of performance, risk, and risk-adjusted performance. The results provide insights into success effects of prevalent strategies and business models along different success dimensions and their interrelationships.

There is strong empirical evidence that strategies and business models with broader scope perform better than alternative approaches with narrower scope, also after adjustment for risk. The findings on risk itself are less clear, though Study 2 found strong evidence that broader scope is associated with less risk. Similar, though statistically much less significant, results were obtained in Study 1. Nonetheless, the results indicate that higher performance goes along with lower risk, such that firms do not face tradeoffs between these two desirable objectives.

7.2 Implications

This thesis established that a software firm's choice of a strategy and business model determines the firm's success. This finding seems intuitive or even trivial, as differences between firms are apparent from their balance sheets and stock market valuations. Nonetheless, the results are rich in implications for researchers and practitioners, as they (1) indicate specific success factors, (2) describe competitive approaches, and (3) develop configurational analysis. The implications from these points are discussed in turn.

7.2.1 Success factors in the software industry

A major finding of this thesis is the strong empirical evidence suggesting that a software firm's scope has a positive impact on success in terms of performance, risk-adjusted performance, and possibly risk. I.e., strategies and business models with broader scope perform better, show higher risk-adjusted performance, and are sometimes less risky than competitive approaches with narrower scope. Simply stated, it appears that broader scope is equivalent to more success in the software industry. The positive impact of scope in the software industry can well be explained with the economic properties of software markets. With increasing market and product scope, firms increase their sales and reduce their average item costs through high economies of scale and scope, thus effectively increasing their margin. Moreover, network effects increase with higher sales, thus making the product more valuable to its users, allowing the firm to charge higher prices and increase its margin.

There is a second lesson in the findings on success effects: It appears that there are no conflicts across different success dimensions, such that software firms do not face goal conflicts and do not have to make tradeoffs between high performance and low risk when choosing a competitive approach. This finding resembles Bowman's paradox and poses an interesting research question: Why do firms choose different competitive approaches and not the one that leads to the highest performance, lowest risk, and highest risk-adjusted performance? Indeed, the results in this thesis do not allow for an answer to this question, though a starting point could be the difficulty associated with the transition from one competitive approach to another, as it takes certain resources and capabilities that may be difficult to acquire. More difficulties may stem from interdependencies between firm characteristics, making it necessary to change many characteristics at once, which may be too difficult to achieve.

From the importance of scope for firm success it should not be followed that other firm decisions unrelated to scope are not important. Study 3 in particular identified multiple variables which are not associated with scope but have a significant impact on firm performance, such as a firm's emphasis on differentation towards its competitors. However, though other decisions than a firm's scope will affect its success, the results indicate that scope is sufficient to determine which competitive approaches will be more successful than others.

The positive impact of broader scope on firm success seems to be apparent to decision makers in the software industry as can be seen from the high mergers and acquisitions (M&A) activity of software firms (Schief, 2013, p. 14-15). The effects of scope should be further of interest to analysts and investors with an interest in software firms, who should favor firms following competitive approaches with broader scope. Policy makers, however, should be wary of the implications, as software firms striving for the broadest possible scope essentially develop monopolistic tendencies.

It should be of particular interest to researchers and practitioners alike to understand how decision makers can expand the scope of their firms to be more successful. Though causalities between individual characteristics are not the focus of this thesis, the obtained results provide certain insights nevertheless, offering avenues for further research. For instance, Study 1 suggests that broader scope is associated with infrastructure software and higher usage of indirect sales channels. These variables can be directly influenced by decision makers and result in higher sales volumes, thus effectively increasing the scope of a firm.

7.2.2 Competitive approaches in the software industry

A common scheme that emerged in all studies is that strategies and business models can be classified according to their scope broadness. Though a classification along just one dimension appears crude, the significant success effects confirm the importance of scope as a dimension to describe strategies and business models in the software industry. I.e., the empirical results have shown that the most successful competitive approaches can be associated with broader scope and the least successful ones with narrower scope. The relative definition is key here. In order to describe strategies and business models, this thesis uses relative statements. Rather than establishing a threshold to differentiate between broad and narrow scopes, it is stated whether a competitive approach has a broader or narrower scope than others. Whereas such relative definitions appear less elegant than absolute ones, it is plausible that in order to be more successful than its competitors, a firm has to consider its positioning towards its competitors rather than absolute thresholds.

Scope is a valuable descriptor of competitive approaches not just because of its significant impact on firm success, but also because it can be associated with a multitude of firm decisions. For instance, firms with the largest sales volumes often have more international sales and target many industries, which allows them to target more markets to place their sales. Study 1 even indicated that firms facilitate broader scope with indirect sales, which provides them with better scalability as they expand their scope. Moreover, scope can provide insights into the actual product offering, as infrastructure products, such as operating systems, can reach higher sales volumes, because they are less specialized and have thus more potential uses than application products. Scope is therefore a combination of firm decisions. The logic of interdependencies between decisions allows for an interpretation of how firms compete.

The results suggest that practitioners should seek to increase their scope to be more successful than their competititors. The descriptive statistics of competitive approaches in this thesis provide practitioners with insights into which characteristics need to be adjusted in order to position themselves for broader scope. The example in the previous paragraph indicated how multiple characteristics and their interdependencies should be taken into account to achieve that.

For researchers, the findings suggest that scope should be regarded as a major dimension in conceptualizations of software firm strategies and business models. The literature review of extant business model studies in Section 2.4.2 indicates that the need to include scope has not been recognized in configurational business model research yet. This is different from strategic group research, where market scope is regarded as a necessary dimension to be taken into account. Also, evaluations of software firm performance should control for scope in general, as it is a highly significant impact factor.

The comparisons of strategies and business models in this thesis show that both concepts lead to comparable results. However, differences exist. The strategy concept is well researched and easier to capture than the detailed business model concept. As a result, data is easier to obtain for analyzes of firm strategies. Moreover, strategic variables show stronger effects than business model variables. This could be interpreted as an indicator that the strategy concept is more founded in general as a determinant of firm success. On the downside, the strategy concept provides less insights into detailed firm decisions, thus limiting the interpretability and applicability of the results. In conclusion, researchers who are not interested in detailed firm decisions but are seeking to control for the main determinants of software firm performance may restrict themselves to strategic decisions.

Though consistent results were obtained for measures from different data sources and assessment modes, different datasets proved to be more applicable in certain settings. Primary data proved to be the only way to collect the detailed data required by the SBMF. Without the option to obtain detailed data required for industry-specific business model measurement, researchers will have to rely on primary data further on, as has been done in this thesis and previous configurational business model studies. The downside is that such datasets are often not available to other research teams and thus do not allow for reproducibility. Also, without the availability of secondary datasources, the possibilities for method triangulation are limited, thus reducing the overall confidence in the empirical results. The difficulties with collecting primary data further often result in a small sample size, thus limiting the potential for statistical analyzes.

7.2.3 Configurational analysis of strategies and business models

This thesis fills two gaps in extant configurational research. (1) Previous studies on strategic groups provided conclusive empirical evidence for multiple industries with the exception of the software industry, which has not been analyzed in detail thus far. The results in this thesis confirm previous findings, such as the existence of strategic groups and performance differences across them. Further findings on industry-specific effects, such as the importance of scope and its interdependencies with other firm characteristics, additionally confirm the meaningfulness of the strategic group concept in the context of the software industry and the applicability of strategic group analysis to software firms. (2) Previous studies on business models only partly applied methods from configurational analysis to derive configurations in the software industry. This thesis demonstrated the applicability of the full methodology to business models of software firms. Detailed conceptualizations, e.g., Study 3 used 43 variables to measure business models, allow for deep analyzes of configurations in the software industry.

By applying configurational analysis to software firms, prevailing strategies and business models could be identified, along with their characteristics and success indicators. The significant results and their interpretability confirm the value of the configurational approach to improve our understanding of how firms compete. As opposed to conventional methods like regression analysis, configurations identify sets of characteristics that prevail together and thus help to identify interdependencies.

The course of analysis suggests that configurational analysis may well be applicable to concepts other than strategies and business models in the context of the software industry. As discussed in Section 1, organizational configurations may cover a wide set of concepts. Thus, researchers and practitioners in the field of IS can apply configurational analysis to organizational configurations in general. The methodology outlined in this thesis and the presented results suggest that this could be a fruitful ground for research and lead to meaningful mental models for practitioners. Moreover, the methodology can be applied to other industries beyond software firms.

7.3 Limitations

There are several limitations to this thesis. These should be taken into account by researchers and practitioners in their interpretation of the results. They also provide avenues for further research.

The major limitation stems from the limited sample sizes in the empirical studies, ranging from 63 to 133 firms. These samples sizes are caused by the high data requirements of configurational analysis as well as difficulties to obtain sufficiently detailed software-specific data. Nonetheless, the numbers are comparable to those of most previous strategic group studies. Though most configurational business model studies used larger samples, they did not adhere to the same high data requirements of configurational analysis. The sample sizes in this thesis probably limited the ability to detect further effects and larger samples would certainly improve our confidence in the empirical results. In a similar vain, in order to keep up the overall sample sizes, each study had to make tradeoffs in the scope of data to be collected. For example, Study 1 limited the number of variables and periods, Study 2 limited the number of variables, and Study 3 limited the number of periods for which data was collected.

Because of the different datasets in the three empirical studies, they also vary in the applied method, as some steps couldn't be carried out. For instance, only Study 2 allowed to test for stable strategic periods and carry out the full methodology. On the other hand, the other two studies build upon more detailed data and thus allow for deeper interpretations. These differences in data limit the comparability of the studies' results. Preferably, a single large dataset should be used that covers objective data from secondary sources as well as perceptual data from primary sources. However, due to the mentioned difficulties with obtaining such data for software firms, different datasets had to be used. As a result, all studies adhere to the most critical steps of configurational analysis, such as two-stage clustering, but only Study 2 executes all methodological steps. As a consequence, business models could not be evaluated with the full methodology in this thesis.

Related to the low sample size in Study 3, which is the only study to use perceptual success measures, the study does not detect pair-wise differences in perceived performance and risk. This might indicate that the operationalizations are insufficient or that respondents found it difficult to provide data on these measures. Further research should collect more perceptual data on performance and risk to evaluate effects on perceived success.

Another major limitation is that practitioners are given limited guidance as to how to design their actions in order to modify their strategies and business models. Though they are provided with relevant firm characteristics that need to be modified, the question remains how the modifications can be carried out. Moreover, this thesis uses entire firms as a unit of analysis, but firms with multiple products may apply different strategies and business models to each product. In order to ensure that the obtained results are in fact comparable, Study 1 and Study 2 used very focused samples and Study 3 asked the recipients to provide data for their main products.

Although this thesis clearly focused on configurational analysis, it is desirable to combine this kind of analysis with conventional analytical methods, such as multiple regression. This could lead to further insights into which characteristics have an effect in configurations only or apply to all firms in general.

Section 8

Conclusion

Knowing the determinants of firm success can be essential to practitioners as being successful is at the very heart of economic activity. It is also crucial to researchers, as differences in performance across firms offer explanations to how firms behave, revealing reasonings behind their actions. This thesis suggests that success factors in the software industry are rooted in its industry-specific economic properties, such as high economies of scale, high economies of scope, and network effects. Software firms following strategies and business models that facilitate broader scope benefit from these forces by becoming more successful than their competitors.

Configurational analysis was chosen to evaluate strategies and business models as it offers a rich methodology to identify prevalent competitive approaches and their characteristics. Three empirical studies were performed on distinct datasets, evaluating strategic groups and business model configurations in the software industry. The first notable finding is that identifiable competitive approaches exist, thus allowing to draw further conclusions on their characteristics and effects. The numbers of abstract strategies and business models range from three to five, each associated with a unique combination of firm characteristics. A major finding is that competitive approaches can be described as having a broader or narrower scope. The characteristics associated with scope provide insights into firm decisions to be made to increase scope, such as increasing sales, expand internationally, and target more industries. The results suggest that broader scope is desirable as it ultimately leads to higher performance. The effects on higher performance are still present after adjustment for risk. The effects on risk itself remain less clear, though some evidence can be found that broader scope also results in lower risk. This finding is interesting, because it suggests that firms do not face a tradeoff between different objectives, but can possibly increase their performance while reducing their risk at the same time.

This thesis contributes to multiple research fields. Strategic group research benefits from the detailed empirical results on the software industry which wasn't subject of similar studies thus far. Business model research benefits from the transfer of theories and methods from strategies to business models. IS and software business research benefit from the descriptions of software firm value creating activities, strategies, business models, statistics on their characteristics, as well as theoretical and methodological considerations from configurational research that can be applied for further analyzes of the software industry.

Practitioners will find the empirical findings on the effects of scope particularly useful. Decision makers in software firms can use the characteristics associated with scope as decision variables in order to implement more successful strategies and business models. Investors and analysts with an interest in the software industry can use the characteristics and their effects to forecast future success. Policy makers should be wary of the implications, as the economic properties of software seem to favor strategies and business models with monopolistic tendencies.

The main limitation to this thesis is rooted in the small sample sizes of the three empirical studies and the third study in particular. Though the numbers are easily explained by the high data requirements of the applied methods, the confidence in the empirical results would increase with larger sample sizes. Moreover, it can be expected that further effects could be detected with additional data. This should be particularly helpful for perceptual measures that proved to be difficult to obtain and showed less significant results than objective measures. Future research should thus focus on collecting additional software-specific and longitudinal data for further configurational analyzes of the software industry.

Appendix A

Appendix

A.1 Overview of strategic group studies

Reference	Theory	Industry	Performance scope			Risk scope	Assessment	Data source
			Financial	**Non-financial**	**Other**			
Cheng and Chang (2009)	Cognitive models	Semiconductor	ROA; profit margin	–	–	–	Objective	Secondary (annual reports)
Claver-Cortés et al. (2004)	Mobility barriers	Construction	ROA; ROS	–	–	–	Objective	Secondary (Ardan)
Cool and Schendel (1987)	Mobility barriers	Pharma-ceuticals	AROS; risk-adjusted AROS	Market share; risk-adjusted market share; seg-ment share; risk-adjusted segment share	–	All measures risk-adjusted with their variance	Objective	Secondary (IMS)
Cool and Schendel (1988)	Mobility barriers	Pharma-ceuticals	AROS; risk-adjusted AROS	–	–	AROS vari-ance	Objective	Secondary (IMS)
Deephouse (1999)	(Cognitive models)	Banking	Relative ROA	–	–	–	Objective	Secondary (Call Re-ports)

Table A.1 – continued on next page

Table A.1 – continued from previous page

Reference	**Theory**	**Industry**	**Performance scope**			**Risk scope**	**Assessment**	**Data source**
			Financial	**Non-financial**	**Other**			
DeSarbo et al. (2009)	Cognitive models	Banking	Tobin's q; market-to-book ratio; dividend yield; PER	–	–	–	Objective	Secondary (Compustat)
Desarbo and Grewal (2008)	(Mobility barriers)	Banking	Tobin's q; market-to-book ratio; dividend yield; PER	–	–	–	Objective	Secondary (Compustat)
Ebbes et al. (2010)	(Mobility barriers)	Banking	Tobin's q; market-to-book ratio; dividend yield; PER	–	–	–	Objective	Secondary (Compustat)
Ferguson et al. (2000)	Mobility barriers	Insurance	–	Specific (insurance reputation): loss ratio; expense ratio	–	–	Objective	Secondary (OneSource)

Table A.1 – continued on next page

Table A.1 – continued from previous page

Reference	**Theory**	**Industry**	**Performance scope**			**Risk scope**	**Assessment**	**Data source**
			Financial	**Non-financial**	**Other**			
Fiegenbaum et al. (1990)	–	Pharmaceuticals	–	–	–	–	Objective	Secondary (Compustat)
Fiegenbaum and Thomas (1995)	–	Insurance	–	–	–	–	Objective	Secondary
Grover and Saeed (2004)	Mobility barriers	Internet-based businesses	ROA	–	–	–	Objective	Secondary (public statements)
Houthoofd (2009)	Mobility barriers	Construction	Subjective financial-based	–	–	–	Subjective	Primary (survey)
Houthoofd and Heene (1997)	Mobility barriers	Brewing	Risk-adj. ROA	–	–	ROA risk-adjusted with ROA variance	Objective	Secondary and primary (interview)
Ketchen et al. (1993)	(Mobility barriers)	Hospitals	ROA; ROE; Specific: net patient revenue; profit-per-discharge	Occupancy	–	–	Objective	Secondary

Table A.1 – continued on next page

Table A.1 – continued from previous page

Reference	Theory	Industry	Performance scope			Risk scope	Assessment	Data source
			Financial	**Non-financial**	**Other**			
Lawless et al. (1989)	(Mobility barriers)	Manufacturing	Profit margin; ROA; ROE	–	–	–	Objective	Secondary
Leask and Parker (2006)	Mobility barriers	Pharmaceuticals	–	–	–	–	Objective	Secondary
Leask and Parker (2007)	Mobility barriers	Pharmaceuticals	–	Market share; weighted market share	–	–	Objective	Secondary (IMS)
Mas-Ruiz and Ruiz-Moreno (2011)	(Mobility barriers)	Banking	–	Lerner index	–	–	Objective	Secondary
McNamara et al. (2003)	Mobility barriers	Banking	ROA; ROE; OPM	–	–	(ROA variance); (ROE variance); (OPM variance)	Objective	Secondary

Table A.1 – continued on next page

Table A.1 – continued from previous page

Reference	**Theory**	**Industry**	**Performance scope**			**Risk scope**	**Assessment**	**Data source**
			Financial	**Non-financial**	**Other**			
Mehra (1996)	Mobility barriers	Banking	ROA; relative PER; net profit/employees	–	–	–	Objective	Secondary
Murthi et al. (2013)	Mobility barriers	Airlines	OP	–	–	–	Objective	Secondary
Nair and Kotha (2001)	Mobility barriers	Steel	ROS; ROA	–	–	–	Objective	Secondary
Nath and Gruca (1997)	Cognitive models, mobility barriers	Hospital	ROA	Occupancy; market share of admissions	–	–	Objective	Primary (survey), secondary
Neill and Rose (2006)	Cognitive models	Wholesale distribution	Subjective financial-based	Subjective efficiency- and customer-based	–	–	Subjective	Primary (survey)
O'Regan et al. (2011)	Mobility barriers	Plastics	ROA	–	–	–	Objective	Secondary (AMI)

Table A.1 – continued on next page

Table A.1 – continued from previous page

Reference	Theory	Industry	Performance scope			Risk scope	Assessment	Data source
			Financial	**Non-financial**	**Other**			
Osborne et al. (2001)	Cognitive models	Pharmaceuticals	–	–	–	–	Objective	Secondary (annual reports)
Pandian et al. (2006)	(Mobility barriers)	Pharmaceuticals	ROA; ROIC; CRET; CBRET	–	–	CBRET	Objective	Secondary (CRSP, Compustat)
Porac et al. (1989)	Cognitive models	Knitwear manufacturers	–	–	–	–	–	–
Porter (1979)	(Mobility barriers)	Consumer goods	ROE	–	–	–	Objective	Secondary
Reger and Huff (1993)	Cognitive models	Banking	ROA	–	–	–	Objective	Primary (interviews)
Schimmer and Brauer (2012)	Cognitive models	Insurance	ROA; dividend/assets	–	–	–	Objective	Secondary
Short et al. (2007)	Mobility barriers	12 (including pre-packaged software and pharmaceuticals)	ROA; Tobin's q	–	Altman's Z	–	Objective	Secondary

Table A.1 – continued on next page

Table A.1 – continued from previous page

Reference	Theory	Industry	Performance scope			Risk scope	Assessment	Data source
			Financial	**Non-financial**	**Other**			
Veliyath and Ferris (1997)	Mobility barriers	Airline; computer equipment; pharmaceuticals	ROE	–	–	Beta; market returns variance	Objective	Secondary
Wiggins and Ruefli (1995)	Mobility barriers	Pharmaceuticals; paints and allied products; primary metals; office equipment and computing machinery; general merchandize stores	ROA; Tobin's q	–	–	–	Objective	Secondary (Compustat)
Zúñiga-Vicente et al. (2004)	Mobility barriers	Banking	ROA; ROE	Market share growth	Equity/liabilities; assets growth	–	Objective	Secondary

Table A.1: Structured overview of strategic group studies.

A.2 Software Industry Survey 2013 questionnaire

Software Industry Survey 2013

This questionnaire is part of the annual *Software Industry Survey* conducted by the Technische Universität Darmstadt. The results will be available in a public report and will not reveal individual companies. By responding to this survey, you help us to analyse the current state and perspectives of the software industry in Germany. The results of the survey and a firm-specific analysis will be delivered to responding companies after the results are published. **All your responses are confidential!**

Instructions:

- **Fill the questionnaire according to the current state of your firm.**
- **Estimate an answer if information is required that is difficult to obtain.**
- **Also partial responses are useful.**

See https://www.softwareindustrysurvey.de for more information.

There are 23 questions in this survey

Firm information

1 []Which one of the following options best describes your firm business? *

Please choose **only one** of the following:

- ○ Standard software product firm (including software-as-a-service providers)
- ○ Individual / custom software development firm
- ○ Implementation and implementation consulting firm (e.g. installation and modification of software products)
- ○ Other software-related firm (e.g. embedded software, business consulting, reseller)
- ○ Firm that is not in any way related to the software industry

2 []What year was your firm founded?

Only numbers may be entered in this field.

Please write your answer here:

Main product performance and risk

3 []

For all remaining questions, please think of the *product which is responsible for the main share* of your firm's revenues. If your firm offers many different products, please think of one particular important or representative product.

The term *product* is further used to mean both:

- **software products or**
- **software-related services**

4 []What share of your firm's revenues can be attributed to your main product?

Please choose **only one** of the following:

- ○ <50%
- ○ 50-70%
- ○ >70%

5 []How do you estimate the following figures for your product compared to the products of your main competitors?

Please choose the appropriate response for each item:

	Much lower			Same			Much higher
Absolute sales (in EUR)	○	○	○	○	○	○	○
Sales growth (in percent)	○	○	○	○	○	○	○
Absolute profits (in EUR)	○	○	○	○	○	○	○
Profit growth (in percent)	○	○	○	○	○	○	○
Profit margin (in percent)	○	○	○	○	○	○	○

6 []How volatile are the following figures for your product (compared to the products of your main competitors)?

Please choose the appropriate response for each item:

	Much lower			Same			Much higher
Absolute sales (in EUR)	○	○	○	○	○	○	○
Sales growth (in percent)	○	○	○	○	○	○	○
Absolute profits (in EUR)	○	○	○	○	○	○	○
Profit growth (in percent)	○	○	○	○	○	○	○
Profit margin (in percent)	○	○	○	○	○	○	○

Business strategy of the main product

7 []Indicate the importance of the following strategies to your firm.

Please choose the appropriate response for each item:

	Not important at all			Moderately important			Very important
Build a strong sales and marketing force	○	○	○	○	○	○	○
Be unique in our industry	○	○	○	○	○	○	○
Invest in research	○	○	○	○	○	○	○
Build a strong brand image	○	○	○	○	○	○	○
Provide higher quality than competitors	○	○	○	○	○	○	○
Fulfil more requirements than competitors	○	○	○	○	○	○	○
Invest in user interface design and usability	○	○	○	○	○	○	○
Minimize costs in general	○	○	○	○	○	○	○
Minimize development costs	○	○	○	○	○	○	○
Minimize advertising expenses	○	○	○	○	○	○	○
Provide at lower costs than competitors	○	○	○	○	○	○	○
Provide at lower prices than competitors	○	○	○	○	○	○	○
Emphasize economies of scale and scope	○	○	○	○	○	○	○

Business model (A)

8 []How well do the following statements describe your product strategy?

Please choose the appropriate response for each item:

	Strongly disagree			Do not agree or disagree			Strongly agree
High growth is the most important objective of our firm	○	○	○	○	○	○	○
At the moment, we see no need for strong growth	○	○	○	○	○	○	○
Our firm must grow even if we need to sacrifice profits	○	○	○	○	○	○	○
High profits are the most important objective of our firm	○	○	○	○	○	○	○
At the moment, we see no need to generate high profits	○	○	○	○	○	○	○
Our firm must generate high profits even if we need to sacrifice growth	○	○	○	○	○	○	○
Cooperations with other firms are an important part of our strategy	○	○	○	○	○	○	○
We engage in considerable exchange, sharing, or co-development with other firms	○	○	○	○	○	○	○
We maintain many cooperations with other firms	○	○	○	○	○	○	○

Business model (B)

9 []What percentage of your operating costs is spent on the following items?

Please choose the appropriate response for each item:

	0-20%	20-40%	40-60%	60-80%	80-100%
Salaries and other personnel costs	○	○	○	○	○
Products or services purchased from third parties (e.g. subcontracting, licenses, hardware)	○	○	○	○	○
What percentage of your revenues do you pay as a fee to your distributor?	○	○	○	○	○

10 []Which are the 3 core activities that add value to your product?

Please select at most 3 answers

Please choose **all** that apply:

- ☐ Research (algorithms, technologies, patents)
- ☐ Development (requirements, coding, testing, documentation)
- ☐ Maintenance (incremental modifications, bug fixes)
- ☐ Production (printing, physical packaging)
- ☐ Marketing (launch, price, place, ...)
- ☐ Implementation (installation, configuration, modification)
- ☐ Education (training, certification)
- ☐ Support (phone hotline, email support)
- ☐ Operations (hosting, monitoring, backup, upgrade)
- ☐ Replacement of the product at it's life end

Business model (C1)

11 []Which of the extremes describes your product best?

Please choose the appropriate response for each item:

	1	2	3	4	5	6	7	
End customers generate most of our revenues (e.g. license sales, maintenance, consulting)	○	○	○	○	○	○	○	Third parties generate most of our revenues (e.g. advertising)
Charge based on **usage-dependent** metrics (e.g. memory usage)	○	○	○	○	○	○	○	Charge based on **usage-independent** metrics (e.g. named user)
Charge through a single payment (e.g. up-front)	○	○	○	○	○	○	○	Charge through recurring payments (e.g. subscription)
The product is designed to help performing specific tasks (e.g. ERP, accounting, office, media, games)	○	○	○	○	○	○	○	The product is designed to integrate IT systems (e.g. operating system, middleware, engineering, security, servers)

7 of 13

Business model (C2)

12 []Which of the extremes describes your product best?

Please choose the appropriate response for each item:

	1	2	3	4	5	6	7	
Our product is tailor-made for each customer	○	○	○	○	○	○	○	Our product is the same for all customers
The implementation effort (installation and modification) is low	○	○	○	○	○	○	○	The implementation effort is high
The product is operated on-premise (i.e. installation and execution on local user systems)	○	○	○	○	○	○	○	The product is operated on-demand (i.e. installation and execution on a central hosting platform supporting access via internet, e.g. Saas)
Each customer has a completely individualized support contract (e.g. customer-specific service level agreements)	○	○	○	○	○	○	○	All customers have the same support contract

8 of 13

Business model (D)

13 []How important are the following licences for your product?

Please choose the appropriate response for each item:

	Not important at all			Moderately important			Very important
Open source copyleft licenses (e.g. GPL)	○	○	○	○	○	○	○
Open source permissive licenses (e.g. BSD)	○	○	○	○	○	○	○
Proprietary licenses	○	○	○	○	○	○	○
All rights are transferred to our customers	○	○	○	○	○	○	○

14 []How important are the following platforms for your product?

Please choose the appropriate response for each item:

	Not important at all			Moderately important			Very important
Mobile (e.g. iOS, Android)	○	○	○	○	○	○	○
Desktop or laptop computers (e.g. Windows, Linux)	○	○	○	○	○	○	○
Servers (e.g. mainframes)	○	○	○	○	○	○	○
Cloud computing (e.g. Force.com)	○	○	○	○	○	○	○
Social media (e.g. Facebook)	○	○	○	○	○	○	○
Game consoles (e.g. XBox)	○	○	○	○	○	○	○
Embedded systems	○	○	○	○	○	○	○

Business model (E)

15 []How important are the following geographic regions for your product sales?

Please choose the appropriate response for each item:

	Not important at all			Moderately important			Very important
Germany	○	○	○	○	○	○	○
Rest of Europe and Russia	○	○	○	○	○	○	○
Middle East and Africa	○	○	○	○	○	○	○
Asia and South Pacific	○	○	○	○	○	○	○
North, South, and Central America	○	○	○	○	○	○	○

16 []What percentage of your operating costs is spent on the following items?

Please choose the appropriate response for each item:

	0-20%	20-40%	40-60%	60-80%	80-100%
Product functionality, design, quality and security (e.g. developer salaries and bonuses)	○	○	○	○	○
Product marketing and sales (e.g. salaries and bonuses for sales staff)	○	○	○	○	○
Technical infrastructure and equipment (e.g. hardware, costs for providing the product as software as a service)	○	○	○	○	○

Business model (F)

17 []How important are the following channels for the distribution of your product?

Please choose the appropriate response for each item:

	Not important at all			Moderately important			Very important
Sales agents	○	○	○	○	○	○	○
Events	○	○	○	○	○	○	○
Telesales	○	○	○	○	○	○	○
Online shop	○	○	○	○	○	○	○
Retail stores	○	○	○	○	○	○	○

18 []How important are the following industries for your product sales?

Please choose the appropriate response for each item:

	Not important at all			Moderately important			Very important
Manufacturing, automobiles	○	○	○	○	○	○	○
Information- and communication	○	○	○	○	○	○	○
Finance and insurance	○	○	○	○	○	○	○
Wholesale and retail	○	○	○	○	○	○	○
Pharma and chemicals	○	○	○	○	○	○	○
Logistics, transport and storage	○	○	○	○	○	○	○
Health and education	○	○	○	○	○	○	○

Business model (G)

19 []Indicate the importance of the following customer segments to your firm.

Please choose the appropriate response for each item:

	Not important at all			Moderately important			Very important
Small and medium organizations (250 employees or less)	○	○	○	○	○	○	○
Large organizations (more than 250 employees)	○	○	○	○	○	○	○
Consumers (private individuals)	○	○	○	○	○	○	○

20 []Answer the following questions about the releases of your product (a release is defined as a major version, not bug fix or patch).

Please write your answer(s) here:

How often do you provide new product releases (in weeks, e.g. 4 is equivalent to monthly releases, 52 to yearly releases)?	
How many releases of the product do you maintain in the marketplace? (e.g. Microsoft maintains 4 Windows releases, being Windows XP, Vista, 7, and 8)	

Firm key figures

21 []

Please answer the remaining questions for your entire firm (not only the main product).

22 []Estimate the following figures for your firm. (You can use common abbreviations such as "m" for "million" and "bn" for "billion".)

Please write your answer(s) here:

Number of personnel at the end of **2012**	
Revenue (EUR) in **2012**	
Number of personnel at the end of **2011**	
Revenues (EUR) in **2011**	

23 []Estimate your *operating profit margin* in 2012 (profit before extraordinary items, interests, and taxes, divided by sales).

Please choose **only one** of the following:

- ○ <0%
- ○ 0-10%
- ○ 10-20%
- ○ 20-30%
- ○ >30%

13 of 13

A.3 Independent variables used in Study 3

#	Variable	SBMF component	Description
1	Differentiation	Unique selling proposition	Combination of multiple Likert items on a scale from 1 (low) to 7 (high)
2	Cost leadership	Unique selling proposition	Combination of multiple Likert items on a scale from 1 (low) to 7 (high)
3	Growth orientation	Investment horizon	Combination of multiple Likert items on a scale from 1 (low) to 7 (high)
4	Profit orientation	Investment horizon	Combination of multiple Likert items on a scale from 1 (low) to 7 (high)
5	Research activity	Value chain activity	Boolean, indicates whether it's a main activity of the firm
6	Development activity	Value chain activity	Boolean, indicates whether it's a main activity of the firm
7	Maintenance activity	Value chain activity	Boolean, indicates whether it's a main activity of the firm
8	Production activity	Value chain activity	Boolean, indicates whether it's a main activity of the firm
9	Marketing activity	Value chain activity	Boolean, indicates whether it's a main activity of the firm
10	Implementation activity	Value chain activity	Boolean, indicates whether it's a main activity of the firm
11	Education activity	Value chain activity	Boolean, indicates whether it's a main activity of the firm
12	Support activity	Value chain activity	Boolean, indicates whether it's a main activity of the firm
13	Operations activity	Value chain activity	Boolean, indicates whether it's a main activity of the firm
14	Replacement activity	Value chain activity	Boolean, indicates whether it's a main activity of the firm
15	Vertical integration	Degree of vertical integration	Calculated as personnel costs divided by the sum of costs for personnel and purchases
16	Cooperation	Number of cooperation partners	Likert scale from 1 (few) to 7 (many)
17	Size	Sales volume	Natural logarithm of sales

Table A.2 – continued on next page

Table A.2 – continued from previous page

#	Variable	SBMF component	Description
18	Revenue source	Revenue source	Bipolar scale from 1 (end customer) to 7 (third party)
19	Usage pricing	Pricing assessment base	Bipolar scale from 1 (usage-independent) to 7 (usage dependent)
20	Recurring payments	Structure of payment flows	Bipolar scale from 1 (single payment) to 7 (recurring payments)
21	Revenue sharing	Revenue distribution model	Five intervals indicating the percentage of shared revenues (0-20%, 20-40%, ...)
22	Software stack layer	Software stack layer	Bipolar scale from 1 (software for specific tasks) to 7 (software to integrate IT systems)
23	Cloud platform	Platform	Likert scale indicating platform importance from 1 (not important at all) to 7 (very important)
24	Open source license	License model	Likert scale indicating license importance from 1 (not important at all) to 7 (very important)
25	Standardized product	Degree of standardization	Bipolar scale from 1 (tailor-made) to 7 (same for all customers)
26	Product costs	Key cost driver	Five intervals indicating the share at overall costs (0-20%, 20-40%, ...)
27	Sales costs	Key cost driver	Five intervals indicating the share at overall costs (0-20%, 20-40%, ...)
28	Infrastructure costs	Key cost driver	Five intervals indicating the share at overall costs (0-20%, 20-40%, ...)
29	Domestic scope	Localization	Calculated as the importance of the German market divided by the sum of importances of all markets (each item measured on a Likert scale from 1 to 7)
30	SME customers	Target customer	Likert scale indicating the importance from 1 (not important at all) to 7 (very important)
31	Private customers	Target customer / target users	Likert scale indicating the importance from 1 (not important at all) to 7 (very important)
32	Large customers	Target customer	Likert scale indicating the importance from 1 (not important at all) to 7 (very important)

Table A.2 – continued on next page

Table A.2 – continued from previous page

#	Variable	SBMF component	Description
33	Verticalization	Target industry	Calculated measure, high (low) values correspond to high (low) verticalization
34	Agents channel	Channel	Likert scale indicating the importance from 1 (not important at all) to 7 (very important)
35	Events channel	Channel	Likert scale indicating the importance from 1 (not important at all) to 7 (very important)
36	Telesales channel	Channel	Likert scale indicating the importance from 1 (not important at all) to 7 (very important)
37	Online channel	Channel	Likert scale indicating the importance from 1 (not important at all) to 7 (very important)
38	Retail channel	Channel	Likert scale indicating the importance from 1 (not important at all) to 7 (very important)
39	Implementation effort	Implementation effort	Likert scale from 1 (low) to 7 (high)
40	On-demand operation	Operating model	Bipolar scale from 1 (on-premise) to 7 (on-demand)
41	Releases per year	Maintenance model	Number of releases per year
42	Standardized support	Support model	Bipolar scale from 1 (individualized support contracts) to 7 (same support contracts for all customers)
43	Parallel releases	Replacement strategy	Number of parallel releases

Table A.2: Independent variables used in Study 3 to measure business models and strategic groups.

Bibliography

Al-Debei, M. M. and Avison, D. (2010). Developing a unified framework of the business model concept. *European Journal of Information Systems*, 19(3):359–376.

Amit, R. and Wernerfelt, B. (1990). Why do firms reduce business risk? *Academy of Management Journal*, 33(3):520–533.

Anderberg, M. (1973). *Cluster analysis for applications.* Academic Press, New York.

Andersen, T. J., Denrell, J., and Bettis, R. A. (2007). Strategic responsiveness and Bowman's risk-return paradox. *Strategic Management Journal*, 28(4):407–429.

Anderson, J. C., Jain, D. C., and Chintagunta, P. K. (1993). Customer value assessment in business markets: a state-of-practice study. *Journal of Business-to-Business Marketing*, 1(1):3–29.

Aubert, B. A., Beaurivage, G., Croteau, A.-M., and Rivard, S. (2009). Firm strategic profile and IT outsourcing. In *Information Systems Outsourcing*, pages 217–240. Springer-Verlag, Berlin.

Bain, J. (1956). *Barriers to New Competition.* Harvard University Press, Cambridge.

Barnes, S. J. (2002). The mobile commerce value chain: analysis and future developments. *International Journal of Information Management*, 22(2):91–108.

Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1):99–120.

Barney, J. B. and Hoskisson, R. E. (1990). Strategic groups: untested assertions and research proposals. *Managerial and Decision Economics*, 11(3):187–198.

Bazerman, M. H. (1984). The relevance of Kahneman and Tversky's concept of framing to organizational behavior. *Journal of Management*, 10(4):333–343.

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression diagnostics: identifying influential data and sources of collinearity.* John Wiley & Sons, Hoboken, NJ.

Bettis, R. A. (2012). The search for asterisks: compromised statistical tests and flawed theories. *Strategic Management Journal*, 33(1):108–113.

Bonaccorsi, A., Giannangeli, S., and Rossi, C. (2006). Entry strategies under competing standards: hybrid business models in the open source software industry. *Management Science*, 52(7):1085–1098.

Bowman, E. H. (1980). A risk/return paradox for strategic management. *Sloan Management Review*, 21(3):17–31.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 3(1):185–216.

Bromiley, P. (1991). Testing a causal model of corporate risk taking and performance. *Academy of Management Journal*, 34(1):37–59.

Bromiley, P., Miller, K. D., and Rau, D. (2001). Risk in strategic management research. In Hitt, M. A., Freeman, R. E., and Harrison, J. S., editors, *The Blackwell Handbook of Strategic Management*, pages 259–288. Blackwell, Malden.

Burkhart, T., Werth, D., Krumeich, J., and Loos, P. (2011). Analyzing the business model concept - a comprehensive classification. In *International Conference on Information Systems*, pages 1–19, Shanghai.

Buxmann, P., Diefenbach, H., and Hess, T. (2012). *The software industry: economic principles, strategies, perspectives.* Springer, Berlin.

Capon, N., Farley, J. U., and Hoenig, S. (1990). Determinants of financial performance: a meta-analysis. *Management Science*, 36(10):1143–1159.

Chakrabarti, S. (2003). *Mining the web: discovering knowledge from hypertext data.* Morgan Kaufmann Publishers, San Francisco.

Cheng, S.-L. and Chang, H.-C. (2009). Performance implications of cognitive complexity: an empirical study of cognitive strategic groups in semiconductor industry. *Journal of Business Research*, 62(12):1311–1320.

Choi, S.-Y., Stahl, D. O., and Whinston, A. B. (1997). *The economics of electronic commerce.* Macmillan Technical Publishing, New York.

Claver-Cortés, E., Molina-Azorín, J. F., and Quer-Ramón, D. (2004). An analysis of intergroup and intragroup differences. *Management Research*, 2(1):81–89.

Collis, D. (1994). How valuable are organizational capabilities? *Strategic Management Journal*, 15(1):143–152.

Cool, K. O. and Schendel, D. (1987). Strategic group formation and performance: the case of the U.S. pharmaceutical industry, 1963-1982. *Management Science*, 33(9):1102–1124.

Cool, K. O. and Schendel, D. (1988). Performance differences among strategic group members. *Strategic Management Journal*, 9(3):207–223.

Courtland, R. (2015). Gordon Moore: the man whose name means progress. http://spectrum.ieee.org/computing/hardware/gordon-moore-the-man-whose-name-means-progress, accessed 25.04.2015.

Covin, J. (1991). Entrepreneurial versus conservative firms: a comparison of strategies and performance. *Journal of Management Studies*, 28(5):339–462.

Crnkovic, I., Larsson, S., and Chaudron, M. (2005). Component-based development process and component lifecycle. *Journal of Computing and Information Technology*, 13(4):321–327.

Crook, T. R., Ketchen, D. J., Combs, J. G., and Todd, S. Y. (2008). Strategic resources and performance: a meta-analysis. *Strategic Management Journal*, 29(2):1151–1154.

Cusumano, M. A. (2004). *The business of software*. Free Press, New York.

Dalkey, N. and Helmer, O. (1963). An experimental application of the delphi method to the use of experts. *Management Science*, 9(3):458–467.

Dechow, P. M. and Skinner, D. S. (2000). Earnings management: reconciling the view of accounting, academics, practitioners, and regulators. *Accounting Horizons*, 14(2):235–250.

Deephouse, D. L. (1999). To be different, or to be the same? It's a question (and theory) of strategic balance. *Strategic Management Journal*, 20(2):147–166.

Desarbo, W. S. and Grewal, R. (2008). Hybrid strategic groups. *Strategic Management Journal*, 29:293–317.

DeSarbo, W. S., Grewal, R., and Wang, R. (2009). Dynamic strategic groups: deriving spatial evolutionary paths. *Strategic Management Journal*, 30(13):1420–1439.

Dhama, H. (1995). Quantitative Models of Cohesion and Coupling in Software. *Journal of Systems and Software*, 1212(94):65–74.

Dillman, D. A., Smyth, J. D., and Christian, L. M. (2009). *Internet, Mail, and Mixed-Mode Surveys - The Tailored Design Method*. Wiley, Hoboken, 3rd edition.

Ebbes, P., Grewal, R., and DeSarbo, W. S. (2010). Modeling strategic group dynamics: A hidden Markov approach. *Quantitative Marketing and Economics*, 8(2):241–274.

Engelhardt, L. (2004). Entrepreneurial models and the software sector. *Competition and Change*, 8(4):391–410.

Engelhardt, S. (2008). The economic properties of software. Research paper, Friedrich Schiller University Jena, Jena.

Ferguson, T. D., Deephouse, D. L., and Ferguson, W. L. (2000). Do strategic groups differ in reputation? *Strategic Management Journal*, 21(12):1195–1214.

Fiegenbaum, A. (1990). Prospect theory and the risk-return association: an empirical examination in 85 industries. *Journal of Economic Behavior and Organization*, 14(2):187–203.

Fiegenbaum, A., Sudharshan, D., and Thomas, H. (1990). Strategic time periods and strategic groups research: concepts and an empirical example. *Journal of Management Studies*, 27(2):133–148.

Fiegenbaum, A. and Thomas, H. (1986). Dynamic and risk measurement perspectives on Bowman's risk-return paradox for strategic management: an empirical study. *Strategic Management Journal*, 7(5):395–407.

Fiegenbaum, A. and Thomas, H. (1995). Strategic groups as reference groups: theory, modeling and empirical examination of industry and competitive strategy. *Strategic Management Journal*, 16(6):461–476.

Finch, H. (2005). Comparison of distance measures in cluster analysis with dichotomous data. *Journal of Data Science*, 3:85–100.

Fisher, F. M. and McGowan, J. J. (1983). On the misuse of accounting rates of return to infer monopoly profits. *The American Economic Review*, 73(1):82–97.

Fiss, P. C., Marx, A., and Cambré, B. (2013). Configurational theory and methods in organizational research: introduction. In Fiss, P. C., Marx, A., and Cambré, B., editors, *Configurational theory and methods in organizational research*, page 319. Emerald Group Publishing, Bingley.

Forward, A. and Lethbridge, T. C. (2008). A taxonomy of software types to facilitate search and evidence-based software engineering. In *Proceedings of the 2008 Conference of the Center for Advanced Studies on Collaborative Research: Meeting of Minds*, CASCON '08, pages 14:179–14:191, New York. ACM.

Fox, J. and Monette, G. (1992). Generalized collinearity diagnostics. *Journal of American Statistical Association*, 87(417):178–183.

Gooding, R. Z., Goel, S., and Wiseman, R. M. (1996). Fixed versus variable reference points in the risk-return relationship. *Journal of Economic Behavior and Organization*, 29(2):331–350.

Grover, V. and Saeed, K. A. (2004). Strategic orientation and performance of internet-based businesses. *Information Systems Journal*, 14(1):23–42.

Haesen, R., Snoeck, M., Lemahieu, W., and Poelmans, S. (2008). On the definition of service granularity and its architectural impact. *Lecture Notes in Computer Science*, 5074/2008:375–389.

Hartman, S. J. and Nelson, B. H. (1996). Group decision making in the negative domain. *Group & Organization Management*, 21(2):146–162.

Hatten, K. J. and Hatten, M. L. (1987). Strategic groups, asymmetrical mobility barriers and contestability. *Strategic Management Journal*, 8(4):329–342.

Hess, T., Loos, P., Buxmann, P., Erek, K., Frank, U., Gallmann, J., Gersch, M., Zarnekow, R., and Zencke, P. (2012). ICT providers: a relevant topic for business and information systems engineering? *Business & Information Systems Engineering*, 4(6):367–373.

Hoskisson, R. E., Hitt, M. A., Johnson, R. A., and Moesel, D. D. (1993). Construct validity of an objective (entropy) categorical measure of diversification strategy. *Strategic Management Journal*, 14(3):215–235.

Houthoofd, N. (2009). Business definition and performance implications: the case of the Belgian construction sector. *Construction Management and Economics*, 27(7):639–652.

Houthoofd, N. and Heene, A. (1997). Strategic groups as subsets of strategic scope groups in the Belgian brewing. *Strategic Management Journal*, 18(8):653–666.

Hunt, M. S. (1972). *Competition in the major home appliance industry, 1960-1970.* PhD thesis, Harvard University, Cambridge.

Hutzschenreuter, T. and Gröne, F. (2009). Changing Vertical Integration Strategies under Pressure from Foreign Competition: The Case of US and German Multinationals. *Journal of Management Studies*, 46(2):269–307.

Izci, E. and Schiereck, D. (2010). Programmierte Wertgenerierung durch M&A in der Business Software-Industrie? *Mergers and Acquisitions Review*, (2):69–74.

Jansen, S., Brinkkemper, S., and Finkelstein, A. (2007). Providing transperancy in the business of software: a modeling technique for software supply networks. In *IFIP International Federation for Information Processing*, volume 243, pages 677–686. Springer.

Kahneman, D. and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, 47(2):263–292.

Katz, M. L. and Shapiro, C. (1994). Systems competition and network effects. *The Journal of Economic Perspectives*, 8(2):93–115.

Keats, B. W. and Hitt, M. a. (1988). A causal model of linkages among environmental dimensions, macro organizational characteristics, and performance. *Academy of Management Journal*, 31(3):570–598.

Ketchen, D. J., Combs, J. G., Russell, C. J., Shook, C. L., Dean, M. A., Runge, J., Lohrke, F. T., Naumann, S. E., Haptonstahl, D. E., Baker, R., Beckstein, B. a., Handler, C., Honig, H., and Lamoureux, S. (1997). Organizational configurations and performance: a meta-analysis. *Academy of Management Journal*, 40(1):223–240.

Ketchen, D. J., Hult, G. T. M., and Slater, S. F. (2007). Toward greater understanding of market orientation and the resource-based view. *Strategic Management Journal*, 28(9):961–964.

Ketchen, D. J. and Shook, C. L. (1996). The application of cluster analysis in strategic management research: an analysis and critique. *Strategic Management Journal*, 17(6):441–458.

Ketchen, D. J., Thomas, J. B., and Snow, C. C. (1993). Organizational configurations and performance: a comparison of theoretical approaches. *Academy of Management Journal*, 36(6):1278–1313.

Klosterberg, M. (2010). Die Bewertung von Softwareunternehmen. In Drukarczyk, J. and Ernst, D., editors, *Branchenorientierte Unternehmensbewertung*, pages 255–273. Franz Vahlen, München, 3 edition.

Kontio, J., Jokinen, J.-P., Mäkelä, M. M., and Leino, V. (2005). Current practices and research opportunities in software business models. In *Proceedings of the Seventh International Workshop on Economics-driven Software Engineering Research*, EDSER '05, pages 1–4, New York. ACM.

Krumeich, J., Burkhart, T., Werth, D., and Loos, P. (2012). Towards a component-based description of business models: A state-of-the-art analysis. In *AMCIS 2012 Proceedings*, pages 1–13.

Lambert, S. C. (2015). The importance of classification to business model research. *Journal of Business Models*, 3(1):49–61.

Lambert, S. C. and Davidson, R. A. (2012). Applications of the business model in studies of enterprise success, innovation and classification: An analysis of empirical research from 1996 to 2010. *European Management Journal*.

Landeta, J. (2006). Current validity of the delphi method in social sciences. *Technological Forecasting and Social Change*, 73(5):467–482.

Lavie, D. (2007). Alliance Portfolios and Firm Performance: A Study of Value Creation and Appropriation in the U.S. Software Industry. *Strategic Management Journal*, 28(12):1187–1212.

Lawless, M. W., Bergh, D. D., and Wilsted, W. D. (1989). Performance variations among strategic group members: an examination of individual firm capability. *Journal of Management*, 15(4):649–661.

Leask, G. and Parker, D. (2006). Strategic group theory: review, examination and application in the UK pharmaceutical industry. *Journal of Management Development*, 25(4):386–408.

Leask, G. and Parker, D. (2007). Strategic groups, competitive groups and performance within the U.K. pharmaceutical industry: Improving our understanding of the competitive process. *Strategic Management Journal*, 28(7):723–745.

Leask, G. and Parnell, J. A. (2005). Integrating strategic groups and the resource based perspective: understanding the competitive process. *European Management Journal*, 23(4):458–470.

Léger, P.-M. and Quach, L. (2009). Post-merger performance in the software industry: the impact of characteristics of the software product portfolio. *Technovation*, 29(10):704–713.

Léger, P.-M., Yang, S., and Leger, P.-M. (2005). Network effects and the creation of shareholders' wealth in the context of software firm mergers and acquisitions. In *ECIS 2005 Proceedings*, Regensburg.

Lehmann, S. and Buxmann, P. (2009). Pricing strategies of software vendors. *Business & Information Systems Engineering*, 1(6):452–462.

Li, F. and Whalley, J. (2002). Deconstruction of the telecommunications industry: from value chains to value networks. *Telecommunications Policy*, 26(9-10):451–472.

Lovelock, J.-D., Graham, K. H., O'Connell, A., Hahn, W. L., Atwal, R., and Graham, C. (2015). Forecast alert: IT spending, worldwide, 1Q15 update. Report, https://www.gartner.com/doc/3024720, accessed 05.05.2015.

Malone, T. W., Weill, P., Lai, R. K., D'Urso, V. T., Herman, G., Apel, T. G., and Woerner, S. L. (2006). Do some business models perform better than others? Research paper, MIT Sloan School of Management, Cambridge.

March, J. G. and Shapira, Z. (1987). Managerial perspectives on risk and risk taking. *Management Science*, 33(11):1404–1418.

Mas-Ruiz, F. J. and Ruiz-Moreno, F. (2011). Rivalry within strategic groups and consequences for performance: the firm-size effects. *Strategic Management Journal*, 32(12):1286–1308.

Mason, E. (1939). Price and production policies of large- scale enterprises. *American Economic Review*, 29(1):61–74.

McGee, J. and Thomas, H. (1986). Strategic groups: theory, research and taxonomy. *Strategic Management Journal*, 7(2):141–160.

McNamara, G., Deephouse, D. L., and Luce, R. A. (2003). Competitive positioning within and across a strategic group structure: the performance of core, secondary, and solitary firms. *Strategic Management Journal*, 24(2):161–181.

Mehra, A. (1996). Resource and market based determinants of performance in the U.S. banking industry. *Strategic Management Journal*, 17(4):307–322.

Messerschmitt, D. G. and Szyperski, C. (2003). *Software ecosystem*. The MIT Press, Cambridge.

Miles, R. E. and Snow, C. C. (1978). *Organizational strategy, structure, and process*. McGraw-Hill, New York.

Milligan, G. W. and Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2):159–179.

Mojena, R. (1977). Hierarchical grouping methods and stopping rules: an evaluation. *Computer Journal*, 20(4):359–363.

Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics Magazine*, 8(38):114–117.

Morris, M. H., Shirokova, G., and Shatalov, A. (2013). The Business Model and Firm Performance: The Case of Russian Food Service Ventures. *Journal of Small Business Management*, 51(1):46–65.

Murthi, B. P. S., Rasheed, A. A., and Goll, I. (2013). An empirical analysis of strategic groups in the airline indusry using latent class regressions. *Managerial and Decision Economics*, 34(2):59–73.

Nair, A. and Kotha, S. (2001). Does group membership matter? Evidence from the Japanese steel industry. *Strategic Management Journal*, 22(3):221–235.

Nath, D. and Gruca, T. S. (1997). Convergence across alternative methods for forming strategic groups. *Strategic Management Journal*, 18(9):745–760.

Neill, S. and Rose, G. M. (2006). The effect of strategic complexity on marketing strategy and organizational performance. *Journal of Business Research*, 59(1):1–10.

Nickel, M. N. and Rodriguez, M. C. (2002). A review of research on the negative accounting relationship between risk and return: Bowman's paradox. *The International Journal of Management Science*, 30(1):1–18.

Okoli, C. and Pawlowski, S. D. (2004). The delphi method as a research tool: an example, design considerations and applications. *Information & Management*, 42(1):15–29.

O'Regan, N., Kluth, C., and Parnell, J. A. (2011). The demise of strategic groups as an influence on firm performance: lessons from the UK plastics industry. *Strategic Change*, 20:111–126.

Osborne, J. D., Stubbart, C. I., and Ramaprasad, A. (2001). Strategic groups and competitive enactment: a study of dynamic relationships between mental models and performance. *Strategic Management Journal*, 454(5):435–454.

Pandian, J. R., Thomas, H., Furrer, O., and Bogner, W. C. (2006). Performance differences across strategic groups: an examination of financial market-based performance measures. *Strategic Change*, 15(7-8):373–383.

Parolini, C. (1999). *The value net.* John Wiley & Sons, Chichester.

Penrose, E. T. (1959). *The theory of the growth of the firm.* Oxford University Press, Oxford.

Peteraf, M. and Shanley, M. (1997). Getting to know you: a theory of strategic group identity. *Strategic Management Journal*, 18(S1):165–186.

Pil, F. K. and Holweg, M. (2006). Evolving from value chain to value grid. *MIT Sloan Management Review*, 47(4):72–80.

Porac, J. F., Thomas, H., and Baden-Fuller, C. (1989). Competitive groups as cognitive communities: the case of Scottish knitwear manufacturers. *Journal of Management Studies*, 26(4):397–416.

Porter, M. E. (1979). The structure within industries and companies' performance. *The Review of Economics and Statistics*, 61(2):214–227.

Porter, M. E. (1980). *Competitive strategy.* Free Press, New York.

Porter, M. E. (1985). *Competitive advantage.* Free Press, London.

Porter, M. E. (1991). Towards a dynamic theory of strategy. *Strategic Management Journal*, 12(S2):95–117.

PricewaterhouseCoopers (2014). Global top 100 companies by market capitalisation. Report, http://www.pwc.com/gx/en/audit-services/capital-market/publications/top100-market-capitalisation.jhtml, accessed 10.11.2014.

Punj, G. and Stewart, D. W. (1983). Cluster analysis in marketing research: review and suggestions for application. *Journal of Marketing Research*, 20(2):134–148.

Pussep, A., Schief, M., and Widjaja, T. (2012). The Software Value Chain: Methods for Construction and Their Application. In *ECIS 2012 Proceedings*, Barcelona.

Pussep, A., Schief, M., Widjaja, T., Buxmann, P., and Wolf, C. M. (2011). The Software Value Chain as an Analytical Framework for the Software Industry and Its Exemplary Application for Vertical Integration Measurement. In *AMCIS 2011 Proceedings*, pages 1–8, Detroit.

Ragin, C. C. (2013). Foreword: the distinctiveness of configurational research. In Fiss, P. C., Cambré, B., and Marx, A., editors, *Configurational theory and methods in organizational research*, page 319. Emerald Group Publishing, Bingley.

Rajala, R. and Westerlund, M. (2007). Business models – a new perspective on firms' assets and capabilities. *Entrepreneurship and Innovation*, 8(2):115–125.

Reger, R. K. and Huff, A. S. (1993). Strategic groups: a cognitive perspective. *Strategic Management Journal*, 14(2):103–123.

Ritchey, T. (1998). General morphological analysis. In *Proceedings of the 16th EURO Conference on Operational Analysis*, pages 1–10.

Royce, W. W. (1970). Managing the development of large software systems. In *Proceedings of IEEE WESCON*, volume 26, pages 1–9, Los Alamitos.

Ruefli, T. W. (1990). Mean-variance approaches to risk-return relationships in strategy: paradox lost. *Management Science*, 36(3):368–380.

Schief, M. (2013). *Business models in the software industry: the impact on firm and M&A performance.* Dissertation, Technische Universität Darmstadt, Darmstadt.

Schief, M. and Buxmann, P. (2012). Business models in the software industry. In *HICSS 2012 Proceedings*, pages 3328–3337, Wailea.

Schief, M., Pussep, A., and Buxmann, P. (2012). Performance of business models: empirical insights from the software industry. In *PACIS 2012 Proceedings*, Ho Chi Minh City.

Schief, M., Pussep, A., and Buxmann, P. (2013). The Impact of Software Business Model Characteristics on Firm Performance. In *ICSOB 2013 Proceedings*, Berlin.

Schimmer, M. and Brauer, M. (2012). Firm performance and aspiration levels as determinants of a firm's strategic repositioning within strategic group structures. *Strategic Organization*, 10(4):406–435.

Schmidt, R. C. (1997). Managing delphi surveys using nonparametric statistical techniques. *Decision Sciences*, 28(3):763–774.

Scott, W. R. (1998). *Organizations: rational, natural and open systems.* Prentice Hall, Englewood Cliffs.

Sharpe, W. F. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442.

Short, J. C., Ketchen, D. J., Palmer, T. B., and Hult, G. T. M. (2007). Firm, strategic group, and industry influences on performance. *Strategic Management Journal*, 28(2):147–167.

Short, J. C., Payne, G. T., and Ketchen, D. J. (2008). Research on organizational configurations: past accomplishments and future challenges. *Journal of Management*, 34(6):1053–1079.

Sinha, T. (1994). Prospect theory and the risk return association: another look. *Journal of Economic Behavior and Organization*, 24(2):225–231.

Slater, S. F. and Olson, E. M. (2000). Strategy type and performance: the influence of sales force management. *Strategic Management Journal*, 21(8):813–829.

Sokal, R. R. and Sneath, P. H. (1963). *Principles of numerical taxonomy*, volume 57 of *A Series of books in biology.* Freeman, San Francisco.

Spanos, Y. E. and Lioukas, S. (2001). An examination into the causal logic of rent generation: contrasting Porter's competitive strategy framework and the resource-based perspective. *Strategic Management Journal*, 22(10):907–934.

Spekman, R. E., Kamauff, J. W., and Salmond, D. J. (1994). At last purchasing is becoming strategic. *Long Range Planning*, 27(2):76–84.

Stabell, C. B. and Fjeldstad, Ø. D. (1998). Configuring value for competitive advantage: on chains, shops, and networks. *Strategic Management Journal*, 19(5):413–437.

Stanoevska-Slabeva, K., Talamanca, C. F., Thanos, G., and Zsigri, C. (2007). Development of a generic value chain for the grid industry. *Lecture Notes in Computer Science*, 4685/2007:44–57.

Statista (2016). The 100 largest companies in the world by market value in 2015. Report, https://www.statista.com/statistics/263264/top-companies-in-the-world-by-market-value, accessed 16.11.2016.

Stelzer, D. (2004). Produktion digitaler Güter. In Braßler, A. and Corsten, H., editors, *Entwicklungen im Produktionsprozessmanagement*, pages 233–250. Vahlen, München.

Thomas, H. and Venkatraman, N. (1988). Research on strategic groups: progress and prognosis. *Journal of Management Studies*, 25(6):537–555.

Tversky, A. and Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business*, 59(4):251–278.

Valtakoski, A. and Rönkkö, M. (2010). Diversity of business models in software industry. In Tyrväinen, P., Jansen, S., and Cusumano, M. A., editors, *ICSOB 2010 Proceedings*, pages 1–12, Berlin, Heidelberg. Springer-Verlag.

Veit, D., Clemons, E., Benlian, A., Buxmann, P., Hess, T., Kundisch, D., Leimeister, J. M., Loos, P., and Spann, M. (2014). Business models – an information systems research agenda. *Business & Information Systems Engineering*, 6(1):1–17.

Veliyath, R. and Ferris, S. P. (1997). Agency influences on risk reduction and operating performance: an empirical investigation among strategic groups. *Journal of Business Research*, 39(3):219–230.

Venkatraman, N. and Ramanujam, V. (1986). Measurement of business performance in strategy research: a comparison of approaches. *The Academy of Management Review*, 11(4):801–814.

Venkatraman, N. and Ramanujam, V. (1987). Measurement of business economic performance: an examination of method convergence. *Journal of Management*, 13(1):109–122.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, 5(2):171–180.

Wiggins, R. R. and Ruefli, T. W. (1995). Necessary conditions for the predictive validity of strategic groups: analysis without reliance on clustering techniques. *Academy of Management Journal*, 38(6):1635–1656.

Williamson, O. E. (1991). Comparative economic organization. *Administrative Science Quarterly*, 36(2):269–296.

Wirtz, B. W., Pistoia, A., Ullrich, S., and Göttel, V. (2015). Business Models: Origin, Development and Future Research Perspectives. *Long Range Planning*, 4(1):1–19.

Witten, I. H. and Frank, E. (2005). *Data mining: practical machine learning tools and techniques.* Elsevier Ltd, Oxford, 2nd edition.

Xu, L. and Brinkkemper, S. (2007). Concepts of product software. *European Journal of Information Systems*, 16(5):531–541.

Zott, C. and Amit, R. (2007). Business model design and the performance of entrepreneurial firms. *Organization Science*, 18(2):181–199.

Zott, C. and Amit, R. (2008). The fit between product market strategy and business model: implications for firm performance. *Strategic Management Journal*, 26(August 2007):1–26.

Zott, C. and Amit, R. (2010). Business model design: an activity system perspective. *Long Range Planning*, 43(2-3):216–226.

Zott, C., Amit, R., and Massa, L. (2011). The business model: recent developments and future research. *Journal of Management*, 37(4):1019–1042.

Zúñiga-Vicente, J. Á., de la Fuente-Sabaté, J. M., and Suárez-González, I. (2004). Dynamics of the strategic group membership–performance linkage in rapidly changing environments. *Journal of Business Research*, 57(12):1378–1390.

Affidavit

I hereby declare that the dissertation entitled *"Firm Strategies and Business Models in the Software Industry: A Configurational Approach"* is my own work. I have used only the sources indicated and have not made unauthorized use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given. I have not presented this thesis or parts thereof to a university as part of an examination or degree.

Anton Pussep, Düsseldorf