Brands and customers as drivers of firm value

Dissertation

Zur Erlangung der Würde des Doktors
der Wirtschafts- und Sozialwissenschaften
des Fachbereichs Betriebswirtschaftslehre
der Universität Hamburg

Vorgelegt von
Dipl.-Kffr. Nicole Martin
aus Hamburg / São Paulo

Hamburg, den 29.08.2012

Prüfungsvorsitzende: Prof. Dr. Nicola Berg

Erstgutachter: Prof. Dr. Henrik Sattler

Zweitgutachter: Prof. Dr. Michel Clement

Disputationstermin: 21.09.2012

Table of Contents

Brands and customers as drivers of firm value

I. Oral presentation and attendances at scientific conferences and seminars

I. Oral presentation and attendances at scientific conferences and seminars

Oral presentation

02 / 2008	Crowdsourcing: Systematisierung praktischer Ausprägungen und verwandter Konzepte, Multikonferenz Wirtschaftsinformatik, München

Scientific conferences

12 / 2011	Marketing Camp, University of Cologne
09 / 2011	SALTY, 13th Conference "Quantitative Marketing", University of Cologne
01 / 2011	Marketing Camp, University of Hamburg
09 / 2010	SALTY, 12th Conference "Quantitative Marketing", University of Mannheim
08 / 2010	PhD Seminar "Quantitative Marketing", University of Hamburg
01 / 2010	Marketing Camp, University of Cologne
09 / 2009	SALTY, 11th Conference "Quantitative Marketing", University of Kiel
07 / 2009	PhD Seminar "Quantitative Marketing", University of Hamburg
05 / 2009	38th Annual Conference of the European Marketing Academy (EMAC), Nantes
01 / 2009	Marketing Camp, University of Hamburg
07 / 2008	PhD Seminar "Quantitative Marketing", University of Cologne
02 / 2008	Multikonferenz Wirtschaftsinformatik, München

Scientific seminars

10 / 2011- 02 / 2012	Lecture course, "Multivariate Research Methods", Prof. Dr. S. Dobnič, University of Hamburg
11 / 2011	Doctoral workshop, "Experteninterviews", Prof. Dr. U. Nagel, University of Hamburg
10 / 2011	Doctoral workshop, "PLS Path Modeling: Methodology and Application", Prof. Dr. C. M. Ringle, University of Hamburg
10 / 2010	Doctoral workshop, "International Research Workshop", Akademie of Sankelmark / University of Southern Denmark: • Data Analysis with R • Introduction to the SOEP / Analysing Panel Data • Structural Equation Modeling with Amos
09 / 2009	Scientific and practical workshop, „Managing Brands and Customers for Profit", Prof. Dr. V. Kumar (Georgia State University), DMC Dialogmarketing Consulting
2009 – 2010	„BASIS Qualifikation für Lehrende der UHH": • „Wissenschaftliches Schreiben – Ins Schreiben kommen" • „Stimm- und Sprechtraining" • „Diskussion anregen und moderieren" • „Umgang mit Lampenfieber"

II. Synopsis: Brands and customers as drivers of firm value

II. Synopsis: Brands and customers as drivers of firm value

Introduction

As CEOs and financial investors focus on maximizing shareholder value, marketing managers are held accountable for showing how their activities generate long-term value to the firm (e.g., Lehmann et al., 2006; Srinivasan & Hanssens, 2009). Marketers who do not identify and measure marketing effectiveness risk their standing within the firm. Moreover, if marketers cannot demonstrate their worth to their firms, marketing will be perceived as less important, will gain less respect in the boardroom, and will have less influence on strategic business decisions (e.g., Rust et al., 2004a; Verhoef & Leeflang, 2009). Therefore, the importance of justifying marketing investments and the metrics necessary to measure marketing effectiveness have gained importance (Leone et al., 2006). In this regard, brand-related and customer-related measurement approaches have come into focus in recent years by academics as well as practitioners (e.g., Ambler et al., 2002).

Brands are one of the most valuable resources of a firm as they can help to increase the level of the firm's cash flow, to realize cash flows earlier, to extend the duration of cash flows or reduce the risk to the firm's cash flows (Madden, 2006; Srivastava et al., 1998), and, thus, are positively linked to firm value. The value of a brand is generally defined as "the marketing effects uniquely attributable to the brand - for example, when certain outcomes result from the marketing of a product or service because of its brand name that would not occur if the same product or service did not have that name" (Keller, 1993, p. 1). Recent research has operationalized brand equity based on a long-term and financially relevant perspective, that is, brand-specific present and future economic benefits (e. g., Barth et al., 1998). Hereby, this perspective adequately reflects the monetary value of marketing effects, attributable to the brand on a financial basis. Brand equity has been linked, for instance, to stock return and systematical risk (Madden, 2006), share prices and return (Barth et al., 1998), market-to-book ratios (Kerin & Sethuraman, 1998), replacement value of firms (Simon & Sullivan, 1993), and firm's stock prices (Aaker & Jacobson, 1994).

From a customer-focused perspective, the effectiveness of marketing activities is reflected by the value that the customer provides to the firm (Berger & Nasr, 1998). As customers are one of the most valuable target group of the firm (e.g., Schulze et al., 2012), the value of customers has been widely discussed and prior research has pointed out various issues in understanding how marketing investments are related to the value of customers (e.g., Berger et al., 2002). In this regard, the value of customers is measured in terms of customer equity which is defined as the sum of customer lifetime values of all customers or group of

customers (e.g., Rust et al., 2004a). Recent research has also related customer equity as a proxy of shareholder value. More precisely, researchers have regressed measures of customer equity on stock prices (e. g., Kumar & Shah, 2009) or compared the value of current and future customer segments against shareholder value (Schulze et al., 2012; Gupta et al., 2004; Rust et al., 2004a).

Even though marketing research has extensively centered the brand and customer perspective, a holistic view of both concepts as well as single facets of each concept have not sufficiently been investigated, as demonstrated in the following:

Marketing research has provided substantial knowledge on the conceptualization and measurement of assets influenced by marketing. However, there is still little research at the marketing-accounting interface (e.g., Hyman & Mathur, 2005) which is needed to improve marketers' standing within a firm (e.g., Madden, 2006). Thus, researchers and practitioners are challenged to value marketing investments in clear financial terms, consistently with accounting standards (Bell et al., 2002). The lack of a marketing-accounting perspective leads to a variety of unconsidered issues, as for instance, the unclarity of how assets influenced by marketing relate to each other including two very prominent ones, namely brands and customers (e. g. Keller & Lehmann, 2006; Kumar et al., 2006).

From a brand-focussed perspective, prior research has focussed on customer-based brand equity approaches which are based on the customers' brand familiarity and brand associations (Keller, 1993). However, the customer-based brand equity perspective can also be related to other relationships than the customer-firm relationship. On the one hand, this perspective can be extended to the context of place branding. Place marketers focus more and more on establishing the city as a brand (Braun, 2008) and try to promote their brand to its different target groups (e. g., citizens). Not only for firms but also for cities, the efficiency measurement of marketing investments (taxpayers' money) is considerably of high importance. Unfortunately, proper brand and target group related efficiency measurements in place marketing practice remain missing (Jacobsen, 2009). On the other hand, the brand-centred perspective can be extended to the firm-employer relationship. Prior research has indicated that the firms' ability to recruit and maintain employees being one of the most valuable resources of the firm can be positively influenced by a favourable, strong and unique employer branding (e. g., Cable & Turban, 2003). In this regard, precise measurement approaches for the added value of employer branding activities and drivers of the employer-based brand equity have been investigated, but only on a rudimentary level.

From a customer-focussed perspective, researchers have concentrated on the value that the customer provides to the firm and, thus, on the allocation of resources away from low-value customers towards high-value customers (e.g., Bell et al., 2002). Predictive statistical methods, referred to as scorecards, are used to assign customers to classes, and provide decision support for appropriate actions or interventions. However, several issues concerning the assessment of these scorecards remain unsolved (e.g., Hand, 2006). A different customer-focused perspective is reflected by concepts that assign customers a more active role than providing the firm with cash flows. Firms have recognized to open their boundaries and leverage external resources (e. g., customers) to the internal organization of the product and/or service creation process with the objective to drive internal growth and create competitive advantage (e. g., Open Innovation; Huston & Sakkab, 2006). Based on recent developments of the web 2.0 and information and communication technologies further concepts of the interactive value creation process have been developed (e. g., Crowdsourcing). However, research has paid little attention to this phenomenon yet.

Against this background, this dissertation addresses the above mentioned research areas by comprising the following six[1] dissertation projects. These dissertation projects focus on: (1) customer-related assets and their contribution to firm value, (2) measuring success in place marketing and branding, (3) measures and drivers of employer-based brand equity, (4) the assessment of scorecard performances, (5) the applicability of the Cost Curve methodology as an assessment instrument of scorecards, and (6) the concept of Crowdsourcing. Valuable implications and gaps for further research are provided. An overview over the research objectives and main results of each dissertation project is given in *table 1*.

Overview of dissertation projects

Customer-related assets and their contribution to firm value: A theoretical framework and empirical application. What is the value of marketing? Many CEOs want an answer to this question and pressure marketers to identify the assets generated by marketing and to monetize the contributions of these assets to firm value. The extant research on marketing-related assets has been limited in two respects: the lack of a comprehensive, *non-overlapping* and *measurable* framework for these assets and the unknown applicability of marketing-related assets to financial accounting. Building on the work of Srivastava et al. (1998) and Rust et al. (2004b), this research offers a customer-related assets (CRAs) framework that uses a customer-centric perspective to identify a comprehensive and mutually exclusive set of

[1] During the period of research, the author of this dissertation has worked on six dissertation projects. Nevertheless, the dissertation project "Measures and drivers of employer-based brand equity" is still in progress and hence, not included in this dissertation.

customer-related assets and that integrates these assets with financial accounting standards. In addition to offering a method of monetizing the contribution of marketing to firm value, the CRAs framework clarifies the controversial relationship between brand equity and customer equity. The authors demonstrate the practical use of this framework for a major European corporation, combining a large-scale empirical study involving survey data with information from the firm's internal databases.

Measuring success in place marketing and branding. As the competition between cities increases, cities focus more and more on establishing themselves as brands. Consequently, cities invest an extensive amount of taxpayers' money into their marketing activities. Unfortunately, cities still lack a proper success measurement, which has raised questions regarding the efficient and effective use of the taxpayers' money. With this contribution the authors want to highlight some existing, but primarily new possibilities for a complex success measurement in place marketing, referring to the extant literature on place marketing and the general field of marketing. Therewith, the authors strive to translate different concepts like customer equity or customer satisfaction into the lexicon of place marketing, thus identifying empirical gaps for further research, as well as existing fruitful approaches.

Measures and drivers of employer-based brand equity. In recent years, firms have expended considerable resources on employer branding activities to attract potential and to maintain current employees. This study focuses on the measurement of the importance of employer branding activities in terms of brand equity. Although brand equity measures for the customer-firm-relationship have attracted a large body of research, the question of how employer-based brand equity can be reliably measured and explained has not been adequately investigated. The authors introduce a model to empirically measure and explain employer-based brand equity with a sample of 1,126 potential employers across 3 work areas. It is suggested that not only low-qualified employees but also high-qualified employees are willing to be paid less and work more for a strong, favorable and positive employer brand. A functional relationship between brand equity measures and brand functions (information function, risk reduction function, internal signaling, social signaling, and professional signaling) is investigated.

Assessing scorecard performances: A literature review and classification. The assessment of scorecard performance in the field of credit scoring is of major relevance to firms. This study presents the first systematic academic literature review of how empirical benchmark studies assess scorecard performance in the field of credit scoring. By analysing 62 comparative studies, this study provides two main contributions. First, this study provides a systematic

overview of the assessment-related decisions of all the reviewed studies based on a classification framework. Second, the assessment criteria of *consistency, application fit*, and *transparency* are introduced and used to discuss the observed assessment-related decisions. As the findings show, researchers often pay insufficient attention to ensuring the consistent assessment of scorecard performance. Moreover, the majority of the reviewed studies choose performance indicators that failed to fit the application context and provided non-transparent assessment documentation. In conclusion, these researchers pay a great deal of attention to the development of scorecards, but they often fail to implement a straightforward assessment procedure.

Applicability of the Cost Curve methodology (Bewertung und Auswahl von Scorecards im Kreditwesen: Eine Analyse zur Eignung von Kosten-Kurven). Credit scorecards are routinely used in the financial service industry to guide decision making in marketing and risk management. The paper is concerned with the problem of identifying an appropriate scorecard among a set of alternatives. To that end, a requirement specification for scorecard assessment in the credit industry is developed. Examining the compliance of current assessment practices with these requirements, the authors find that standard performance measures suffer important limitations. The Cost Curve (Drummond & Holte, 2006) methodology is introduced as a more powerful tool for scorecard selection in credit scoring applications. Its unique advantages are illustrated by means of an empirical study. A key implication of the paper is that Cost Curves facilitate a business oriented scorecard selection and, thereby, contribute toward increasing decision quality in scorecard-supported business processes.

Crowdsourcing (Systematisierung praktischer Ausprägungen und verwandter Konzepte). This paper focusses on the concept of Crowdsourcing which refers to an organizational concept of an interactive product and/or service creation process based on web 2.0 technology. By analyzing different Crowdsourcing communities, the authors introduce the first academic definition of the Crowdsourcing concept and a classification framework to distinguish between different types of Crowdsourcing communities. Accordingly, systematical differences between Crowdsourcing and related concepts, namely Open Source and Open Innovation are detected. It is argued that Crowdsourcing generalizes Open Source and Open Innovation regarding: the motivation of the included persons, the organization of the product and/or service creation process, the aimed objective and the project initiation.

Table 1: Overview of dissertation projects

Author(s)	Status of the Paper	Research Objective	Sample	Method(s)	Main Results
Customer-related assets and their contribution to firm value: A theoretical framework and empirical application					
Paul, P., Martin, N., Sattler, H., & Hennig-Thurau, T. (2012)	Working paper, Targeted to submission to *Journal of Marketing*	This study introduces a customer-related asset framework to identify a comprehensive and mutually exclusive set of customer-related assets and to integrate these with accounting standards. To offering a method of monetizing the contribution of marketing to firm value, the framework clarifies the controversial relationship between brand equity and customer equity.	n = 1,281 adults, representative sample	CBC, HB, OLS, Forecasting of future DCFs	The authors include three types of customer-related assets that drive customers' decisions to buy products (brand, value, relationship management assets). All three assets fulfill the criteria of identifiability, controllability and reliable measurement and, thus, can be referred to as assets from an accounting perspective. As a result, an important basis for researchers and practitioners is provided to value assets in clear financial terms that are consistent with an accounting perspective. For marketers and accountants, this is essential to avoid conceptual overlap and double counting.
Measuring success in place marketing and branding					
Zenker, S., & Martin, N. (2011)	Published, *Journal of Place Branding and Public Diplomacy*	The authors highlight existing but primarily new possibilities for a complex success measurement in place marketing, referring to the extant literature on place marketing and the general field of marketing. This study translates different marketing concepts into the lexicon of place marketing and identifies empirical gaps for further research.	-	Literature review	The absence of a comprehensive performance measurement system becomes obvious as current success measurability typically disregard characteristics of places: the diverse target groups and the complexity of the product itself. The introduced framework of success measurements gives place marketers practical suggestions for measuring the impact of their work. The authors identify gaps for further empirical research and develop a research agenda for place marketing theory.
Measures and drivers of employer-based brand equity					
Erfgen, C., Martin, N., & Sattler, H. (2012)	Project not included in this dissertation, work in progress	The authors investigate the relationship between employer-based brand equity measures and brand functions. Furthermore, it is analyzed if high-qualified employees are willing to be paid less and work more for a strong, favorable and positive employer brand.	n = 1,126 adults, across work areas: automobile industry, manufacturer of sportswear and media enterprises	CBC, HB, OLS	-
Assessing scorecard performances: A literature review and classification					
Martin, N. (2012)	Working paper, To be submitted soon to *Expert Systems with Applications*	It is analyzed how empirical benchmark studies assess scorecard performances in the field of credit scoring. Based on an introduced classification framework, this study provides a systematical overview of assessment-related decisions of empirical benchmark studies. Further, assessment criteria are introduced and used to critically discuss observed assessment related decisions for further research.	n = 62 empirical studies	Literature review	A great lack of consistency in assessment related decisions is detected. It is remarkable that especially those performance indicators are frequently used that have great disadvantages due to assumptions and calculation restrictions. Further, uncertainties and missing transparency about definitions and calculations of well-established and frequently used indicators are observed. It is demonstrated that further research is needed in order to make cost-benefit-wise decisions.
Bewertung und Auswahl von Scorecards im Kreditwesen: Eine Analyse zur Eignung von Kosten-Kurven					
Martin, N., & Lessmann, S. (2012)	Working paper, Submitted to *Zeitschrift für betriebswirt-schaftliche Forschung*	The authors aim at identifying the weaknesses of commonly applied performance indicators with respect to quality criteria and present the Cost Curve methodology as a preferable alternative for the application context of credit scoring. An empirical simulation study is undertaken to evidence the particular suitability of cost curves for credit scoring.	n = 1,000 adults, credit applicants sample (German Credit Data, UCI Machine Learning Repository)	Logistic Regression, C.4.5 Decision Tree, Random-Forest, Cost Curves	As demonstrated by the simulation study, Cost Curves are a proper assessment tool for credit scoring application as this instrument accounts for asymmetric distributions of class and cost information, for different levels of cost and class specific information and time varying changes of these parameters. Cost Curves provide a high degree of information for the assessment of scorecard performances.
Crowdsourcing: Systematisierung praktischer Ausprägungen und verwandter Konzepte					
Martin, N., Lessmann, S., & Voß, S. (2008)	Published, *Multikonferenz Wirtschaftsinformatik Proceedings*	By analyzing different Crowdsourcing communities, the authors introduce the first academic definition of the Crowdsourcing concept and a classification framework to distinguish between different types of Crowdsourcing communities. Accordingly, systematical differences between Crowdsourcing and related concepts, namely Open Source and Open Innovation are detected.	n = 10 Crowdsourcing communities	Literature review	This study identifies characteristics of Crowdsourcing communities and identifies differences and potential overlap between Crowdsourcing and related concepts. More precisely, it is argued that Crowdsourcing generalizes discussed concepts concerning the motivation of the included persons, the organization of the product and/or service creation process, the aimed objective and the project initiation.
Note. CBC = Choice-based conjoint; HB = Hierarchical bayes algorithm; OLS = Ordinary least squares; DCFs = Discounted cash flows					

References

Aaker, D. A., & Jacobson, R. (1994). The Financial Information Content of Perceived Quality. *Journal of Marketing Research*, *31*(2), 191–201.

Ambler, T., Bhattacharya, C. B., Edell, J., Keller, K. L., Lemon, K. N., & Mittal, V. (2002). Relating Brand and Customer Perspectives on Marketing Management. *Journal of Service Research*, *5*(1), 13-25.

Barth, M. E., Clement, M. B., Foster, G., & Kasznik, R. (1998). Brand Values and Capital Market Valuation. *Review of Accounting Studies*, *3*(1), 41–68.

Bell, D., Deighton, J., Reinartz, W. J., Rust, R. T., & Swartz, G. (2002). Seven Barriers to Customer Equity Management. *Journal of Service Research*, *5*(1), 77-85.

Berger, P. D., & Nasr, N. I. (1998). Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing*, *12*(1), 17-30.

Berger, P. D., Bolton, R. N., Bowman, D., Briggs, E., Kumar, V., Parasuraman, A., & Terry, C. (2002). Marketing Actions and the Value of Customer Assets. *Journal of Service Research*, *5*(1), 39–54.

Braun, E. (2008). *City Marketing: Towards an Integrated Approach.* Erasmus Research Institute of Management (ERIM), Erasmus University Rotterdam.

Cable, D. M., & Turban, D. B. (2003). The Value of Organizational Reputation in the Recruitment Context: A Brand-Equity Perspective. *Journal of Applied Social Psychology*, *33*(11), 2244–2266.

Drummond, C., & Holte, R. C. (2006). Cost Curves: An improved method for visualizing classifier performance. *Machine Learning*, *65*(1), 95–130.

Gupta, S., Lehmann, D. R., & Stuart, J. A. (2004). Valuing Customers. *Journal of Marketing Research*, *41*(1), 7–18.

Hand, D. J. (2006). Classifier Technology and the Illusion of Progress. *Statistical Science*, *21*(1), 1–14.

Huston, L., & Sakkab, N. (2006). Connect and Develop (cover story). *Harvard Business Review*, *84*(3), 58–66.

Hyman, M. R., & Mathur, I. (2005). Retrospective and Prospective Views on the Marketing/Finance Interface. *Journal of the Academy of Marketing Science*, *33*(4), 390–400.

Jacobsen, B. P. (2009). Investor-based place brand equity: A theoretical framework. *Journal of Place Management and Development*, *2*(1), 70–84.

Keller, K. L. (1993). Conceptualizing, Measuring, Managing Customer-Based Brand Equity. *Journal of Marketing*, *57*(1), 1–22.

Keller, K. L., & Lehmann, D. R. (2006). Brands and Branding: Research Findings and Future Priorities. *Marketing Science*, *25*(6), 740–759.

Kerin, R. A., & Sethuraman, R. (1998). Exploring the Brand Value-Shareholder Value Nexus for Consumer Goods Companies. *Journal of the Academy of Marketing Science*, *26*(4), 260–273.

Kumar, V., & Shah, D. (2009). Expanding the Role of Marketing: From Customer Equity to Market Capitalization. *Journal of Marketing*, *73*(6), 119–136.

Kumar, V., Lemon, K. N., & Parasuraman, A. (2006). Managing Customers for Value: An Overview and Research Agenda. *Journal of Service Research*, *9*(2), 87–94.

Lehmann, Donald R., & Reibstein, D. J. (2006). Marketing Metrics and Financial Performance. *Marketing Science Institute Monograph*, *2006*.

Leone, R. P., Rao, V. R., Keller, K. L., Luo, A. M., McAlister, L., & Srivastava, R. (2006). Linking Brand Equity to Customer Equity. *Journal of Service Research*, *9*(2), 125–138.

Madden, T. J. (2006). Brands Matter: An Empirical Demonstration of the Creation of Shareholder Value Through Branding. *Journal of the Academy of Marketing Science*, *34*(2), 224–235.

Martin, N., Lessmann, S., & Voß, S. (2008). Crowdsourcing: Systematisierung praktischer Ausprägungen und verwandter Konzepte. In: Martin Bichler, Thomas Hess, Helmut Krcmar, Ulrike Lechner, Florian Matthes, Arnold Picot, … (Eds.), *Multikonferenz Wirtschaftsinformatik, MKWI 2008, München, 26.2.2008 - 28.2.2008, Proceedings.* GITO-Verlag, Berlin.

Rust, R. T., Lemon, K. N., & Zeithaml, V. A. (2004a). Return on Marketing: Using Customer Equity to Focus Marketing Strategy. *Journal of Marketing*, *68*(1), 109–127.

Rust, R.T., Zeithaml, V. A., & Lemon, K. N. (2004b), Customer-Centered Brand Management, *Harvard Business Review*, *82*(9), 110–118.

Schulze, C., Skiera, B., & Wiesel, T. (2012). Linking Customer and Financial Metrics to Shareholder Value: The Leverage Effect in Customer-Based Valuation. *Journal of Marketing*, *76*(2), 17–32.

Simon, C. J., & Sullivan, M. W. (1993). The Measurement and Determinants of Brand Equity: A Financial Approach. *Marketing Science*, *12*(1), 28–52.

Srinivasan, S., & Hanssens, D. M. (2009). Marketing and Firm Value: Metrics, Methods, Findings, and Future Directions. *Journal of Marketing Research*, *46*(3), 293–312.

Srivastava, R. K., Shervani, T. A., & Fahey, L. (1998). Market-Based Assets and Shareholder Value: A Framework for Analysis. *Journal of Marketing*, *62*(1), 2–18.

Verhoef, P. C., & Leeflang, P. S. H. (2009). Understanding the Marketing Department's Influence Within the Firm. *Journal of Marketing*, *73*(2), 14–37.

Zenker, S., & Martin, N. (2011). Measuring success in place marketing and branding. *Place Branding & Public Diplomacy*, *7*(1), 32–41.

III. Customer-related assets and their contribution to firm value: A theoretical framework and empirical application

Authors: **Michael Paul**

Nicole Martin

Henrik Sattler

Thorsten Hennig-Thurau

Year: *2012*

Bibliographic Status: *Working Paper,*

Targeted to submission to

Journal of Marketing

Customer-related assets and their contribution to firm value:

A theoretical framework and its empirical application

Michael Paul*
Marketing Center Muenster
University of Muenster, Germany
Phone (+49) 251 83 22046
Fax (+49) 251 83 22032
Email: michael.paul@uni-muenster.de

Nicole Martin
Institute of Marketing and Media
University of Hamburg, Germany
Phone (+49) 40 42838 8713
Fax (+49) 40 42838 8715
Email: nicole.martin@wiso.uni-hamburg.de

Henrik Sattler
Institute of Marketing and Media
University of Hamburg, Germany
Phone (+49) 40 42838 8714
Fax (+49) 40 42838 8715
Email: henriksattler@googlemail.com

Thorsten Hennig-Thurau
Marketing Center Muenster
University of Muenster, Germany
Phone (+49) 251 83 29954
Fax (+49) 251 83 22032
Email: thorsten@hennig-thurau.de

***corresponding author**

Acknowledgments: The authors thank Britta Lehmann for her support with the data collection.

Working Paper
August 27, 2012

Customer-related assets and their contribution to firm value:

A theoretical framework and its empirical application

Abstract

What is the value of marketing? Many CEOs want an answer to this question and pressure marketers to identify the assets generated by marketing and to monetize the contributions of these assets to firm value. The extant research on marketing-related assets has been limited in two respects: the lack of a comprehensive, *non-overlapping* and *measurable* framework for these assets and the unknown applicability of marketing-related assets to financial accounting. Building on the work of Srivastava et al. (1998) and Rust et al. (2004c), this research offers a customer-related assets (CRAs) framework that uses a customer-centric perspective to identify a comprehensive and mutually exclusive set of customer-related assets and that integrates these assets with financial accounting standards. In addition to offering a method of monetizing the contribution of marketing to firm value, the CRAs framework clarifies the controversial relationship between brand equity and customer equity. The authors demonstrate the practical use of this framework for a major European corporation, combining a large-scale empirical study involving survey data with information from the firm's internal databases.

Keywords: Customer-related asset, accountability, brand equity, customer equity

Customer-related assets and their contribution to firm value: A theoretical framework and its empirical application

Introduction

Because CEOs and financial investors focus on maximizing shareholder value, marketing managers are held accountable for demonstrating the ways in which their activities generate assets that add to the value of the firm (e.g., Lehmann et al. 2006; Srinivasan & Hanssens, 2009). Marketers who do not identify and measure the assets that they create and maintain risk their standing within their firms. Moreover, if marketers cannot demonstrate their worth to their firms, marketing will be perceived as less important, will gain less respect in the boardroom, and will have less influence on strategic business decisions (e.g., Rust et al., 2004a; Verhoef & Leeflang, 2009). Thus, one of the greatest challenges for marketing today is to convincingly establish its contributions to firm value (e.g., Hanssens et al., 2009). We argue that accomplishing this task requires the definition and measurement of marketing-related assets in a comprehensive and mutually exclusive way that is consistent with accounting standards. These requirements are essential because they enable marketers and accountants to avoid conceptual overlaps and double counting.

The extant research on marketing has been limited by the lack of a mutually exclusive way to define and measure marketing-related assets. Srivastava et al. (1998) introduced a widely acknowledged "marketing-finance framework" that demonstrates how marketing instruments generate assets and how these assets relate to shareholder value. However, these authors did not seek to identify a comprehensive set of marketing-related assets. Instead, they named exemplary market-based assets and concluded that they "[…] are far from developing a theory that […]

identifies the range and extent of such assets" (Srivastava et al., 1998, pp. 14). As a result, a considerable lack of clarity exists with respect to the relationships among marketing-related assets; in particular, it is unclear how the two very prominent assets of brand equity and customer equity relate to each other (e.g., Keller & Lehmann, 2006; Kumar et al., 2006). For instance, Ambler et al. (2002) note that "[t]here remains much confusion […] regarding the definitions of brand equity and customer equity and the extent to which the two are related or distinct" (Ambler et al., 2002, p. 14). Furthermore, we agree with Bell et al. (2002), who state that marketing has provided substantial knowledge on the conceptualization and measurement of marketing-related assets but that "[m]oving forward, it will be important to move toward the development of some accounting standards for the measures of [these types of assets] that could lead to their placement on the balance sheet […]" (Bell et al., 2002, p. 81). Unfortunately, there is generally little research that occurs at the interface between marketing and accounting (e.g., Hyman & Mathur, 2005). To manage this interface, marketers must provide definitions of generated assets that are consistent with the definitions that are set forth by accountants.

From an accounting perspective, a marketing-related asset is typically an intangible asset that accountants define as "an identifiable non-monetary asset with physical substance" (IAS 38.8). Intangible assets should fulfill the criteria of identifiability and controllability, which are necessary for reliable asset measurability In particular, an asset is identifiable if it is separable or if it arises from contractual or other legal rights (IAS 38.11-12), and it is controllable if a firm not only controls the asset itself but can also ensure that the future economic benefits from the asset in question will flow to the firm (IAS 38.13). In addition, assets must be reliably measurable. Assets are required to fulfill all of these aforementioned criteria to be recognized by accountants. From a marketing perspective, the literature has not established whether marketing-

related assets are identifiable, controllable or reliably measurable. With respect to the identifiability of assets, Gupta & Zeithaml (2006) note that "[m]any researchers are confused about the differences between brand equity and [customer equity]. Is one a subset of the other? Does brand equity affect [customer equity], or is it the other way around?" (Gupta & Zeithaml, 2006, p. 735). In addition, the marketing literature does not address the controllability of an asset by the firm. Thus, for instance, the concept of customer equity (e.g., Blattberg & Deighton, 1996; Blattberg et al., 2001) is widely discussed as an asset within the marketing literature (e.g., Wiesel et al., 2008). However, it remains unclear whether accountants recognize customer equity as an asset because, aside from contractual rights that involve customers, the customer base itself cannot be controlled by the firm.

Building on the work of Srivastava et al. (1998) and Rust et al. (2004c), our research aims to offer a framework of customer-related assets (CRAs). Customer-related assets refer to the benefits that a firm offers to the customer and that directly influence the customer's decision to buy the firm's products. This framework meets accountants' requirements (i.e., identifiability, controllability, and reliable measurability) for assets and can therefore be used to determine the contributions of marketing to a firm's value. In particular, we adopt a customer-centric perspective (e.g., Shah et al., 2006) to introduce the concept of customer-related assets (CRAs). Drawing on prior research (Rust et al., 2004a; Rust et al., 2004b; Rust et al., 2004c; Srivastava et al., 1998), we define three types of customer-related assets, namely, the *brand asset*, *value asset*, and *relationship management asset*. We concentrate on these customer-related assets because they reflect benefits that directly influence the customer's decision to buy the products (here and in the remainder of the paper the term 'products' comprises either physical products or services) that are offered by the firm. Because our definitions and the measurement approaches with

respect to these three marketing assets are in line with international accounting standards (e.g., IAS/IFRS), these perspectives provide a basis for including marketing-related assets in financial accounting.

Overall, this research contributes to the literature because it marks the first attempt to develop a framework of a measurable, comprehensive, and mutually exclusive set of customer-related assets and to integrate these assets with financial accounting standards. In addition to offering a means of monetizing the contributions of marketing to firm value, the CRAs framework clarifies the controversial relationship between brand equity and customer equity. We provide evidence for the applicability of our framework by empirically demonstrating the uses of this framework in the context of the analysis of a large European corporation; this empirical examination combines a large-scale empirical study involving survey data with information from the firm's internal databases.

The remainder of this article is organized as follows. First, we review existing research on assets in accounting and marketing and define the concept of customer-related assets as it is used in this research. We then introduce our conceptual framework and outline how customer-related assets can be measured. Next, we illustrate the applicability of our approach. We conclude with a discussion that considers the implications of this approach for marketing theory and practice.

Asset definitions in accounting and marketing

The accounting perspective

An important element of financial reporting is the provision of information about a firm's assets. The International Accounting Standards Board (IASB) framework builds on the premise that firms typically use their assets to produce goods or services. The goods and services of a firm are capable of satisfying the needs of customers; as a result, these customers are willing to

pay for these goods, a process that contributes to the firm's cash flows. Accounting standards, such as IAS/IFRS, generally pursue an accounting strategy that is based on the conservatism and creditor protection principles (e.g., Gray, 1988).

Although tangible assets that have physical substance (e.g., property, plant, and equipment) are recognized on the balance sheets, financial accounting rarely captures intangible assets, despite the fact that these assets often make up the majority of a firm's value (Lehmann, 2004). Intangible assets, including customer-related assets, can only be recognized if they meet the requirements of IAS 38.11-13 and IAS 38.21. In particular, intangible assets must be identifiable, which is fulfilled if an asset "(a) is separable, i.e., is capable of being separated or divided from the entity and sold, transferred, licensed, rented or exchanged, either individually or together with a related contract, identifiable asset or liability, regardless of whether the entity intends to do so; or (b) arises from contractual or other legal rights, regardless of whether those rights are transferable or separable from the entity or from other rights and obligations" (IAS 38.11). The recognition of an intangible asset also depends on the controllability of the asset, which is provided if and only if "[…] it is probable that the expected future economic benefits that are attributable to the asset will flow to the entity" (IAS 38.21). Additionally, an intangible asset must be reliably measurable (IAS 38.21). Thus, from the accounting perspective, intangible assets are required to fulfill the criteria of identifiability, controllability, and reliable measurability.

Typically, intangible assets can only be activated if they are acquired from third parties, which are activated on the basis of their purchase prices (IAS 38.25-32). Thus, internally generated intangible assets, such as customer lists, are typically not recognized as assets by accounting standards (IAS 38.63). This condition applies because most intangible assets cannot

fulfill the criteria of identifiability, controllability, and reliable measurability. For instance, the International Accounting Standards Board (2007) emphasizes that customer-related assets often have "[…] insufficient information for valuation to recognize them separately from goodwill" (International Accounting Standards Board, 2007, p. 8).

The marketing perspective

A close examination of the extant marketing literature reveals that, with only a few exceptions, marketers have refrained from offering a rigorous definition of the concept of an asset. Furthermore, within the marketing literature, inconsistent terminology has been chosen, thus preventing marketing scholars from developing an asset systematization scheme that is recognized by accountants.

There are three main literature streams within the marketing field that focus on assets or on related (and often synonymously used) concepts, such as "equity" or "value". The first marketing literature stream utilizes a non-monetary "asset" perspective that is based on the customer's mindset (e.g., Aaker, 1996; Keller, 1993; Zeithaml, 1988). Although this research is important for understanding consumer behavior and identifying drivers of customer-related assets, it is not useful for financial valuation purposes because of its non-monetary perspective. A second literature stream adopts a monetary "asset" perspective that focuses on short-term monetary benefits. For instance, a widely applied method for measuring the monetary value of a brand is the price premium approach, that is, the ability of a brand to charge a higher price than an equivalent unbranded product (e.g., Ailawadi et al., 2003; Aaker, 1996). Because this stream of research neglects the future economic benefits of an asset, it is not suitable for financial accounting purposes (e.g., IAS 38.13). A third literature stream focuses on marketing investments in terms of the financial value of an asset, considering both the present and future

economic benefits of the asset in question. Within this perspective, several marketing concepts have been discussed, especially the concepts of brand equity (e.g., Mahajan et al., 1994; Simon & Sullivan, 1993) and customer equity (e.g., Kumar et al., 2006; Gupta & Zeithaml, 2006). To address these concepts, Rust et al. (2004c) introduced a framework that emphasizes customer equity as a key marketing concept that is driven by three constructs, namely, brand equity, value equity, and relationship equity. Rust et al. (2004c) define these drivers as subordinate to the broader notion of customer equity. These authors focus on the discussion of whether "brands exist to serve the customers" (Rust et al., 2004c, p. 110) or whether the converse relationship holds, and they emphasize the move from a brand-centric perspective toward a customer-centric approach. However, they do not answer the question of whether the aforementioned customer equity drivers can be referred to as assets from an accounting perspective. In particular, it remains unclear whether customer equity drivers and the concept of customer equity itself meet the accounting requirements of identifiability, controllability, and reliable measurability.

Only a few marketers explicitly offer a definition of an "asset" (Srivastava et al., 1998; Rust et al., 2004a). Srivastava et al. (1998) introduced the concept of "market-based assets" and define these assets as "any physical, organizational, or human attribute that enables the firm to generate and implement strategies that improve its efficiency and effectiveness in the marketplace" (Srivastava et al., 1998, p. 4). Rust et al. (2004a) define assets as "customer-focused measures of the value of the firm (and its offerings) that may enhance the firm's long-term value" (Rust et al., 2004a, p. 78). Both of these research teams consider marketing assets to reflect benefits that directly influence the customer's decision to buy a firm's products; however, neither set of researchers focuses on the characteristics of identifiability, controllability, and reliable measurability, which are the key traits of an asset from the accounting perspective.

In general, from the perspective of many marketing researchers, the concept of assets is not separated from the outcomes that are generated by these assets (i.e., cash flows); in addition, "equity" is the umbrella term that is used to merge assets and cash flows into a joint concept. This practice has already been criticized by certain marketing scholars, and Leone et al. (2006) note that customer equity is "capturing value 'created' by branding activities under the value 'extracted' from customers" (Leone et al., 2006, p. 128). Ambler et al. (2002) found that the "customer equity perspective focuses on the customer's profitability, but the profitability is often driven by what the consumer thinks of the brand" (Ambler et al., 2002, p. 14). Ambler et al. (2002) also suggest that "[…] brand equity is […] an asset and customer equity […] the financial (dollar) value of an asset […]" (Ambler et al., 2002, p. 14). Thus, customer equity is not separated from the other assets of the firm (e.g., brand) and therefore does not meet the criterion of being identifiable.

In accordance with the reasoning that is set forth above, the traditional accounting perspective typically suspects that a potential overlap exists "in the value between a brand and a customer relationship, and possibly goodwill" (International Accounting Standards Board, 2008, p. 13). Accountants have argued that assets that have no contractual basis and are simply created and maintained by marketing are typically not separable because the cash flows generated from these assets are inextricably linked to the cash flows of the business as a whole (IAS 38.51). These inconsistencies may explain why established customer metrics from marketing, such as customer equity or brand equity, have not been included in accounting standards and accounting frameworks (e.g., Gleaves et al., 2008).

A customer-centric framework of customer-related assets

An overview of the framework

As the conceptual basis of our framework, we build on the idea of customer centricity, which combines several streams in marketing research that believe that a firm's thinking and actions should focus on the wants and needs of its customers (e.g., Shah et al., 2006). This idea provides the conceptual basis for the development of a framework of customer-related assets. *Figure 1* illustrates our conceptual framework for identifying and structuring customer-related assets. With the following descriptions, we start from the right side of *figure 1* and end at the left side of the *figure*.

[Insert Figure 1]

Part I (figure 1): Our framework implies that the sum of all assets that directly influence the customer's behavior (customer-related assets) approximates the firm's operating value. The firm's operating value is the outcome of the transactions with all existing and potential customers from operating activities (*part I*, *figure 1*) because these transactions produce the great majority of a firm's total value (e.g., Gupta et al., 2004; Bauer & Hammerschmidt, 2005). Although a firm may also receive cash from other activities, for example, financing activities that lie outside its core business, the sale of products is typically a firm's most important source of cash flows (e.g., Damodaran, 2006). In our framework, we exclude non-operating assets (for instance, securities that are transferable for speculative purposes) because we assume that the main value of a firm share is represented by the sum of its customer-related assets.

Part II (figure 1): Cash flows from customers are directly generated by customer-related assets[2], which are characterized by the benefits that these assets offer to customers. Thus, customer-related assets can be regarded as drivers of firm value. Referring to assets as the center of this framework, we build on prior marketing research (Srivastava et al., 1998; Rust et al., 2004c) and include three types of customer-related assets that are comprehensive and non-overlapping. In particular, our framework distinguishes between *brand assets*, which we define as the benefits that a firm offers its customers through its brand; *value assets*, which are defined as the benefits that a firm offers its customers through the quality of the core attributes of its products; and *relationship management assets*, which we define as the benefits that a firm offers to its customers through its use of individual customer information to manage its relationships with these customers.[3]

Part III (figure 1): We argue that non-customer-related assets are only indirectly related to customer cash flows. Non-customer-related assets are not necessarily perceived by customers or considered to be beneficial, but these assets can contribute to *brand*, *value*, and/or *relationship management assets*. Non-customer-related assets help to create or leverage customer-related assets and can therefore be referred to as drivers of customer-related assets. Because we assume that customer-related assets are a proxy for the total firm value, separate estimation of the values of customer-related and non-customer-related assets would produce double-counting effects; therefore, we instead assume that the impact of non-customer-related assets is accounted for by

[2] We focus on cash inflows from customers, but we acknowledge that cash outflows to customers also exist. The inflow focus appears to be appropriate; individual customer revenues (i.e., cash inflows) vary to a large extent, whereas costs (i.e., cash outflows) are typically very similar across customers. In our valuation model, we therefore model cash outflows on the aggregate level (as an average across customers).

[3] Contractual relationships, that is, customer relationships that are ensured by a contract, play a predominant role in the accounting literature. In accordance with IAS 38.12, these contracts provide sufficient information for contractual relationships to be recognized as intangible assets. However, the economic benefits of customer contracts are causal products of a firm's brand, the quality of its products and/or its relationship management efforts. Therefore, we focus on our three assets and assign contractual relationships to their original driver. Theoretically, the value of the customer contract (as an asset) could also be identified separately. However, the usage of both estimation procedures is not appropriate because this approach would lead to double counting.

the value of customer-related assets. For example, production plants, patents, and skilled and motivated employees influence *value assets* and are typically created or maintained by business functions other than marketing (e.g., procurement, production, or human resources). As a consequence, a firm's marketing department contributes to the quantity of customer-related assets by developing a strong brand, ensuring high product value, and using customer information to manage relationships with customers (e.g., by offering valuable loyalty programs). At the same time, other departments will also contribute through the management of various non-customer-related assets that influence a firm's quantity of customer-related assets.

Part IV (figure 1): All of the different types of assets require investments and activities from various business functions, such as marketing, to be developed and maintained. *Brand assets* will primarily be affected by particular marketing investments, such as advertising that increases brand awareness and brand image. Other functions, such as procurement, R&D, and human resources, are indirectly involved in the creation of brand assets in that they are required aspects of performing brand management activities. *Value assets* will primarily be created by marketing product management activities (e.g., market research for product innovation). *Relationship management assets* typically result from other particular marketing functions, such as customer service and sales, in which information on customers is gathered and then used for direct marketing, loyalty programs or other relationship management activities. In the following discussion, we will concentrate on cash flows (*part I*, *figure 1*) and customer-related assets (*part II*, *figure 1*), which constitute the center of the proposed framework.

The link between cash flows and customer-related assets

Customer-related assets drive customers' decisions to purchase products. Through the use of customer-related assets, firms can generate future cash flows from existing or future

customers who buy existing products or any new products these firms may eventually create. The economic value of all of these customer-related assets is reflected by the present value of the future benefits that flow from the customer to the firm.

From an accounting perspective, three approaches have been discussed specifically with respect to the valuation of these types of benefits: the cost approach, the market approach, and the income approach (e.g., DIN ISO 10668, 2010). The cost approach has strong limitations (e.g., Reily & Schweihs, 1999), in particular because it is based on historical costs or replacement costs that typically do not consider future benefit developments across the economic lifetime of an asset. The market approach assumes the existence of appropriate information that has been generated by third party transactions involving comparable assets. For the relatively frequent situation in which there is an absence of comparable data from other transactions, marketers and many accountants recommend the use of the income approach (e.g., Day & Fahey, 1988; DIN ISO 10668, 2010). The income approach measures the present value of the economic benefits that are expected to flow to the firm over the remaining economic life of a particular asset (an approach that corresponds to DIN ISO 10668, 2010).

We focus on the idea of the income approach and estimate the sum of future economic benefits via the discounted cash flow approach (DCF). DCF analysis is a widely accepted valuation technique in finance (e.g., Rappaport, 1986), accounting (e.g., Husmann & Schmidt, 2008), and marketing (e.g., Day & Fahey, 1988; Srivastava et al., 1998; Srivastava et al., 1999). Because many accountants use DCF methods in their own decision making, the use of the DCF method provides marketing managers with a common language that they can use to communicate the importance of customer-related assets to their peers in accounting and finance (Gruca & Rego, 2005; DIN ISO 10668, 2010).

Within our framework, we are interested in splitting up the estimated cash flows to the different cells depending on the combinations of existing or potential new product(s) and existing or potential new customers of the firm (*part II*, *figure 1*). Equation (1) states that the total discounted cash flows across all four cells DCF^{I-IV} approximate the total (operating) value of the firm[4] (in the following discussion, we use the term cash flows to indicate operating cash flows):

$$\textbf{(1)} \qquad Total\ value = \sum_{I}^{IV} DCF^{I-IV}$$

Each quadrant *Q* from *I-IV* builds on the following estimation assumptions. We assume a planning horizon within which we explicitly estimate future cash flows. Beyond the scope of this assumption, we value the long-term expected growth of an asset on the basis of its remaining useful economic life. The useful life is the remaining period during which an asset is expected to be available for use by the firm (IAS 38.8). If appropriate, an indefinite lifetime can be assumed (DIN ISO 10668, 2010). Furthermore, risks that are not captured by future cash flows must be considered in terms of the discount rate. As recommended by DIN ISO 10668 (2010) and Damodaran (2006), we discount future expected cash flows by a risk-adjusted discount rate in the form of the weighted average cost of capital. Moreover, for reasons of simplicity, we neglect both depreciation and changes in net working capital. Against this background, each of the quadrants (*part I*, *figure 1*) is estimated as follows:

The *first type of cash flow* (*Q I*) results from existing customers who buy existing products from the firm. These customers make these purchases because they have already experienced the product and its benefits, and these purchases reflect the brand, the value and/or

[4] More precisely, customer cash flows approximate the value of a firm's operating assets. Thus, for firms that own significant cash or non-operating assets, these assets should be added to the value of the operating assets to arrive at the firm value. However, in most situations, operating assets are a valid proxy for firm value (e.g., Damodaran, 2006).

the relationship management assets of the firm. Cash fls from existing customers who buy existing products are the cornerstone of all equity models, such as customer equity (e.g., Gupta et al., 2004; Kumar et al., 2006) and brand equity (e.g., Srivastava et al., 1998; Bahadir et al., 2008). With respect to this concept, the first type of cash flow can be measured as follows:

$$DCF^{I} = \sum_{i=1}^{N_I} \left(\sum_{p=1}^{P} \left(\frac{EBIT_1 \times \widehat{K}_{i,p} \times (1 - T_p)}{(1 + d_1)^p} \right) + \frac{EBIT_1 \times \widehat{K}_{i,P+1} \times (1 - T_{P+1})}{(1 + d_1)^{P+1}} \times \frac{1}{d_1} \right) \quad (2)$$

In the above equation, DCF^I describes the discounted cash flows across all customers $i = 1 \ldots N_I$ in cell I, $EBIT_1$ is the earnings before interest and the tax margin for existing products, $\widehat{K}_{i,p}$ and $\widehat{K}_{i,P+1}$ are the expected revenues for customer i, d_1 is the discount rate for the existing product, and T_p and T_{P+1} are the tax margins that are involved in the purchase. As a standard case, we estimate the discounted cash flows to be infinite, but we differentiate between a planning horizon *P,* in which benefits are explicitly estimated for each customer and each period, and the time period after the planning horizon, *P+1,* in which cash flows are modeled as a perpetuity using the cash flows of a customer *i* from the first year after the planning horizon.

The *second type of cash flow* (*Q II*) results from potential new customers who buy existing products. New customers can be acquired through channels such as advertising, direct marketing, sales promotions, or word-of-mouth. Several customer equity approaches consider cash flows from potential new customers (e.g., Gupta et al., 2004; Kumar & Shah, 2009). Certain approaches also account for the role of existing customers in attracting new customers for a firm through referrals (e.g., Villanueva et al., 2008). The customer acquisition potential of brands is accounted for by brand equity models. For instance, strong brands can attract new customers for existing products (e.g., Keller, 2007). The cash flows from potential new customers buying existing products are captured by equation 3:

(3) $$DCF^{II} = \sum_{i=1}^{N_{II}} \left(\sum_{p=1}^{P} \left(\frac{EBIT_1 \times \widehat{K}_{i,p} \times ra_{i,p} \times (1 - T_p)}{(1 + d_1)^p} \right) + \frac{EBIT_1 \times \widehat{K}_{i,P+1} \times ra_{i,P+1} \times (1 - T_{P+1})}{(1 + d)^{P+1}} \times \frac{1}{d_1} \right)$$

DCF^{II} is the sum of discounted cash flows from all customers $i = 1 \ldots N_{II}$ in cell II. The parameters that were introduced in the discussion of $Q\ I$ are used in the above equation; furthermore, the parameters $ra_{i,p}$ and $ra_{i,P+1}$ are introduced to account for the probability that a new customer will buy an existing product in each of the examined periods. This probability ranges between 0 and 1.

The *third type of cash flow (Q III)* results from existing customers who buy potential new products from the firm. A firm may expect that its existing customers are interested in buying new products after they have experienced the firm's existing products and consider the brands, values, and relationship management approaches of the firm to be beneficial. Potential new products are typically not incorporated into customer equity models. By contrast, there are several brand equity models that include this third type of cash flow in the form of the (option) value of brand extensions (e.g., Park & Srinivasan, 1994). The value of *Q III* is measured as follows (equation 4):

(4) $$DCF^{III} = \sum_{i=1}^{N_{III}} \left(\sum_{p=1}^{P} \left(\frac{EBIT_2 \times \widehat{K}_{i,p} \times (1 - T_p)}{(1 + d_2)^p} \right) + \frac{EBIT_2 \times \widehat{K}_{i,P+1} \times (1 - T_{P+1})}{(1 + d_2)^{P+1}} \times \frac{1}{d_2} \right) \times rm - C$$

In the above equation, DCF^{III} is the sum of discounted cash flows from all of the customers $i = 1 \ldots N_{III}$ in cell III, $EBIT_2$ is the earnings before interest and the tax margin for the new product, and d_2 is the corresponding discount rate. In addition, rm accounts for the probability that a new product will be introduced into the market. This parameter ranges from 0 to 1. The sum of the discounted cash flows across the customers is reduced by the costs of developing and introducing the new product into the market *C*. The discounting of *C* is not necessary if *C* spans a short time period.

Finally, the *fourth type of cash flow (Q IV)* includes potential new customers who buy potential new products. New customers for these potential products can be acquired through advertising, direct marketing, sales promotions, or word-of-mouth references from customers regarding existing products (e.g., Arndt, 1967). This cash flow type is associated with the highest level of uncertainty. Similar to the third type of cash flow, cash flows from new products and new customers are typically not incorporated into customer equity models, whereas brand equity models may include these cash flows in the form of the (option) value of brand extensions. Equation 5 represents the sum of the discounted cash flows DCF^{IV} from new customers $i = 1 \ldots N_{IV}$ in cell IV who buy a new product:

$$\textbf{(5)} \quad DCF^{IV} = \sum_{i=1}^{N_{IV}} \left(\sum_{p=1}^{P} \left(\frac{EBIT_2 \times \widehat{K}_{i,p} \times ra_{i,p} \times (1 - T_p)}{(1 + d_2)^p} \right) + \frac{EBIT_2 \times \widehat{K}_{i,P+1} \times ra_{i,P+1} \times (1 - T_{P+1})}{(1 + d_2)^{P+1}} \times \frac{1}{d_2} \right) \times rm - C$$

Customer-related assets: Brand, value, and relationship management

Building on the extant research, we focus our analysis on three types of customer-related assets (i.e., brand, value, and relationship management), based on discussions within the marketing (Rust et al., 2004c) and accounting (e.g., Barth et al., 1998; Kallapur & Kwan, 2002) literature. The terms brand, value, and relationship management are based on Rust et al. (2004c) but will be extended in their definitions and measurement approaches to meet our objective of addressing the accounting perspective.

Brand assets. *Brand assets* refer to the benefits that a brand adds to a product as perceived by the potential and/or existing customers of a firm (e.g., Rust et al., 2004c; Farquhar, 1989). As described above, a commonly applied method of isolating the monetary value of a brand is the brand price premium approach, which is based on the ability of a brand to charge a higher price than an equivalent unbranded product (e.g., Ailawadi et al., 2003; Aaker, 1996).

Following Park & Srinivasan (1994) and Srinivasan (1979), we argue that this price premium can be measured as the difference between a consumer's (a) perceived overall brand benefit and (b) perceived benefits on the basis of objectively measured product attribute levels. Thus, we distinguish between the benefits of the brand and the benefits that a firm offers to its customers through the perceptions of the objectively considered product attribute levels. Based on the brand price premium, we predict brand-specific cash flows and, in accordance with a widely accepted financial approach, measure brand benefits as the sum of future discounted brand-specific cash flows (e.g., Bahadir et al., 2008).

Our conceptualization of a *brand asset* is consistent with the accounting definition of an intangible asset: From an accounting perspective, one critical prerequisite is the identifiability of *brand assets*; that is, brand-specific cash flows should be separable from those of other assets. Brands are capable of being separated or divided from the firm. In practice, there are several examples in which brands have been sold or licensed, for instance to access new markets or to strengthen their position in current markets (e.g., Mahajan et al., 1994). Furthermore, brand-specific cash flows can be controlled by a firm. A firm has the power to obtain brand-specific cash flows and restrict the access of other firms to those cash flows. In financial accounting, trade names and trademarks are used as synonyms for the brand elements that identify one firm's products as being distinct from the products of others; these elements can be legally protected.

Value assets. In accordance with multi-attribute models of product quality (e.g., Fishbein, 1963), we define *value assets* as the benefits that a firm offers to its customers through the perceptions of the objectively considered quality of the core product and its attributes (Park & Srinivasan, 1994; Rust et al., 2004c; Srinivasan, 1979). High quality is beneficial to customers and has been extensively shown to positively influence customer purchase decisions by fostering

favorable behavioral intentions and by reducing the likelihood of switching to another firm (e.g., Gupta & Zeithaml, 2006; Zeithaml et al., 1996). Similar to the *brand asset* perspective, we suggest calculating value benefits as the sum of future discounted value-specific cash flows. These cash flows are isolated based on a price premium approach that reflects the ability to charge a higher product price for a product with a certain product quality level than for a product with the same brand benefits (and the same relationship management benefits; see below) but a minimum level of objectively measured product quality.

Our definition of a *value asset* is consistent with the general accounting definition of an asset. Researchers have provided evidence that product quality is identifiable, although this task is challenging (e.g., Tellis & Johnson, 2007). The *value asset* meets the separability criterion because this asset can be separated from the firm and sold, potentially in combination with other assets, as demonstrated through discussions in the accounting literature (Financial Accounting Standards Board, 2007, 141.A22.b). As an example, assume the case that a firm owns a registered brand and the documented but unpatented technical expertise for producing the product attributes of the branded product. Furthermore, assume that to transfer the ownership of the brand the owner can be required to transfer everything that is required to produce the branded product to the new owner. The unpatented technical production expertise must be separated from the original owner and may be sold if the related brand is sold; this phenomenon implies that the key competences of the product meet the separability criterion. Therefore, we argue that, at least in many industries, the quality of a product is typically dependent on firm-specific patented or unpatented production expertise or plants and is therefore identifiable. Furthermore, this asset is controlled by the firm because a firm has control over the asset and its future economic benefits will flow to the firm. Firms have the ability to restrict the access of others to those benefits.

Relationship management assets. *Relationship management assets* use customer information to generate benefits. Customer relationship management activities (e.g., loyalty programs) create customer benefits that cannot be directly attributed to the objectively considered quality of the physical product itself or to a firm's brand. These relationship management benefits result from a retrospective database that may include descriptive characteristics of customers in the form of personal information (e.g., the customer's name and address) as well as sales and/or usage history data. As previous literature has shown, the benefits that arise from a firm's relationship management activities are conceptually distinct from other types of benefits (Hennig-Thurau et al., 2002). As for *brand* and *value assets*, we measure the *relationship management assets* as the sum of the discounted earnings that are generated by the *relationship management assets*. Again, these cash flows are isolated based on a price premium approach that reflects the considered benefit of these assets, which is measured as the ability to charge a higher product price due to a certain level of customer relationship management activities in comparison to a product with the same brand benefits and the same levels of objectively quality but with a minimum level of customer relationship management activities. Thus, this measurement approach includes benefits that exceed the value derived from the physical product itself or from the brand.

Although a *relationship management asset* does not necessarily arise from contractual or other legal rights, it is consistent with the accounting definition of an intangible asset. The asset itself and the future economic benefits of the *relationship management asset* are identifiable, given that this customer information can be separated from the firm and, if desired, can be leased, exchanged or sold to others (for instance, the factoring option enables firms to sell their accounts receivable to a factoring company in exchange for cash (e.g., Soufani, 2002)). From an

accounting perspective, one example of a *relationship management asset* is a customer list (e.g., IAS 38.9) that can be acquired in a business combination (e.g., a merger) and therefore typically meets the separability criterion (e.g., Financial Accounting Standards Board, 2007, 142.A37). Furthermore, this asset can be controlled by the firm, as typically no other firm can use the specific information a firm has about its own customers (e.g., purchase histories and detailed preferences).

An illustrative application

Data collection

Focusing on the relationship between customer-related assets and future benefits, we present results from estimating both parts (*part I and II*, *figure 1*) of the proposed framework. The illustrative application is based on the analysis of one of the largest companies in the European personal transport industry. This analysis was conducted with that firm's cooperation. Within this firm, we focus on the long-distance railway passenger segment, which is the most important segment for this firm. The data collection procedure is based on two different sources of information, namely, customer survey data (regarding buying behavior and preference structure) and financial firm data from the cooperating business (*figure 2*).

[Insert Figure 2]

Customer survey data

We used data from an online survey collected in May and June 2010. Potential respondents were contacted at random by recruiters by email based on a stratified random sampling of the German population (14+ years old) with age and gender as interlocked strata. Respondents were selected only if they planned to purchase a railway passenger travel (> 100

km) offer (an existing product segment) from the considered firm within the next 12 months. We sampled three key customer segments, which were categorized on the basis of the firm's internal customer segmentation standards, to capture typical buying behavior. The respondents were either "heavy users" who had travelled at least three times within the past 12 months, "normal users" who had travelled once or twice within the last 12 months, or "potential users" who had not travelled with the considered firm within the last 12 months but who planned to purchase one or more travel offers from the considered firm within the next 12 months.

The survey contained two parts. First, we obtained a sample of 1,663 respondents who answered questions about their buying behavior regarding offers of passenger railway transport (of >100 km) from the considered firm. This behavior included their individual past, present and future purchase behaviors between 2007 and 2012; their previous usage of discount cards and loyalty programs; specific contractual information; and their reasons for purchasing these offers and engaging in travel. In addition, these respondents were confronted with decisions regarding their intended choice behavior with respect to certain airline flights, including a hypothetical option to choose a flight from the examined transport firm; this option constituted brand an extension X, with the core brand of the firm as the parent brand. According to expert interviews with company managers, this brand extension is a potentially important new business segment for the transport firm in question. A total of 831 out of the 1,663 respondents answered questions about the hypothetical brand extension X and the flight segment.

Second, two within-subject choice-based conjoint experiments were conducted, one regarding the passenger railway segment and the other within the passenger flight segment. In both experiments, we gave respondents ten choice sets and instructed them to choose among travel alternatives. Each choice set contained three alternatives and a "none" purchase

alternative. The setting of each choice-based conjoint experiment incorporated the design principles of minimal overlap and orthogonality (e.g., Huber & Zwerina, 1996). We used the Sawtooth CBC/Web 6.4.1 software package with the "complete enumeration" setting for these experiments. Each participant was randomly assigned to the choice sets. For validation purposes, we asked each respondent to perform the same choice task on two holdout choice sets. The selection of attributes and attribute levels (for details, see below) was based on prior market research studies, market observations and qualitative interviews with experts in the industry. The chosen attributes and attribute levels met realistic market conditions, including the central competitive environment of the considered firm for the railway and flight markets.

Existing business segment/existing product: Within the first choice-based conjoint experiment, different railway passenger transport offers for a journey of approx. 300 km (round trip) were presented. These choice sets included the following four attributes (see appendix): all of the attribute levels were chosen to mirror real market conditions as closely as possible, based on in-depth interviews with market experts from the considered firm:

- Brand: We considered three levels, namely the considered brand X, a hypothetical new brand named Speedtrans, and an important European competitor, A.
- Price: We employed three price levels: €38, €55, and €72.
- Value: This attribute describes the core product values that are essential for a travel offer. In this analysis, these attributes included the frequency of city connections per hour, the punctuality of the connection, the journey length, the kindness and competence of the companies' staff, and the comfort and facility attributes of the train. These value attributes were bundled and presented in terms of "low", "middle", and "high" levels (see appendix).

- Relationship management: Each stimulus contained customized services that were characterized by a discount card, a loyalty program, a heavy-user program, and a customer online account. The offer included the service discount card, which can be bought for a certain price and which grants a certain percentage reduction on the full fare for an entire year. Within their particular loyalty program, customers collect points for each trip taken by rail and can redeem points for special rewards. A heavy-user program provides the possibility that customers traveling at a minimum cost of x € each year can collect points and redeem these points for special services. An online customer account provides services to each customer, including access to various features, such as past travel information. These customer service attributes were bundled into "low", "middle", and "high" levels.

New business segment/new product: The second choice-based conjoint experiment described passenger-flight offers with round-trip flight connections between selected European cities. All of the attributes and attribute levels were chosen based on intensive market research and qualitative interviews with firm experts that led to the following attributes and attribute levels (see appendix):

- Brand: The brand attribute was presented at four levels: the brand extension of the considered brand X, a hypothetical new brand named Speedair, and the two main airline competitors within the European market (competitors A and B).
- Price: The experimental design included three price levels: €99, €149, and €199.
- Value: As in the first choice-based conjoint experiment, the core product value of an offer was described by central attributes, such as the frequency of (flight) connections per hour, the punctuality of the connections, the journey length, the kindness and competence of the

companies' staff, and the comfort and facility attributes of the plane. Again, these attributes were bundled and presented at "low", "middle", and "high" levels.

- Relationship management: These values were described by the same attributes that were discussed above and were bundled into "low", "middle", and "high" levels.

We employed several steps of data cleaning. In the first step, we eliminated from the original 1,663 respondents (including 831 respondents for the flight subsample) those who completed the questionnaire in an unrealistic response time (less than 50% and more than 200% of a realistic response time) and who demonstrated (almost) no variance in their response behavior across variables. As a result, we obtained a remaining sample of 1,375 respondents and a subsample of 671 respondents who had answered questions regarding the passenger flight segment. Next, we examined the individually estimated conjoint measurement price premiums for the three assets and excluded (1) respondents with a unrealistically high price premium (compared with the market price, i.e., absolutely more than €100 for the railway segment or more than €300 for the flight segment), and (2) respondents who showed a positive relationship between price and utility (which would lead to an unrealistic estimated price premium). As a result, the final total sample size was 1,281 respondents, including 671 respondents for the subsample of the passenger flight segment. In addition, for cases involving negative price premiums, we tested whether these price premiums (for each respondent) were significantly different from 0; for cases in which the price premium was not significantly different from 0, we set the price premium equal to exactly €0.

Firm data

We combined the information about the individual past, present and future purchase behaviors of the customers with external and internal data from the considered firm. Based on past annual reports of the considered business cooperation and the competitive environment of the hypothetical new business segment, we identified, for both product segments, the income margin before tax and interest (calculated in % as $\frac{EBIT}{Total\ Revenues} \times 100$), the weighted average cost of capital d, and the deferred income tax margin T. In addition, we collected information about future developments in the passenger railway and flight segments through intensive market research. Conducting quality interviews with experts on business cooperation helped us to make the decisions and assumptions involved in our analysis. These assumptions included the validation of estimated future earnings for both segments, the probability of entering into the new business segment with the brand extension X and the corresponding marketing costs C.

Method and results

Based on equations 2-5, we estimated the sum of the discounted cash flows for each quadrant (*Q I-IV*) of the introduced matrix (*part I*, *figure 1*). This procedure included two key estimation steps: In the first step, we estimated $DCF_{i(S)}^{I-IV}$ as the sum of the discounted cash flows on a customer level i, using the knowledge that each customer belongs to a certain customer segment S. The discounted cash flow estimation was based on a planning horizon of six years (2010-2015) and a perpetual annuity (>2015). In a second step, for each relevant customer segment *S (*"heavy users" (*HU*), "normal users" (*NU*), and "potential users" (*PU*)), we estimated the mean of the sum of discounted cash flows across all customers i who belong to a certain sample segment S. The mean of the sum of the discounted cash flows per segment was

multiplied by the number of customers in the segment in the general population. We identified the number of customers per segment in the general population in Germany (POP_s) using market research studies provided by the cooperating business. Next, the sum of discounted cash flows for each customer segment was determined.

The first type of cash flow (Q I): The customer survey and the firm data served as a basis to estimate the sum of discounted cash flows for the existing customers i for the customer segments of "heavy users" (*HU*) and "normal users" (*NU*) (*S={HU;NU})* for the existing passenger transport segment.

Step 1, for a customer *i* belonging to the customer segment S:

(6) $$DCF_{i(S)}^{I} = \sum_{p}^{P}\left(\frac{EBIT_1 \times \widehat{K}_{i,p} \times (1 - T_p)}{(1 + d_1)^p}\right) + \frac{EBIT_1 \times \widehat{K}_{i,P+1} \times (1 - T_{P+1})}{(1 + d_1)^{P+1}} \times \frac{1}{d_1}$$

Step 2:

(7) $$DCF^{I} = mean_{HU}\left(DCF_{i(HU)}^{I}\right) \times POP_{HU} + mean_{NU}\left(DCF_{i(NU)}^{I}\right) \times POP_{NU}$$

The underlying parameters for the first step were estimated as follows: The earnings before interest and the tax margin, $EBIT_1 = 4.68\%$, was measured as the mean *EBIT* margin of the years 2006-2009, the discount rate $d_1 = 9.68\%$ equaled the mean weighted average cost of capital for the years 2006-2009, and the deferred income tax rate $T_p = T_{P+1} = 30.50\%$ was assumed to be constant and equal to the tax rate for the year 2009. To estimate the expected revenues ($\widehat{K}_{i,p}$) and ($\widehat{K}_{i,P+1}$) on an individual customer basis *i*, the number of future purchases, the corresponding purchase prices and weighting factors for age and gender were specified. Based on the reported purchase behavior (2007-2012) of each respondent, we used trend extrapolation to estimate the future number of purchases of travel offers for the years 2013-2015. For the perpetual annuity, we assumed that the number of purchases was equal to the number of

purchases for the year 2015. Analyzing the reported past purchase behavior in terms of the preferred travel distance, train type, and class type, we approximated potential future customer choices. The future usage of a discount card for each respondent was forecasted based on the principle of economically rational customer behavior. As a result, the price per travel unit was determined. In a similar manner as for determining the number of purchases for the perpetual annuity, we assumed that the price per travel unit for the perpetual annuity was equal to the value of this price in the last predicted year (2015). To meet the actual distribution of age and gender of the German customers, we calculated weighting factors for each customer segment. The second step followed the aggregation procedure, as described in equation 7.

The second type of cash flow (Q II): The sum of discounted cash flows for the customers *i* who belong to the potential new customer segment *S={PU}* of the passenger transport segment was estimated as follows:

Step 1, for a customer *i* belonging to customer segment S:

(8) $$DCF_{i(S)}^{II} = \sum_{p}^{P} \left(\frac{EBIT_1 \times \widehat{K}_{i,p} \times ra_{i,p} \times (1 - T_p)}{(1 + d_1)^p} \right) + \frac{EBIT_1 \times \widehat{K}_{i,P+1} \times ra_{i,P+1} \times (1 - T_{P+1})}{(1 + d_1)^{P+1}} \times \frac{1}{d_1}$$

Step 2:

(9) $$DCF^{II} = mean_{PU}(DCF_{i(PU)}^{II}) \times POP_{PU}$$

Similarly to *Q I*, we used the parameters $EBIT_1 = 4.68\%$, $d_1 = 9.68\%$ and $T_p = T_{P+1} = 30.50\%$. The expected revenues ($\widehat{K}_{i,p}$,$\widehat{K}_{i,P+1}$) were based on survey responses regarding future purchases. The price per travel unit was approximated based on an average purchase price for the considered travel offer because no precise information regarding past preferences was available (e.g., train type, train class). As described for *Q I*, the future usage of a discount card was forecasted based on an assumption of economically rational behavior from each respondent. As

described above, we used weighting factors for age and gender. Furthermore, the segment of potential customers was combined with an acquisition probability ($ra_{i,p}$, $ra_{i,P+1}$). Because most of the potential customers had purchased a travel offer from the considered firm previously (but not within the past 12 months), there was the special case that the majority of potential customers had experience with travel offers from the considered firm. Therefore, we based the acquisition risk on two widely used parameters (building scorecards, e.g., Gönül & Hofstede, 2006): First, we built a score (between 0 and 1) that considers the recency in terms of the time between the last purchase and the date of the interview. Next, this recency is combined with the frequency of the (past inquired and future predicted) number of purchases. The combination of both parameters is considered to be an expression of (1 - cash inflow risk). The higher the score, the greater the probability that cash flows are secure. For reasons of simplicity, we assumed the score to be constant over time (a more sophisticated approach could use diffusion modeling to forecast the future behavior of potential customers for each year, e.g., Gupta et al., 2004; Schulze et al., 2012). Thus, the sum of the discounted cash flows of each potential customer was reduced by a risk coefficient before the sum of the discounted cash flows was projected to estimate the potential customer-base in the general population.

The third type of cash flow (Q III): The sum of the discounted cash flows for customers *i* who belong to the customer segments *S={HU;NU}* and the hypothetical new business segment (passenger flight segment) was estimated as demonstrated in equations 10 and 11:

Step 1, for all customers *i* belonging to segment *S*:

(10) $$DCF_{i(S)}^{III} = \sum_{p}^{P} \left(\frac{EBIT_2 \times \widehat{K}_{i,p} \times (1 - T_p)}{(1 + d_2)^p} \right) + \frac{EBIT_2 \times \widehat{K}_{i,P+1} \times (1 - T_{P+1})}{(1 + d_2)^{P+1}} \times \frac{1}{d_2}$$

Step 2:

(11) $$DCF^{III} = \left(mean_{HU}\left(DCF^{III}_{i(HU)}\right) \times POP_{HU} + mean_{NU}(DCF^{III}_{i(NU)}) \times POP_{NU}\right) \times r_m - C$$

The earnings before interest and the tax margin, $EBIT_2$, must be adapted to the hypothetical new business segment. We identified two key competitors, A and B, and analyzed their annual reports from the years 2006-2009. $EBIT_2 = 2.93\%$ is the mean *EBIT* of the two key competitive airlines, and $d_2 = 8.90\%$ is the mean weighted average cost of capital of both competitive airlines; each of these quantities applies across the years 2006-2009. The deferred income tax rate $T_p = T_{P+1} = 30.50\%$ was measured as presented in *Q I*. To estimate the expected revenues ($\widehat{K}_{i,p}$, $\widehat{K}_{i,P+1}$) on an individual customer basis, only those respondents who stated that they intended to travel by plane in 2010 with the considered firm (brand extension X) were included. These respondents answered several questions regarding the potential new business segment. Based on the stated purchases for 2010 to 2012, we used a trend extrapolation to estimate the future number of flights for the years 2013-2015. Beyond the planning horizon, we assumed that the number of purchases was equal to the number of purchases in the year 2015. Based on extensive market research, we determined the average price per flight for the considered round trip type. To meet the actual distribution of age and gender for the considered customers, we used weighting factors for each customer segment. Furthermore, based on in-depth interviews with managers of the considered firm, we specified the probability that the considered firm would enter the new market segment (i.e., the probability of the brand extension) as $r_m = 0.5$. In addition, the managers estimated the costs *C* for developing and introducing the hypothetical brand extension. We split these costs between *Q III* and *Q IV* in a manner proportional to the expected future discounted cash flows of both quadrants. It was assumed that

these costs are one-time costs that occur over a relatively short time period and therefore that these costs do not need to be discounted.

The fourth type of cash flow (Q IV): The last quadrant represents the sum of discounted cash flows for the customers *i* belonging to the potential new customer base *S={PU}* of the new business segment of the considered firm (in this context, these customers are described as "potential users" who have not travelled with the considered firm within the past 12 months).

Step 1, for all customers i:

(12) $$DCF_{i(S)}^{IV} = \sum_{p}^{P} \left(\frac{EBIT_2 \times \widehat{K}_{i,p} \times ra_{i,p} \times (1 - T_p)}{(1 + d_2)^p} \right) + \frac{EBIT_2 \times \widehat{K}_{i,P+1} \times ra_{i,P+1} \times (1 - T_{P+1})}{(1 + d_2)^{P+1}} \times \frac{1}{d_2}$$

Step 2:

(13) $$DCF^{IV} = \left(mean_{PU}\left(DCF_{i(PU)}^{IV} \right) \times POP_{PU} \right) \times r_m - C$$

The parameters $EBIT_2 = 2.93\%$, $d_2 = 8.90\%$, $T = 30.50\%$ and $r_m = 0.5$ were estimated as described for *Q III*. The segment of potential customers is associated with an acquisition probability ($ra_{i,p}, ra_{i,P+1}$) because it is not assured that each potential customer will become a customer of the brand extension X. As in the analysis for *Q III*, we generated a score that lies between 0 and 1. However, because the fourth type of cash flow refers to the potential new market segment, customers have no experience with this brand extension X. Thus, the acquisition probability for these customers was built solely on the future number of inquired (2010-2012) and predicted (2013-2015) purchases. Thus, the sum of the discounted cash flows of each potential customer was reduced by a risk coefficient before the cash flow for this customer segment was projected to picture the potential customer base in the general population. For the second step, the sum of discounted cash flows on an individual basis was aggregated as presented in equation 13. Again, we split the one-time costs *C* for developing and introducing the

hypothetical brand extension between *Q III* and *Q IV* in a manner proportional to the expected future discounted cash flows of both quadrants.

Table 1 gives an overview of the results of the estimation procedure for all four quadrants (*Q I-IV*):

[Insert Table 1]

Customer-related assets. Information from the choice-based conjoint experiments was used to allocate the sum of discounted cash flows (*table 2*) to the customer-related assets (please remember the underlying assumption that the sum of all discounted cash flows attributed to customer-related assets is assumed to be an approximation of the total firm value, e.g., Damodaran, 2006). Using a hierarchical Bayes routine (Arora & Huber, 2001), we computed the conjoint measurement preference structure of the respondents on an individual level. A total of 2,000 preliminary iterations and 1,000 draws per respondent were used to generate parameter estimates (12,000 iterations in total). Every 10^{th} iteration was saved, and each utility was determined by the mean utility across these 1,000 draws. Regarding predictive validity, we computed the aggregate choice shares of both holdout tasks and the predicted model, and we tested the extent to which the model was able to predict correctly the observed choice behavior within the holdout tasks (Huber et al., 1993). In terms of the mean absolute error and root mean square, the predicted shares were close to the actual shares and clearly outperformed the chance model. The choice replication rates of the two holdout tasks were 66.91% (railway segment) and 73.32% (flight segment). Overall, the considered goodness-of-fit measures suggest a reasonably good performance of the utility estimation.

For the estimation procedure of the customer-related assets, we build on the price premium approach (Ailawadi et al., 2003). This step is performed by measuring the ability of a

considered brand, value or customer relationship management level to charge a higher price in comparison to an otherwise equivalent offer with a corresponding minimum reference level of the considered attribute. Accordingly, the *brand assets* are based on the utility difference between the considered brand X and the hypothetical brand, the *value assets* are based on the comparison between the typical value level offered by the cooperation business and the minimum reference level, and the customer *relationship management assets* are based on the comparison between the typical level offered by the cooperation business and the minimum reference level. Each utility difference is divided by the utility per price (e.g., Sattler et al., 2010).

These individually measured price premiums were divided by the sum of all price premiums per respondent to generate a relative distribution (in %) of each customer-related asset. For each respondent, we allocated the sum of discounted cash flows according to the individual asset distribution (in %). Next, we projected the allocated discounted cash flows of each asset based on the (potential) customer base in the general population. This methodology provides a disjunctive asset measurement that is consistent with the provided customer-related asset definitions. *Table 2* provides an overview of the results.

[*Insert Table 2*]

Conclusions and implications

One of the greatest challenges today is to convincingly determine the contribution of marketing investments to firm value. Although it is widely accepted that marketing activities create financial value, researchers and practitioners struggle to monetize the contribution of marketing investments to firm value. The main challenge is to provide a comprehensive, *non-overlapping* and *measurable* framework of marketing-related assets. Such a framework is

particularly important to receive acceptance within financial accounting. Building on Srivastava et al. (1998) and Rust et al. (2004c), this research offers a customer-related assets (CRAs) framework that uses a customer-centric perspective to identify a comprehensive and mutually exclusive set of customer-related assets and that integrates these assets with financial accounting standards. We present the practical use of this framework for a major European firm. Against this background, our proposed framework provides a basis for including customer-related assets in internal and external financial reporting.

Customer-related assets are the center of our framework, and we include three types of customer-related assets that drive customers' decisions to buy products. We define *brand assets* as the benefits a firm offers its customers through the brand, *value assets* offer benefits through the quality of the core attributes of the product, and *relationship management assets* offer benefits through the use of customers' individual information to manage relationships with these customers. A significant contribution of the framework presented in this paper is that brands, value and customer relationships fulfill the criteria of identifiability, controllability and reliable measurability and, thus, can be referred to as assets from an accounting perspective. Furthermore, this paper presents a differentiated measurement approach of customer-related assets, as exemplified by the empirical study presented herein. As a result, an important basis for researchers and practitioners is provided to value assets in clear financial terms that are consistent with an accounting perspective.

With respect to the widely discussed concept of customer equity, we identified many differences between the concepts of customer equity and customer-related assets, especially brand equity. As we seek to understand the controversial relationship between these concepts, it is critical to understand that customers per se (if no contractual rights are provided) cannot be

considered to be an asset. According to IAS 38.11-13 and IAS 21, intangible assets should fulfill the criteria of identifiability, controllability, and reliable measurability. Identifiability assumes that the asset can be separated or divided by the firm and sold, transferred, licensed, rented or exchanged or that the asset results from contractual rights (IAS 38.11). However, the customer himself/herself does not belong to the firm as a resource and hence cannot be sold, transferred, licensed, rented or exchanged. Furthermore, the controllability of the customer by the firm is questionable. As the customer itself is not a resource of a firm and cannot be controlled by the firm, the expected future economic benefits from customers can hardly be restricted from others. Customers are free to purchase products from other firms. Thus, the customer himself/herself typically cannot be considered to be an asset, in contrast to brands. According to the introduced framework and measurement approach, brands can be considered to be resources of firms and can be sold or licensed. Hence, brands are identifiable, an assertion that has been accepted by accounting standards. In addition, brands are controllable in that a firm has the control of the brand-specific cash flows and has the power to obtain brand-specific cash flows and restrict the access of other firms to those cash flows. Marketers and accountants agree that brands belong to the group of intangible assets (e.g., Bahadir et al., 2008).

On the measurement level, there are further differences between the concepts of customer equity and brand equity. Customer equity is defined as the sum of the lifetime values of all customers, including existing and potential customers. More precisely, customer equity encompasses future revenues and costs that relate to acquisition, retention, and cross selling and are adjusted for the time value of money (e.g., Rust et al., 2004b). This definition clearly indicates that the mentioned concept refers to the financial value that flows from the customers to the firm and thus can be assigned to the cash flow matrix (*part I* of the framework).

Nevertheless, as our framework demonstrates, future discounted cash flow also results from potential new products (i.e., brand extension). As a result, brand equity and customer equity are overlapping concepts that can nevertheless be distinguished. Whereas the financial value of a brand potentially covers every quadrant of the presented cash flow matrix, the customer equity concept typically covers quadrants I and II, as only the outcomes of existing products are captured. As Ambler et al. (2002) already noted, "[t]his ability, to extend into new areas and to acquire new customers, is unique to the brand asset" (Ambler et al., 2002, p. 17). In addition, we intend to broaden this perspective, as the financial outcome of existing customers driven by a new product (i.e., brand extensions) is not captured by the concept of customer equity.

Our research has several limitations that represent avenues for future research. This paper provides a powerful framework that aids in the understanding of how customer-related assets can be measured in a comprehensive and non-overlapping way. Nevertheless, we did not investigate in detail how marketing and non-marketing activities influence each customer-related asset. For instance, the *value asset* is influenced by marketing (e.g., product management), human resources (e.g., skilled and motivated employees) and production expertise (e.g., patents). Future research should analyze in greater detail the impacts that marketing and non-marketing activities have on customer-related assets.

Furthermore, although our framework provides an important basis on which to develop accountability norms regarding customer-related assets, we provide no specific accounting standards detailing, for instance, how to put customer-related assets into balance-sheets. The formulation of such standards provides an avenue for future accounting research.

Finally, our proposed framework and the empirical application of this framework are based on several approximations and simplified procedures to provide an easy-to-administer

application. For instance, we relied on self-reported past purchase history instead of observed past purchase history. Furthermore, for quadrants II and IV, the future discounted cash flows were based on a constant acquisition probability as a proxy for future customer behavior. If applicable, more sophisticated methods might be useful. For instance, regarding the constant acquisition probabilities, a diffusion model might be more appropriate (e.g., Gupta et al., 2004; Schulze et al., 2012).

References

Aaker, David A. (1996), Measuring Brand Equity across Products and Markets, *California Management Review*, *38*(3), 102-120.

Ailawadi, Kusum L., Scott A. Neslin, and Donald R. Lehmann (2003), Revenue Premium as an Outcome Measure of Brand Equity, *Journal of Marketing*, *67*(4), 1-17.

Ambler, Tim, C. B. Bhattacharya, Julie Edell, Kevin L. Keller, Katherine N. Lemon, and Vikas Mittal (2002), Relating Brand and Customer Perspectives on Marketing Management, *Journal of Service Research*, *5*(1), 13-25.

Arndt, Johan (1967), Role of Product-Related Conversations in the Diffusion of a New Product, *Journal of Marketing Research (JMR)*, *4*(3), 291-295.

Arora, Neeraj, and Joel Huber (2001), Improving Parameter Estimates and Model Prediction by Aggregate Customization in Choice Experiments, *Journal of Consumer Research*, *28*(2), 273-283.

Bahadir, S. Cem, Sundar G. Bharadwaj and Rajendra K. Srivastava (2008), Financial Value of Brands in Mergers and Acquisitions: Is Value in the Eye of the Beholder?, *Journal of Marketing*, *72*(6), 49-64.

Bauer, Hans H., and Maik Hammerschmidt (2005), Customer-based Corporate Valuation, *Management Decision*, *43*(3), 331-348.

Bell, David, John Deighton, Werner J. Reinartz, Roland T. Rust, and Gordon Swartz (2002), Seven Barriers to Customer Equity Management, *Journal of Service Research*, *5*(1), 77-85.

Blattberg, Robert C., and John Deighton (1996), Managing Marketing by the Customer Equity Test, *Harvard Business Review*, *75*(4), 136-144.

Blattberg, Robert C., Gary Getz, and Jacquelyn S. Thomas (2001), *Customer Equity: Building and Managing Relationships as Valuable Asset,* Boston: Harvard Business School Press.

Damodaran, Aswath (2006), *Damodaran on Valuation: Security Analysis for Investment and Corporate Finance* (2. edition), Hoboken, NJ: Wiley-Blackwell.

Day, George, and Liam Fahey (1988), Valuing Market Strategies, *Journal of Marketing*, *52*(3), 45-57.

DIN ISO 10668 (2010)*, Brand valuation - Requirements for Monetary Brand Valuation.*

Farquhar, Peter H. (1989), Managing Brand Equity, *Marketing Research*, *1*(3), 24-33.

Financial Accounting Standards Board (2007), *Statement of Financial Accounting Standards No. 141: Business Combinations, NO. 299-A*, Norwalk, Connecticut, U.S.A.

Fishbein, Martin (1963), An investigation of the relationship between beliefs about an object and the attitude toward that object, *Human Relations*, *16*(3), 233-239.

Gleaves, Robin, Jamie Burton, Jan Kitshoff, Ken Bates, and Mark Whittington (2008), Accounting is from Mars, marketing is from Venus: establishing common ground for the concept of customer profitability, *Journal of Marketing Management*, *24*(7/8), 825-845.

Gönül, Füsun F. and Frenkel Ter Hofstede (2006), How to Compute Optimal Catalog Mailing Decisions, *Marketing Science, 25(1)*, 65-74.

Gray, Sidney J. (1988), Towards a Theory of Cultural on the Development of Accounting Influence Systems Internationally, *Abacus*, *24*(1), 1-15.

Gruca, Thomas S., and Lopo L. Rego (2005), Customer Satisfaction, Cash Flow, and Shareholder Value, *Journal of Marketing*, *69*(3), 115-130.

Gupta, Sunil, and Valarie Zeithaml (2006), Customer Metrics and Their Impact on Financial Performance, *Marketing Science*, *25*(6), 718-739.

Gupta, Sunil, Donald R. Lehmann, and Jennifer Ames Stuart (2004), Valuing Customers, *Journal of Marketing Research*, *41*(1), 7-18.

Hanssens, Dominique M., Roland T. Rust, and Rajendra K. Srivastava (2009), Marketing Strategy and Wall Street: Nailing Down Marketing's Impact, *Journal of Marketing*, *73*(6), 115-118.

Hennig-Thurau, Thorsten, Kevin P. Gwinner, and Dwayne D. Gremler (2002), Understanding Relationship Marketing Outcomes: An Integration of Relational Benefits and Relationship Quality, *Journal of Service Research*, *4*(3), 230.

Huber, Joel, and Klaus Zwerina (1996), The Importance of Utility Balance in Efficient Choice Designs, *Journal of Marketing Research*, *33*(3), 307-317.

Huber, Joel, Dick R. Wittink, John A. Fiedler, and Richard Miller (1993), The Effectiveness of Alternative Preference Elicitation Procedures in Predicting Choice, *Journal of Marketing Research*, *30*(1), 105-114.

Husmann, Sven, and Martin Schmidt (2008), The Discount Rate: A Note on IAS 36, *Accounting in Europe*, *5*(1), 49-62.

Hyman, Michael R., and Ike Mathur (2005), Retrospective and Prospective Views on the Marketing/Finance Interface, *Journal of the Academy of Marketing Science*, *33*(4), 390-400.

IAS 38, Intangible Assets, International Accounting Standards Board 2010.

International Accounting Standards Board (2007), *Customer-Related Intangible Assets (Agenda Paper 4).*

International Accounting Standards Board (2008), *Customer-Related Intangible Assets (Agenda Paper 11G.75).*

Kallapur, Sanjay, and Sabrina Kwan (2002), *The Value Relevance and Reliability of Brand Assets Recognized by UK Firms*: CIBER, Center for International Business Education and Research, Krannert Graduate School of Management, Purdue University.

Keller, Kevin Lane (1993), Conceptualizing, Measuring, Managing Customer-Based Brand Equity, *Journal of Marketing*, *57*(1), 1-22.

Keller, Kevin Lane (2007), *Strategic brand management: Building, measuring, and managing brand equality* (3rd ed.), Harlow, Upper Saddle River, N.J: Prentice Hall.

Keller, Kevin Lane, and Donald R. Lehmann (2006), Brands and Branding: Research Findings and Future Priorities, *Marketing Science*, *25*(6), 740-759.

Kumar, V., and Denish Shah (2009), Expanding the Role of Marketing: From Customer Equity to Market Capitalization, *Journal of Marketing*, *73*(6), 119-136.

Kumar, V., Katherine N. Lemon, and A. Parasuraman (2006), Managing Customers for Value: An Overview and Research Agenda, *Journal of Service Research*, *9*(2), 87–94.

Lehmann, Donald R. (2004), Metrics for Making Marketing Matter, *Journal of Marketing*, *68*(4), 73–75.

Lehmann, Donald R., and David J. Reibstein (2006), Marketing Metrics and Financial Performance. *Marketing Science Institute Monograph*, Cambridge, U.S.A.

Leone, Robert P., Vithala R. Rao, Kevin L. Keller, Anita M. Luo, Leigh McAlister, and Rajendra Srivastava (2006), Linking Brand Equity to Customer Equity, *Journal of Service Research*, *9*(2), 125–138.

Mahajan, Vijay, Vithala R. Rao, and Rajendra K. Srivastava (1994), An approach to assess the importance of brand equity in acquisition decisions, *Journal of Product Innovation Management*, *11*(3), 221–235.

Park, Chan S. U., and V. Srinivasan (1994), A Survey-Based Method for Measuring and Understanding Brand Equity and Its Extendibility, *Journal of Marketing Research*, *31*(2), 271–288.

Rappaport, Alfred (1986), *Creating shareholder value: The new standard for business performance*, New York: Free Press.

Reily, Robert F., and Robert P. Schweihs (1999), *Valuing intangible assets,* New York: McGraw-Hill.

Rust, Roland T., Tim Ambler, Gregory S. Carpenter, V. Kumar, and Rajendra K. Srivastava (2004a), Measuring Marketing Productivity: Current Knowledge and Future Directions, *Journal of Marketing*, *68*(4), 76–89.

Rust, Roland T., Katherine N. Lemon, and Valerie A. Zeithaml (2004b), Return on Marketing: Using Customer Equity to Focus Marketing Strategy, *Journal of Marketing*, *68*(1), 109–127.

Rust, Roland T., Valerie A. Zeithaml, and Katherine N. Lemon (2004c), Customer-Centered Brand Management, *Harvard Business Review*, *82*(9), 110–118.

Sattler, Henrik, Franziska Völckner, Claudia Riediger, and Christian M. Ringle (2010), The impact of brand extension success drivers on brand extension price premiums, *International Journal of Research in Marketing*, *27*(4), 319–328.

Schulze, Christian, Bernd Skiera, and Thorsten Wiesel (2012), Linking Customer and Financial Metrics to Shareholder Value: The Leverage Effect in Customer-Based Valuation, *Journal of Marketing*, *76*(2), 17–32.

Shah, Denish, Roland T. Rust, A. Parasuraman, Richard Staelin, and George S. Day (2006), The Path to Customer Centricity, *Journal of Service Research*, *9*(2), 113–124.

Simon, Carol J., and Mary W. Sullivan (1993), The Measurement and Determinants of Brand Equity: A Financial Approach, *Marketing Science*, *12*(1), 28–52.

Soufani, Khaled (2002), The decision to finance account receivables: the factoring option, *Managerial and Decision Economics*, *23*(1), 21-32.

Srinivasan, Shuba, and Dominique M. Hanssens (2009), Marketing and Firm Value: Metrics, Methods, Findings, and Future Directions, *Journal of Marketing Research*, *46*(3), 293–312.

Srinivasan, V. (1979), Network Models for estimating Brand-Specific Effects in Multi-Attribute Marketing Models, *Management Science*, *25*(1), 11–21.

Srivastava, Rajenda K., Tasadduq A. Shervani, and Liam Fahey (1998), Market-Based Assets and Shareholder Value: A Framework for Analysis, *Journal of Marketing*, *62*(1), 2–18.

Srivastava, Rajendra K., Tasadduq A. Shervani, and Liam Fahey (1999), Marketing, Business Processes, and Shareholder Value: An Organizationally Embedded View of Marketing Activities and the Discipline of Marketing, *Journal of Marketing*, *63*(4), 168–179.

Tellis, Gerald J., and Joseph Johnson (2007), The Value of Quality, *Marketing Science*, *26*(6), 758–773.

Verhoef, Peter C., and Peter S. H. Leeflang (2009), Understanding the Marketing Department's Influence Within the Firm, *Journal of Marketing*, *73*(2), 14–37.

Villanueva, Julian, Shijin Yoo, and Dominique M. Hanssens (2008), The Impact of Marketing-Induced Versus Word-of-Mouth Customer Acquisition on Customer Equity Growth, *Journal of Marketing Research*, *45*(1), 48–59.

Wiesel, Thorsten, Bernd Skiera, and Julin Villanueva (2008), Customer Equity: An Integral Part of Financial Reporting, *Journal of Marketing*, *72*(2), 1–14.

Zeithaml, Valarie A. (1988), Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence, *Journal of Marketing*, *52*(3), 2–22.

Zeithaml, Valarie A., Leonard L. Berry, and A. Parasuraman (1996), The Behavioral Consequences of Service Quality, *Journal of Marketing*, *60*(2), 31–46.

Figure 1: Framework – Customer-related assets (CRAs)

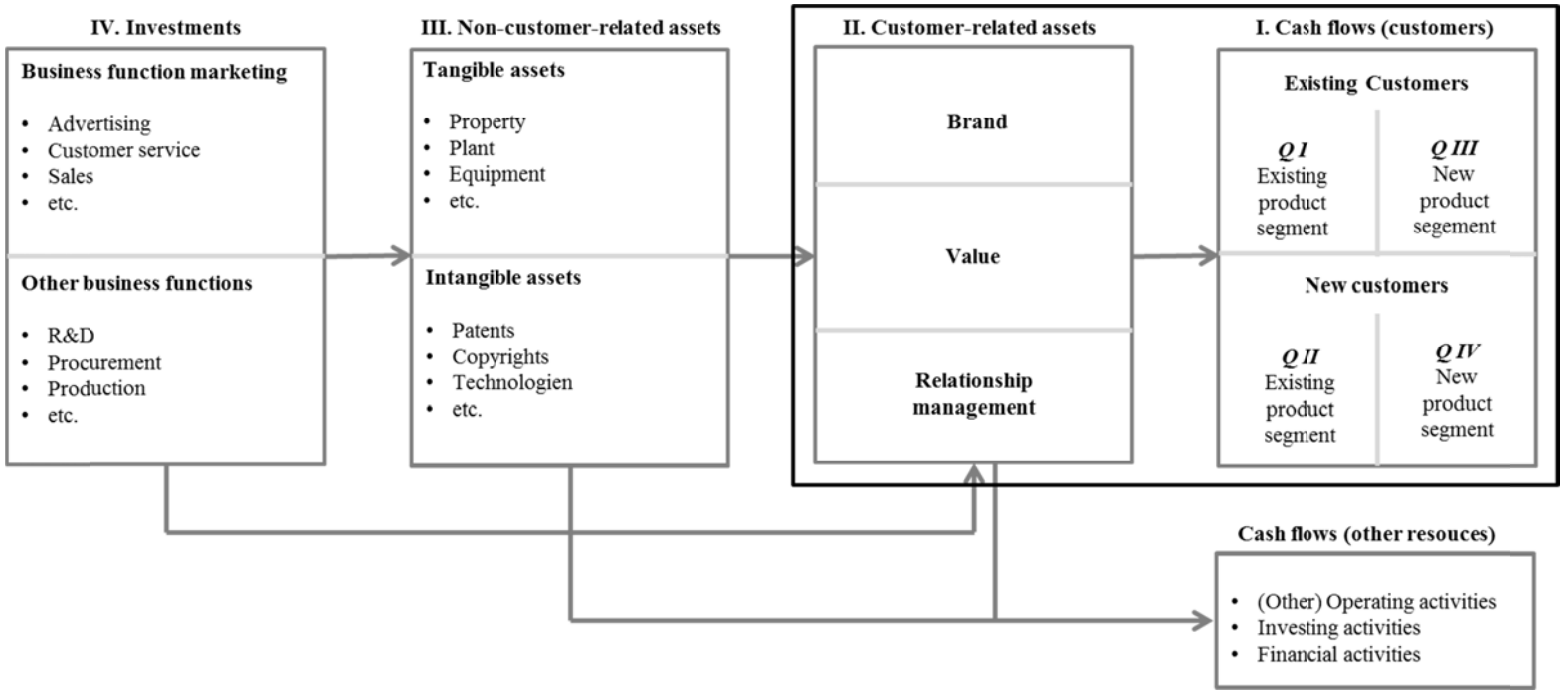


Figure 2: The data collection procedure

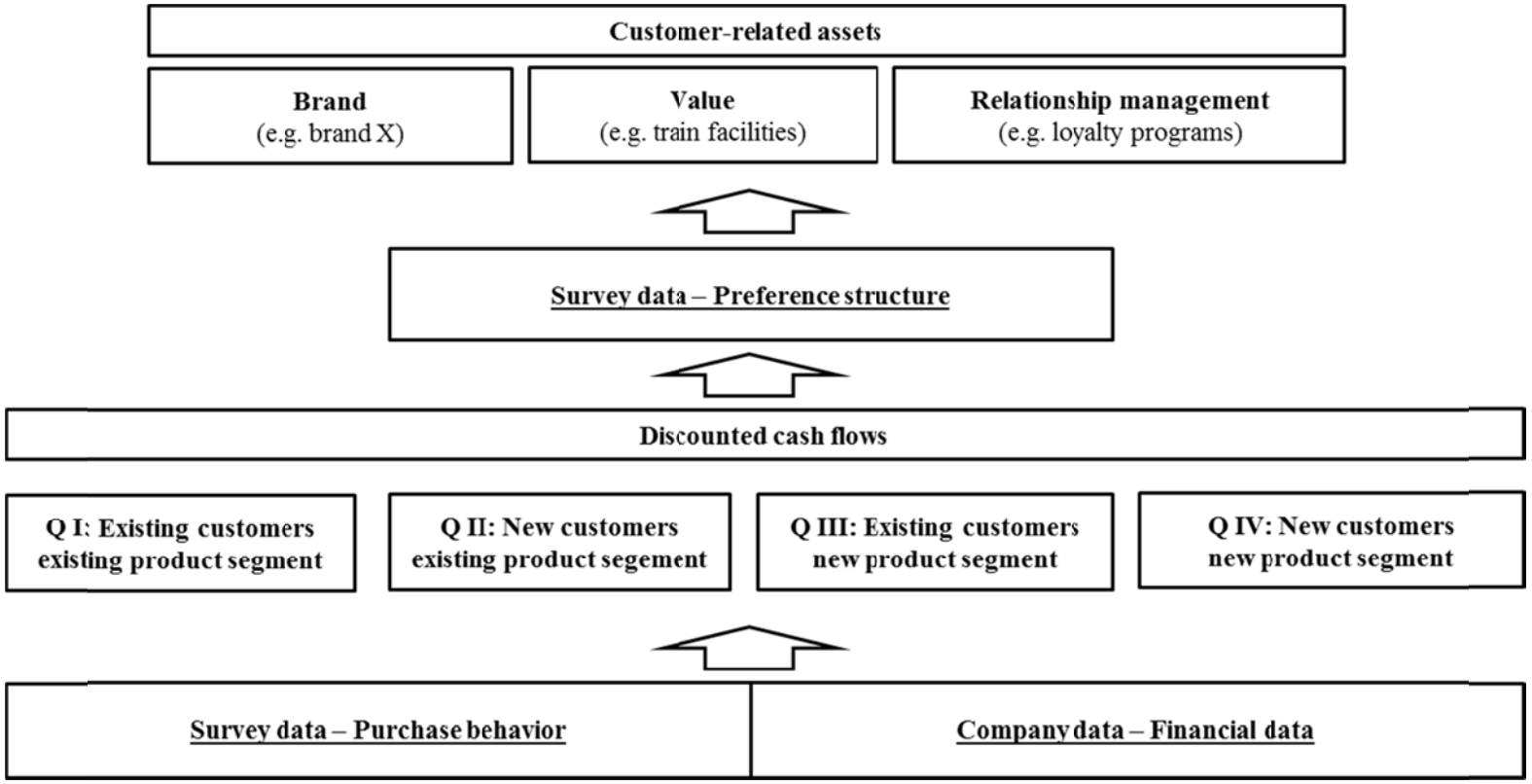


Table 1: Results *Q I-IV* – Discounted cash flows

Quadrants (*Q*)	**Discounted cash flows (in €)**
Q 1	2 580 499 145
Q 2	126 318 028
Q 3	285 119 298
Q 4	76 471 262
Total value	**3 068 407 733**

Table 2: Resulting assets – Discounted cash flows

Customer-related assets	**Asset value (in €)**	**Asset value (in %)**
Brand asset	451 427 511	14.71%
Value asset	2 153 037 511	70.17%
Customer relationship management asset	463 942 711	15.12%
Total asset value	**3 068 407 733**	**100%**

Appendix

Choice-based conjoint experiment: Railway passenger segment (experiment 1):

Attribute	Attribute levels
Brand	- Brand X - Hypothetical new brand Speedtrans - European competitor A
Price	- €38 - €55 - €72
Value	**Low** - Frequency of city connections: Every two hours - Punctuality: 65% of the train connections - Journey length: four hours - Kindness/Competence of the companies' staff: Low - Train facilities/Comfort: Low **Middle** - Frequency of city connections: Every hour - Punctuality: 80% of the train connections - Journey length: three hours - Kindness/Competence of the companies' staff: Medium - Train facilities/Comfort: Medium **High** - Frequency of city connections: Twice per hour - Punctuality: 95% of the train connections - Journey length: two hours - Kindness/Competence of the companies' staff: High - Train facilities/Comfort: High
Customer relationship	**Low** - No Discount card - No Loyalty program - No Heavy-user program - No online account **Middle** - Discount card: for €230 p.a., 50% price reduction on all 2nd class travel offers (or €57 p.a. for a price reduction of 25% or €3,600 p.a. for a price reduction of 100%) - Loyalty program: minimum sales of €500 within the last three years, e.g., voucher of €5 for the train bistro. - Heavy-user program: minimum sales of €2,000 p.a., e.g., access to a lounge area - Online account: Access to personal (past) travel information **High** - Discount card: for €99 p.a., 50% price reduction on all 2nd class travel offers (or €10 p.a. for a price reduction of 25% or €1,900 p.a. for a price reduction of 100%) - Loyalty program: minimum sales of €300 within the last three years, e.g., voucher of €5 for the train bistro - Heavy-user program: minimum sales of €1,200 p.a., e.g., access to a lounge area - Online account: Access to personal (past) travel information and train cancellations communicated by SMS

Appendix

Choice-based conjoint experiment: Flight passenger segment (experiment 2):

Attribute	Attribute levels
Brand	- Brand extension X - Hypothetical new brand Speedair - European competitor A - European competitor B
Price	- €99 - €149 - €199
Value	**Low** - Frequency of city connections: Two per day - Punctuality: 71% of the flight connections - Kindness/Competence of the companies' staff: Low - Plane facilities/Comfort: Low **Middle** - Frequency of city connections: Five per day - Punctuality: 83% of the train connections - Kindness/Competence of the companies' staff: Medium - Plane facilities/Comfort: Medium **High** - Frequency of city connections: Eight per day - Punctuality: 95% of the train connections - Kindness/Competence of the companies' staff: High - Plane facilities/Comfort: High
Customer relationship	**Low** - No Discount card - No Loyalty program - No Heavy-user program - No online account **Middle** - Discount card: for €99 p.a., 25% price reduction on all 2nd class/economy travel offers (or €390 p.a. for a price reduction of 50% or €7,800 p.a. for a price reduction of 100%) - Loyalty program: minimum 40 flights in Europe within the last three years, e.g., flight for free (Europe) - Heavy-user program: minimum sales of 24 flights in Europe p.a., e.g., access to a lounge area - Online account: Access to personal (past) travel information **High** - Discount card: for €49 p.a., 25% price reduction on all 2nd class travel offers/economy travel offers (or €199 p.a. for a price reduction of 50% or €3,900 p.a. for a price reduction of 100%) - Loyalty program: minimum 20 flights in Europe within the last three years, e.g., flight for free (Europe) - Heavy-user program: minimum sales of 12 flights in Europe p.a., e.g., access to a lounge area - Online account: Individual chat with customer service

III. Measuring success in place marketing and branding

Authors: **Sebastian Zenker**

Nicole Martin

Year: *2011*

Bibliographic Status: *Published,*

Journal of Place Branding and Public Diplomacy, Vol. 7 (1), 32-41.

Measuring success in place marketing and branding

Sebastian Zenker*

Nicole Martin**

*Sebastian Zenker,
Institute of Marketing and Media, University of Hamburg, Welckerstraße 8,
D-20354 Hamburg, zenker@placebrand.eu

**Nicole Martin,
Institute of Marketing and Media, University of Hamburg, Welckerstraße 8,
D-20354 Hamburg, nicole.martin@wiso.uni-hamburg.de

Abstract

As the competition between cities increases, cities focus more and more on establishing themselves as brands. Consequently, cities invest an extensive amount of taxpayers' money into their marketing activities. Unfortunately, cities still lack a proper success measurement, which has raised questions regarding the efficient and effective use of the taxpayers' money.
With this contribution we want to highlight some existing, but primarily new possibilities for a complex success measurement in place marketing, referring to the extant literature on place marketing and the general field of marketing. Therewith, we strive to translate different concepts like customer equity or customer satisfaction into the lexicon of place marketing, thus identifying empirical gaps for further research, as well as existing fruitful approaches.

Keywords: Place Marketing, Success Measurement, Citizen Equity, Citizen Satisfaction, Resident Migration Scale, Place Brand Equity

Paper type: Conceptual paper

Measuring success in place marketing and branding

1. Introduction

Cities increasingly compete with each other in an effort to attract tourists, investors, companies, new citizens, and most of all qualified workforce (Anholt, 2004; Kavaratzis, 2005; Zenker, 2009). Place marketers therefore focus more and more on establishing the city as a brand (Braun, 2008) and try to promote their city to its different target groups. As a result, cities invest a considerable amount of taxpayers' money in their marketing activities: Berlin, for example, maintains a marketing budget of 5 million Euros per annum (Jacobsen, 2009). Unfortunately, a proper success measurement in place marketing practice remains missing, thus raising questions regarding the efficient and effective use of the taxpayers' money (Jacobsen, 2009). Additionally, the current academic discussion demonstrates strong shortcomings in this respect, focusing mainly on the explorative description of a certain place (brand) without measuring the impact of place marketing and branding on different target groups. Hence, the aim of this paper is to show suitable concepts of success measurement from the common marketing field and to give place marketers some practical suggestions for measuring the impact of their work. We also want to identify gaps for further empirical research and develop a research agenda for place marketing theory.

2. Success in place marketing and branding

2.1 Defining place marketing and branding

Although examples for city promotion date back to 1850 (Ward, 1998), place marketing is a relatively new field of academic research (Kotler, Haider, & Rein, 1993; O'Leary & Iredal, 1976). The first publications dedicated to place marketing came from regional economists, geographers, and other social scientists (see for an overview: Braun, 2008), but were mostly limited to the promotional aspects of places. Ashworth and Voogd (1990) were

two of the first researchers to widen the scope by trying to develop a strategic planning framework for place marketing. From that point, place marketing was discussed in the broader context of structural change in cities (Berg & Braun, 1999). At the start of the new millennium, the focus in the debate on place marketing shifted in the direction of place branding (e.g. Kavaratzis, 2008). In recent years, the branding of places (and cities in particular) has gained popularity among city officials, illustrated by the development of city brand rankings such as the *Anholt-GMI City Brands Index* (Anholt, 2006) or the *Saffron European City Brand Barometer* (Hildreth, n.d.).

Although the number of contributions on place marketing has increased considerably over the last few years (Lucarelli & Berg, *in press*), researchers are still challenged to separate the different terminologies used in place marketing (see for an overview: Hanna & Rowley, 2008) and, in doing so, find a suitably broad definition. We settled on Braun (2008) as the most fitting provider, with his definition of place marketing (or in this case the synonym of 'city marketing') as "the coordinated use of marketing tools supported by a shared customer-oriented philosophy, for creating, communicating, delivering, and exchanging urban offerings that have value for the city's customers and the city's community at large" (p. 43). Furthermore, its aim is "to maximise the efficient social and economic functioning of the area concerned, in accordance with whatever wider goals have been established" (Ashworth & Voogd, 1990, p. 11). According to Kotler *et al.* (1993), an additional aim for place marketing is to "promote a place's values and image so that potential users are fully aware of its distinctive advantages" (p. 18) – even though practice tends to expose this as the only goal for place marketing activities. Nevertheless, two important aspects should be extracted from these definitions: first, place marketing should aim to increase not only economic stature, but also social functions, like place identification or the satisfaction with a place. Second, place marketing is a customer-orientated approach which should integrate all of a city's customers;

in other words, those activities to increase social function should benefit as many residents as possible instead of one favorable group.

As noted by Zenker and Braun (2010), a place brand is "a network of associations in the consumers' mind based on the visual, verbal, and behavioural expression of a place, which is embodied through the aims, communication, values, and the general culture of the place's stakeholders and the overall place design" (p. 3). According to the authors, the place brand is not the communicated expression or the 'place physics', but the perception of those expressions in the minds of the target audience(s). These perceptions lead to measurable brand effects such as a willingness to stay at a place (Zenker & Gollan, 2010) or resident satisfaction (Insch & Florek, 2008; Zenker, Petersen, & Aholt, 2009), as shown in Figure 1, and they therefore seem worthy of mental note when dealing with success measurement in place marketing. In summary, all of these definitions highlight the complexity of place marketing and branding, which only lends further challenge to the measurement of success.

Figure 1: *The concept of place brand perception*

<Insert Figure 1 about here>

2.2 Special characteristics of place marketing and branding

Place marketing and branding have to consider an assortment of special characteristics, such as the inherent variety of a place's customers. From a theoretical point of view, the main, broadly defined target groups in place marketing are: (1) visitors; (2) residents and workers; and (3) business and industry (Kotler *et al.*, 1993). However, as shown in Figure 2, the groups actually targeted in recent marketing practice are much more specific and diverse (Braun, 2008; Florida, 2004; Hankinson, 2005; Zenker, 2009).

Figure 2: *Different target groups for place marketing*

<Insert Figure 2 about here>

These target groups differ not only in regard to their structure, but also in their particular place needs and demands. Tourists, for example, are searching for leisure time activities like shopping malls or cultural offerings; investors are more interested in business topics; whereas the city's customers need a suitable environment for their purposes rather than simply a ‘dot on the map’. It is of great importance that a proper success measurement parallels these diverse demands, as those measurements must be related to every one of the multiple target groups.

Furthermore, places in themselves are complex products, since a place (offering) can encompass not simply a single location, but a package of locations – sometimes called mega-products (Florek, Insch, & Gnoth, 2006). The product for tourists in Berlin, for instance, overlaps to some extent with the product for the city’s residents. To use an illustrating analogy, a place, like a shopping mall, offers a large assortment of products with each customer filling his or her shopping bag individually. Consequently, it is nearly impossible to measure every incidental aspect of success.

As highlighted before by the definition from Ashworth and Voogd (1990), the general purpose of cities and their governance is not only to aid a ‘buying decision’ (e.g., in terms of visiting or not visiting a city; economic function), but also to fulfil the demands of their target groups, especially their citizens (social function). This concentration on the satisfaction of the ‘customers’ and not on profit for the ‘organization’ creates a crucial difference between place branding and the general field of marketing (where customer satisfaction is often just a necessary condition for future profit). Together, these arguments indicate that some measurement approaches are more suitable than others, and that a variety of success measures must be utilized in the field of place marketing and branding.

2.3 Success measurement in practice

There are indications that place marketing is shifting its focus, from adhering to the above-mentioned definition by Kotler (i.e., using place marketing primarily for promotion) to understanding itself as an integrated management tool (Ashworth & Kavaratzis, 2009). Unfortunately, this point of view is not yet common sense in place marketing practice (Grabow, Hollbach-Grömig, & Birk, 2006), which has led to limited approaches for success measurements. As an additional strain, the marketing budgets of cities are still very restricted (Jacobsen, 2009), creating a financially tense situation for place marketing agencies compared to the general marketing budgets of companies. In spite of very limited budgets, however, cities try to fulfil challenging aims (e.g., image changes) and eagerly pursue many different target groups and target sub-groups. Unfortunately, success measurement is not often performed on a regular basis: marketers mostly limit their data to key figures and indicators (like tourist overnight stays or press clippings) due to the high costs of more comprehensive methods. Thus, the question of efficiency and effectiveness of the place marketing activities remain unanswered, as well as the question of whether taxpayer money is being properly invested.

3. Concepts for measuring success in place marketing

From our point of view, the current measurement metrics typically offer very inadequate information (as said before, only encompassing items like tourist overnight stays or press clippings). Consequently, they also disregard our first and second above-mentioned special characteristics of places: the diverse target groups and the complexity of the product itself. Place marketers therefore need different concepts in order to capture the indicators that underlie a complex success measurement. One concept would likely be insufficient; however: a combination of distinct approaches could – together with the above-mentioned factors – give rich information about the efficiency and effectiveness of place marketing activities.

Thus, we regard the performance evaluation of place marketing activities as an appraisal problem, with multiple perspectives taken into account.

3.1 The perspective of customer-centricity

Citizens, visitors, workers, businesses and industry are central target groups in place marketing and can be considered relevant customers for a place or city. By focusing on fulfilling the customer's needs, the traditional marketing literature assumes the so-called 'customer-centered' perspective, which should be considered for the management of place marketing activities. The nature of customer-centricity "lies not in how to sell products but rather on creating value for the customer and, in the process, creating value for the firm; in other words, customer centricity is concerned with the process of dual value creation" (Shah, Rust, Staelin, & Day, 2006, p. 115). Referring to a relevant target group of a city (in this example, the citizens), this paper presents and discusses two customer-centered metrics below. It is important to mention beforehand that both approaches could also be used for other target groups, such as visitors.

Citizen Equity. In the general field of marketing, the customer-centered concept of customer equity has received increasing attention in recent years. Even though differing ways of calculating customer equity have been introduced, Rust, Lemon and Zeithaml (2004) have provided a well-established definition. They describe customer equity as the sum of lifetime values of all customers, including existing and potential customers. More precisely, customer equity encompasses future revenues and costs that relate to acquisition, retention, and cross selling, and are adjusted for the time value of money. Marketing literature provides various ways of estimating customer lifetime value (on aggregate level: e.g. Gupta, Lehmann, & Suart, 2004; Berger & Nasr, 1998; or on individual level: e.g. Kumar, 2007).

As several studies show, marketing literature links the issue of how to allocate marketing spending to customer equity or customer lifetime value (e. g. Kumar, Lemon, &

Parasuraman, 2006; Kumar, Ramani, & Bohling, 2004; Reinartz & Kumar, 2003). Based on these findings, we believe that the basic idea of customer equity can offer one perspective on the effectiveness and efficiency of place marketing spending. But how can we estimate the customer equity of citizens? To answer this question, we propose the so-called 'citizen equity' (on an aggregate basis), described as follows:

Citizen equity looks at a citizen's value to the place based on predicted future transactions and predicted future costs. Future transactions can be made operationally feasible in terms of customers' taxes: the tax revenues of present and potential customers form the central source of a place's income and become the basis for place actions (e.g., education, culture, administration). These revenues, minus the predicted costs associated with residency (e.g., citizen administration or social benefits), could be considered as average gross contribution. The marketing costs associated with motivating citizens to move to a place (or alternatively motivate them to stay) can be weighted against this contribution, as expenses for acquisition or retention activities, etc., are part of marketing-specific costs. By defining citizen equity as the sum of cumulative cash flows of all customers or customer segments over the entire time of residency, future research can address certain empirical gaps. For instance, marketing literature demonstrates the ongoing desire to enhance marketing activities specifically customized to the customer in order to improve the effectiveness and efficiency of marketing spending (e.g. Kumar, Venkatesan, & Reinartz, 2006). Likewise, place marketers need to focus on how data collection could be extended to provide the empirical basis for estimating customer equity on a more individual level. Further, place marketers ought to be aware that the citizen equity model might include marketing actions beyond acquisition and retention: the cultural goods offered in a place, for example, or the complexity of the 'product' itself should be considered. Additionally, revenues and costs have to be discounted relative to the present value of money, given that citizen equity considers the generated cash flows of each period. However, the discount rate could depend on a variety of risk factors, so it

remains unclear as to how the rate can be adequately formulated and adjusted. Finally, place marketers should use available information to determine the time of residency: the calculation of citizen equity depends on knowing how long certain citizen segments (e.g. students) stay in a place, but this remains unknown.

Citizen Satisfaction. In terms of customer-centricity, it is crucial to capture the value a place presents for the customer. In line with Ashworth and Voogd (1990), who say that the aim of place marketing is to maximise both the economic and social functioning of an area, we believe it is necessary that residents become satisfied with their place of living. Therefore we need additional concepts and new variables to measure the social function of an area in tandem with hard facts such as the revenue and cost perspective (citizen equity). This complex question requires a critical understanding that neither census data nor simple indicators could give satisfactory answers because they only show the actual behaviour of a target group (e.g., citizens migrating or staying at a place), rather than the underlying reasons for the action. For example, sometimes the actual decision to move to another city is strongly influenced by external factors, like the availability of a new job or the closeness to family and friends (Powdthavee, 2008), which are unrelated to the level of satisfaction with the place of living.

In the general field of marketing, the concept of customer satisfaction is widely covered by different customer indices like the American Customer Satisfaction Index *ACSI* (Fornell, Johnson, Anderson, Cha, & Bryant, 1996), the European Customer Satisfaction Index *ECSI* (Cassel & Ekloef, 2001), or the Swiss Index of Customer Satisfaction *SWICS* (Bruhn & Grund, 2000). This concept is frequently linked to related constructs like customer loyalty, commitment, trust, or identification (e.g. Bhattacharya & Sen, 2003; Fullerton, 2003; Garbarino & Johnson, 1999). In some models, commitment or identification results from satisfaction; in others, satisfaction sometimes just serves as an influence. Thus it seems

important to integrate more constructs into a proper performance measurement of citizen satisfaction.

First approaches have been made in this regard: Insch and Florek (2008; 2010) developed a model for place satisfaction from customer satisfaction approaches; and more recently, Insch (2010) devised a tool for "identifying gaps in residents' perceptions of the importance and their satisfaction with aspects of city life that drive and detract from their overall satisfaction" (p.164). Zenker, Petersen and Aholt (2009) tried to translate the satisfaction scales for cities into basic meta-factors that influence the satisfaction of the citizens with their Citizen Satisfaction Index (*CSI*). As another example, Azevedo (2009) used the constructs of attachment, self-esteem, and identity from social psychology to measure 'pride for a place'. Borrowing a construct from the commitment scale for organizations, Zenker and Gollan (2010) developed their Resident Migration Scale (*ReMis*) to measure the 'intention to leave a place of living' and the 'identification with a place'.

For future development, the Customer-Company Identification (*C-C identification*) concept from Bhattacharya and Sen (2003) could be fruitful for understanding when citizens reach the point of identifying with a city. Additionally, the definition of customer loyalty as "a relationship of some sort between an actor and another entity and that the actor displays behavioral or psychological allegiance to that entity in the presence of alternative entities" (Melnyk, Van Osselaer, & Bijmolt, 2009, p. 82) also fits for the citizen-city relationship, and therefore could also be a performance measurement. As a final suggestion, the construct of trust is often raised in conjunction with identification and satisfaction (Bhattacharya & Sen, 2003; Garbarino & Johnson, 1999) and should thus enter into a comprehensive success measurement. All of these approaches, resulting from strong theoretical backgrounds, present a first step in the future development of performance measurement. Although some of them could already be useful from a practitioner's point of view, an underlying model which

explains the relations between these constructs remains missing. Future research should place priority on merging these approaches into a comprehensive model.

3.2 The perspective of brand-centricity

Besides the perspective of customer-centricity, a brand perspective could be regarded as another important dimension of evaluation, since place marketers are keen on establishing the place as a brand. The traditional marketing literature, and especially the definition by Keller (1993), shapes the broad understanding of brand equity – a brand metric of high importance. Keller (1993, p. 8) asserts that "customer-based brand equity is defined as the differential effect of brand knowledge on consumer response to the marketing of the brand". Given this definition, diverse approaches for the measurement of brand equity can be deduced.

Brand Value Driver. The brand value driver affects consumers' response towards a brand and generates valuable information regarding the customers' brand knowledge structure, measured on a non-monetary base (e.g., Keller, 1993). Relevant drivers such as brand awareness (in terms of brand recall and recognition) and brand image (characterized as the favourability, strength and uniqueness of brand associations) offer an overview of the customers' knowledge structure and provide essential information for the brand management (Keller, 1993, p.3). In this regard, the identification and quantification of the brand value driver play an important role for the management of place brands, especially when analysing the changes of driver over time and identifying the interdependences of drivers.

By more closely examining place marketing practice, it can be observed that non-monetary place brand equity metrics (especially image analysis) are already common for success measurement. However, place marketing practice needs an improvement in its tracking systems in order to identify central brand value driver (e.g., place brand personality; Ashworth, 2010) for each target group and capture the complexity of a place.

Place Brand Equity. In order to manage place marketing activities, we need to analyze the influence of a brand (and its value drivers) on outcome variables of the customer-brand relationship (e.g., a citizen's willingness to sacrifice salary for preferred choice of place). From the point of view of place marketing literature, research has only begun to discuss the connection between a place (brand) and the different outcome variables of the customer-brand relationship.

Papadopoulos and Heslop (2002) presented the first evidence for the use of place brand equity from the investor perspective: they translated the idea of country brand equity for products (country-of-origin) to country brand equity for investors (foreign direct investments *FDI*). Jacobsen (2009) developed this idea further and formulated a conceptual framework of drivers for the Investor-based Place Brand Equity (*IPE*), then analysed the linkage between brand value drivers and the decision to invest in an area (*FDI* location preference).

Zenker, Eggers and Farsky (2009) presented another approach: they explored the use of different city (brand) image dimensions, in monetary terms, for the target group of talents. With the help of brand-anchored conjoint (*BAC*) analysis (Louviere & Johnson, 1990) and the Hybrid Individualized Two-Level Choice-Based Conjoint (*HIT-CBC*) method (Eggers & Sattler, 2009), the study measured the percentage of wage that talents were willing to sacrifice for their preferred choice of place. In this approach, the overall willingness to sacrifice (in terms of annual salary) could be employed as an indicator for place brand equity.

With present research work and place marketing practice as a background, further research needs to compose a clearer picture of how to put place brand equity into practice. Even though place brand equity is a future-directed performance indicator that gives important information on the efficiency and effectiveness of marketing spending, it is currently unused by place marketers and even seldom used by companies in general (e.g. PricewaterhouseCoopers, GfK, Sattler, & Markenverband, 2006). The main reason for the lack of usage is that marketing literature has yet to devise a standard for brand equity

measurement. As such, research is also required for the context of places. In order to estimate brand equity, relevant customer-brand outcome variables (including monetary perspective) for each target group need to be clarified. This becomes especially pivotal given that the estimation of the (monetary) brand equity assumes that certain information – such as future place brand-specific cash flows, costs and brand-specific risk factors – have to be estimated. Lastly, present research should also be applied to the context of co-branding and the spill-over effects associated with brand alliances between places.

4. Conclusion

In summary, the absence of a comprehensive performance measurement became obvious to us through literary review. In this respect, the marketing academe exhibits considerable shortcomings. Nevertheless, the first approaches – such as the scale for citizen satisfaction or the model of place brand equity – demonstrate great promise for future developments.

Even though a lot of questions remain unanswered on a theoretical level, place marketing practice could already begin adopting academe's latest developments for their performance measurement. Since urban decision makers should consider resident satisfaction as a top priority, it is crucial that this dimension be integrated into the success measurement for place marketing practice. Advancements in the field also make clear that the construct of citizen satisfaction cannot be used without other related concepts, like commitment or identification. While a complete model that combines these constructs remains missing, an initial partnering of those different scales could nonetheless bequeath rich information for the development of a place – assuming that marketers conduct this performance measurement on a regular basis. By utilizing an online survey, these measurements could be taken with a representative sample on a periodical basis, which would easily be manageable and cost-effective. Such data could then also aid place marketing theory in developing a better

understanding of the relations between these constructs, as well as in building a proper model of place satisfaction alongside other important factors like loyalty, trust or identification.

The approach of citizen equity could help in two ways: first, to identify valuable target segments and calculate the appropriate budget for targeting these customers; and second, to provide a performance measurement tool that can be monitored on a regular basis. On the other hand, the appropriateness of this approach for place marketing remains questionable, due to its concentration on profit. As already mentioned, the underlying objective of a city is not to accumulate profit, but to ensure the satisfaction of as many residents as possible. The intermix of different social stratums is a condition for a functional city, and thus a single concentration on target groups with a high citizen equity could be arguable. That said, cities do need to earn their money in the form of taxes in order to fulfill all of their duties. Therefore, an evaluation of citizen equity could be justified if the attention of governmental activities remains on all existing residents.

Furthermore, the approach of place brand equity could prove valuable for measuring the performance of place marketing activities in the real world. Because one major aim of place branding involves the improvement of the place's image, place brand equity could be a very suitable success measurement tool. In order to render this approach cost-effective, though, place marketing academe is called upon to adapt common marketing techniques and develop the described methods further in order to improve existing place brand equity measurements.

5. References

Ailawadi, K. L., Lehmann, D. R., and Neslin, S. A. (2003) Revenue premium as an outcome measure of brand equity. *Journal of Marketing*, 67(October): 1-17.

Anholt, S. (2004) Chapter 14: branding places and nations. In: R. Clifton and J. Simmons (eds.) *Brands and Branding.* New York, NY: Bloomberg Press, pp. 213-226.

Anholt, S. (2006) *Anholt city brand index – "how the world views its cities"* (2nd ed.). Bellevue: Global Market Insight, WA.

Ashworth, G. J. (2010) Personality associations as an instrument of place branding: possibilities and pitfalls. In: G. J. Ashworth and M. Kavaratzis (eds.) *Towards effective place brand management – branding European cities and regions*. Cheltenham: Edward Elgar Publishing, pp. 222-233.

Ashworth, G. J., and Kavaratzis, M. (2009) Beyond the logo: brand management for cities. *Journal of Brand Management* 16(8): 520-531.

Ashworth, G. J., and Voogd, H. (1990) *Selling the city: marketing approaches in public sector urban planning*. London: Belhaven.

Azevedo, A. (2009) *Are you proud to live here? A residents oriented place marketing audit (attachment, self-esteem and identity).* Paper presented at the 38th European Marketing Academy Conference; 26–29 May, Nantes, France.

Bhattacharya, C. B., and Sen, S. (2003) Consumer-company identification: a framework for understanding consumers' relationships with companies. *Journal of Marketing* 67(April): 76-88.

Berg, L. van den, and Braun, E. (1999) Urban competitiveness, marketing and the need for organising capacity. *Urban Studies* 36: 987-999.

Berger, P. D., and Nasr, M. I. (1998) Customer lifetime value: marketing models and application. *Journal of Interactive Marketing* 12(1): 17-30.

Braun, E. (2008) *City marketing: towards an integrated approach*. Rotterdam: Erasmus Research Institute of Management (ERIM).

Bruhn, M., and Grund, M. A. (2000) Theory, development and implementation of national customer satisfaction indices: the Swiss index of customer satisfaction (SWICS). *Total Quality Management* 11(7): 1017-1028.

Cassel, C., and Ekloef, J. A. (2001) Modelling customer satisfaction and loyalty on aggregate levels: experience from the ECSI pilot study. *Total Quality Management* 12(7): 834-841.

Eggers, F., and Sattler, H. (2009) Hybrid individualized two-level choice-based conjoint (HIT-CBC): a new method for measuring preference structures with many attribute levels. *International Journal of Research in Marketing* 26(2): 108-118.

Florek, M., Insch, A., Gnoth, J. (2006) City council websites as a means of place brand identity communication. *Place Branding* 2(4): 276-296.

Florida, R. (2004) *The rise of the creative class*. New York: Basic Books.

Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J. S., and Bryant, B. E. (1996) The American customer satisfaction index: nature, purpose, and findings. *Journal of Marketing* 60(4): 7-18.

Fullerton, G. (2003) When does commitment lead to loyalty? *Journal of Service Research* 5(4): 333-344.

Garbarino, E., and Johnson, M. S. (1999) The different roles of satisfaction, trust, and commitment in customer relationships. *Journal of Marketing* 63(April): 70-87.

Grabow, B., Hollbach-Grömig, B., and Birk, F. (2006) City marketing – current developments an overview. In: F. Birk, B. Grabow and B. Hollbach-Grömig (eds.) *Stadtmarketing - Status quo und Perspektiven* [*city marketing – status quo and future priorities*]. Berlin: Deutsches Institut für Urbanistik, pp. 19-34.

Gupta, S., Lehmann, D. R., and Suart, J. A. (2004) Valuing customers. *Journal of Marketing Research* XLI: 7-18.

Hankinson, G. (2005) Destination brand images: a business tourism perspective. *Journal of Services Marketing* 19(1): 24-32.

Hanna, S., and Rowley, J. (2008) An analysis of terminology use in place branding. *Place Branding and Public Diplomacy* 4(1): 61-75.

Hildreth, J. (n.d.) *The Saffron European city brand barometer*. Saffron Brand Consultants website, http://saffron-consultants.com/news-views/publications/, accessed 21st December 2010.

Holbrook, M. B. (1992) Product quality, attributes, and brand names as determinants of price: the case of consumer electronics, *Marketing Letters* 3(1): 71-83.

Insch, A. (2010) Managing residents' satisfaction with city life: application of importance–satisfaction analysis. *Journal of Town and City Management* 1(2): 164-174.

Insch, A., and Florek, M. (2008) A great place to live, work and play – Conceptualizing place satisfaction in the case of a city's residents. *Journal of Place Management and Development* 1(2): 138-149.

Insch, A., and Florek, M. (2010) Place satisfaction of city residents: findings and implications for city branding. In: G. J. Ashworth and M. Kavaratzis (eds.) *Towards effective place brand management – branding European cities and regions*. Cheltenham: Edward Elgar Publishing, pp. 191-204.

Jacobsen, B. P. (2009) Investor-based place brand equity: a theoretical framework. *Journal of Place Management and Development* 2(1): 70-84.

Kavaratzis, M. (2005) Place branding: a review of trends and conceptual models. *The Marketing Review* 5: 329-342.

Kavaratzis, M. (2008) *From city marketing to city branding: an interdisciplinary analysis with reference to Amsterdam, Budapest and Athens*, PhD thesis, Groningen: Rijksuniversiteit Groningen.

Keller, K. L. (1993) Conceptualizing, measuring, and managing customer-based brand equity. *Journal of Marketing* 57(January): 1-22.

Kotler, P., Haider, D. H., and Rein, I. (1993) *Marketing places: attracting investment, industry, and tourism to cities, states, and nations*. New York: The Free Press.

Kumar, V. (2007) Customer lifetime value – the path to profitability. *Foundations and Trends® in Marketing* 2(1): 1-96.

Kumar, V., Lemon, K. N., and Parasuraman, A. (2006) Managing customers for value, an overview and research agenda. *Journal of Service Research* 9(2): 87-94.

Kumar, V., Ramani, G., and Bohling, T. (2004) Customer lifetime value approaches and best practices applications. *Journal of Interactive Marketing* 18(3): 60-72.

Kumar, V., Venkatesan, R., and Reinartz, W. (2006) Knowing what to sell, when, to whom. *Harvard Business Review* 84(3): 131-137.

Louviere, J. J., and Johnson, R. D. (1990) Reliability and validity of the brand-anchored conjoint approach to measuring retailer images. *Journal of Retailing* 66(4): 359-382.

Lucarelli, A., and Berg, P. O. (in press) City branding – a state of the art review of the research domain. *Journal of Place Management and Development* 4(1).

Melnyk, V., Van Osselaer, S. M. J., and Bijmolt, T. H. A. (2009) Are women more loyal customers than men? Gender differences in loyalty to firms and individual service providers. *Journal of Marketing* 73(July): 82-96.

O'Leary, R., and Iredale, I. (1976) The marketing concept: quo vadis? *European Journal of Marketing* 10(3): 146-157.

Papadopoulos, N., and Heslop, L. (2002) Country equity and country branding: Problems and prospects. *Journal of Brand Management* 9(4-5): 294-314.

Powdthavee, N. (2008) Putting a price tag on friends, relatives, and neighbours: using surveys of life satisfaction to value social relationships. *The Journal of Socio-Economics* 37: 1459-1480.

PricewaterhouseCoopers, GfK, Sattler, H., and Markenverband (2006) *Praxis von Markenbewertung und Markenmanagement in deutschen Unternehmen: Neue Befragung 2005 [Brand equity and brand management in German companies: New survey 2005]*, PricewaterhouseCoopers AG, Frankfurt am Main.

Reinartz, W., and Kumar, V. (2003) The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing* 67(1): 77-99.

Rust, R. T., Lemon, K. N., and Zeithaml, V. A. (2004) Return on marketing: using customer equity to focus marketing strategy. *Journal of Marketing* 68(1): 109-127.

Shah, D., Rust, R. T., Staelin, R., and Day, G. S. (2006) The path to customer centricity. *Journal of Service Research* 9(133): 113-124.

Ward, S. V. (1998) *Selling places: the marketing and promotion of towns and cities 1850-2000*. London and New York: Spon Press.

Zenker, S. (2009) Who's your target? The creative class as a target group for place branding. *Journal of Place Management and Development* 2(1): 23-32.

Zenker, S., and Braun, E. (2010) *Branding a city – a conceptual approach for place branding and place brand management*. Paper presented at the 39th European Marketing Academy Conference; 1–4 June, Copenhagen, Denmark.

Zenker, S., Eggers, F., and Farsky, M. (2009) *Putting a price tag on cities.* Paper presented at the 38th European Marketing Academy Conference; 26–29 May, Nantes, France.

Zenker, S., and Gollan, T. (2010) Development and implementation of the resident migration scale (ReMiS): measuring success in place marketing. In: E. H. Witte, and T. Gollan (eds.) *Sozialpsychologie und Ökonomie* [*social psychology and economy*]. Lengerich: Pabst Verlag, pp. 156-172.

Zenker, S., Petersen, S., and Aholt, A. (2009) Development and implementation of the citizen satisfaction index (CSI): four basic factors of citizens' satisfaction. *Research Papers on Marketing and Retailing* 39: 1-19.

Figure 1: The concept of place brand perception.

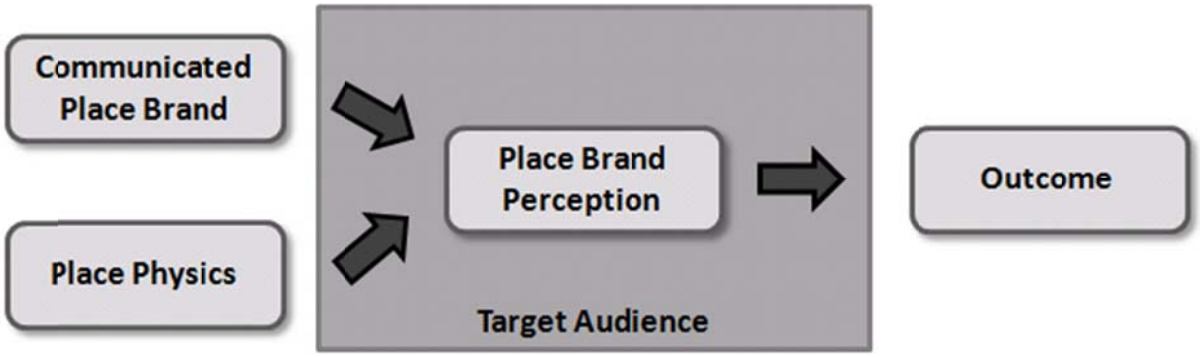
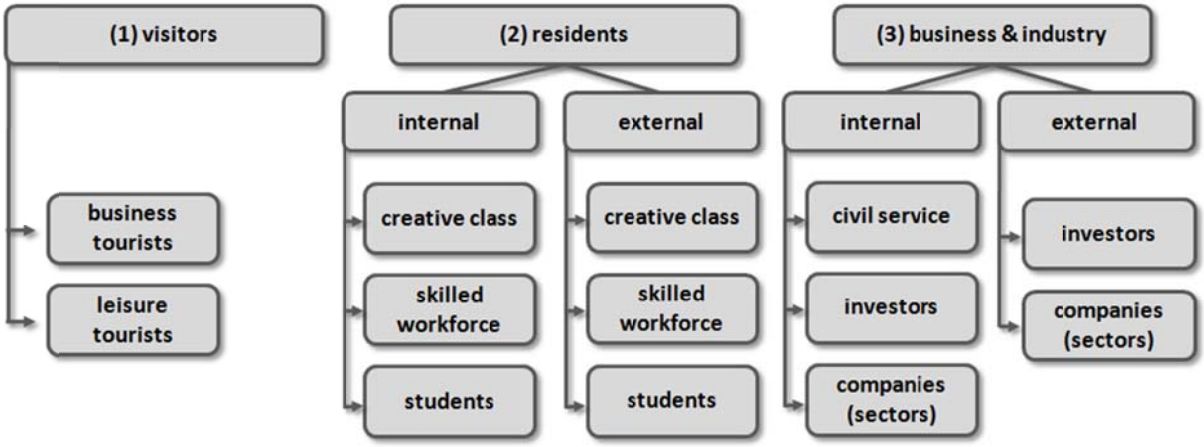


Figure 2: Different target groups for place marketing.



III. Assessing scorecard performances: A literature review and classification

Authors: **Nicole Martin**

Year: *20012*

Bibliographic Status: *Working paper,*

To be submitted soon to

Expert Systems with Applications

Assessing scorecard performance: A literature review and classification

Nicole Martin,
Institute of Marketing and Media, University of Hamburg, Welckerstrasse 8, D-20354 Hamburg, Germany, Tel: + 49 40 42838 8713, Fax: + 49 40 42838 4109,
e-mail: nicole.martin@wiso.uni-hamburg.de

Abstract

The assessment of scorecard performance in the field of credit scoring is of major relevance to firms. This study presents the first systematic academic literature review of how empirical benchmark studies assess scorecard performance in the field of credit scoring. By analysing 62 comparative studies, this study provides two main contributions. First, this study provides a systematic overview of the assessment-related decisions of all the reviewed studies based on a classification framework. Second, the assessment criteria of *consistency*, *application fit*, and *transparency* are introduced and used to discuss the observed assessment-related decisions. As the findings show, researchers often pay insufficient attention to ensuring the consistent assessment of scorecard performance. Moreover, the majority of the reviewed studies choose performance indicators that failed to fit the application context and provided non-transparent assessment documentation. In conclusion, these researchers pay a great deal of attention to the development of scorecards, but they often fail to implement a straightforward assessment procedure.

Keywords: Accuracy, performance indicator, decision support, classification, credit scoring, application scoring

Paper type: Literature review

Assessing scorecard performance: A literature review and classification

1. Introduction

For most financial institutions, credit lending is a principal business activity that also represents a great source of risk. Therefore, financial institutions are actively investigating alternative methods of improving their credit-lending decisions (e.g., Baesens et al., 2003b, p. 627), typically with the use of classification methods (e.g., Hand, 2005, p. 1109). Classification methods, also referred to as scorecards, estimate whether the credit applicants belong to a class of "good" or "bad" credit application risk, whether (existing) customers will accept specific credit offers or continuously use a certain credit product or whether customers are prevented from attrition to other lenders (e.g., Thomas et al., 2005). In general, these methods combine the available information on an applicant or existing customer to generate a numerical score that expresses how likely a customer is to belong to a certain class (e.g., Thomas et al., 2005, p. 1007).

While increasingly complex scorecards are developed (e.g., Yu et al., 2011, p. 15,392), the majority of studies on the subject have attempted to improve classification prediction in terms of the accuracy resulting from the comparison between the predicted scores and the actual paying behaviour of customers (Hand, 2005, p. 1109). In these studies, the researchers are interested in comparing different scorecards and aim to identify the "best" scorecard. The choice of the "best" scorecard is based on the superiority in prediction accuracy, as it is widely accepted that improvements in prediction accuracy might translate into significant future savings (Henley & Hand, 1997). The conclusion that classification predictions have improved depends on how accuracy is measured and implemented.

The related assessment actions and decisions are based on the given degree of information about the classification circumstances, which are characterised by their class distribution: that is, the ratio of "good risk" to "bad risk" applicants' and the costs associated with misclassifying "good risk" or "bad risk" applicants. Both parameters are typically asymmetrically distributed. Their level of information is not necessarily precisely known, and their estimation is linked with high levels of insecurity (e.g., Hand, 2001; West, 2000). On the basis of the assumed or available information, researchers determine the aimed-for performance perspective, which describes the underlying perspective for the assessment process (e.g., in the form of a cost- or error-minimisation perspective). Depending on the

chosen performance perspective, different (types of) performance indicators (Caruana & Niculescu-Mizil, 2004) and performance dimensionalities are available.

Contrary to common practice, a consideration of the characteristics of the application context is recommended; as it may be the case that one scorecard dominates another in some classification circumstances (e.g., cost and class ratio constellations) but is inferior in others (Drummond & Holte, 2006, p. 96). In this regard, comparative studies have often overlooked some characteristics of the context of the real application (e.g., Hand, 2006, p. 3). Accordingly, it has been argued that an imprecise handling of these application-specific challenges can easily lead to incorrect decision-support dimensions and, thus, a misleading scorecard choice (e.g., Drummond & Holte, 2006). Although isolated studies have discussed these relationships, the discussions have been conducted primarily on a theoretical basis (e.g., Adams & Hand, 1999; Hand, 2006). Therefore, the question arises regarding how published studies that benchmark scorecards on the basis of credit scoring data have practically addressed the assessment of the scorecards' performance. It is argued here that assessment-related decisions face several challenges. First, researchers should implement consistent assessment-related decisions such that all the possible combinations between the provided information on classification circumstances and the performance perspective, indicator choice or dimensionality are straightforward and consistent. Second, assessment-related decisions must fit the characteristics of the classification circumstances, as these characteristics have a great influence on the assessment results. Third, researchers should offer a minimum level of transparency to enable the reader to understand each assessment decision and evaluate the assessment's results.

Against this background, this study analyses how the published studies that benchmark scorecards on the basis of credit-scoring data have dealt with the assessment of scorecard performance on a practical level. Thus, this study intends to provide a useful, application-orientated reference for researchers and practitioners in the application of credit scoring decision making. This goal is realised by analysing the studies published on assessment-related decisions in academic journals between 2000 and 2011 on the basis of two approaches. First, a classification framework is developed that demonstrates how assumptions regarding classification circumstances relate theoretically to different assessment-related decisions. In addition, this classification framework is used to classify the reviewed studies. Second, the assessment criteria considering *consistency*, *application fit*, and *transparency* are used to discuss the observed assessment decisions. Accordingly, any misleading assessment decisions and research gaps can be identified.

The remainder of this article is structured as follows. Section 2 presents the methodological framework of this research. In section 3, chosen studies are classified and analysed according to the presented methodological framework. Problematic decisions are discussed, and research gaps are identified. In section 4, the results are summarised, and the limitations and implications are discussed.

2. Methodological framework

2.1 Research area and methodology

The research objective of this study is to identify a framework through which researchers and practitioners can obtain an overview of the assessment of scorecard performances in the area of application scoring. To reach this goal, published articles with a focus on classification analysis based on credit scoring data are considered. The literature search began with the descriptors "credit scoring" and "classification". Of these results, the articles aimed at benchmarking existing scorecards with existing or new scorecards based on application scoring data were selected. Only empirical studies and studies that clearly describe the use of scorecards in credit scoring applications were selected. The application field was further restricted to consumer-based data. Other credit-scoring application fields (e.g., behavioural scoring) and areas (e.g., credit rating) are not included due to the great variety of studies concerning credit scoring applications.

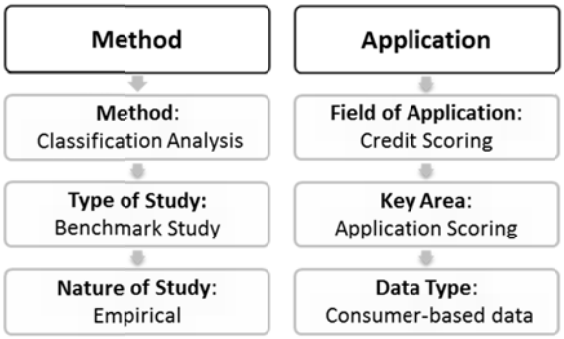


Fig. 1: Identification of research area

To compile consumer credit scoring applications, seven online academic databases – Business Source Premier and Complete, ACM, Emerals Fulltext, Ingenta Journals, Science Direct, Springer Link Journals, and IEEE Xplore/Electronic Library Online – and selected journals related to business intelligence, knowledge discovery, banking and financing between 2000 and 2011 were used as the research basis. Both academics and practitioners refer most frequently to journals; therefore, the author excludes conference papers, Master's

and doctoral dissertations, textbooks and unpublished working papers (analogous to Ngai et al., 2011, p. 560). Based on the great variety of published studies on the subject, these strict criteria led to the selection of 62 studies (for an overview please see appendix A). This should serve as a comprehensive basis that allows the study's central findings to be transferred to other classification problems (e.g., credit rating). Nevertheless, no requirement of completeness was imposed on the study selection.

2.2 Research classification scheme

In this study, the introduced classification scheme is based on the given degrees of information on the classification circumstances, also referred to as the study setting. Depending on the study setting, several assessment options are provided from which the researcher may choose (*fig. 2*). In this regard, it is noted that although the choice of a performance indicator frequently appears to be the central decision in the assessment of scorecard performance, the usefulness of the indicator choice is determined by the given degree of information on the classification circumstances, the chosen performance perspective and dimensionality. Therefore, this study illustrates these relationships as an interrelated causal chain.

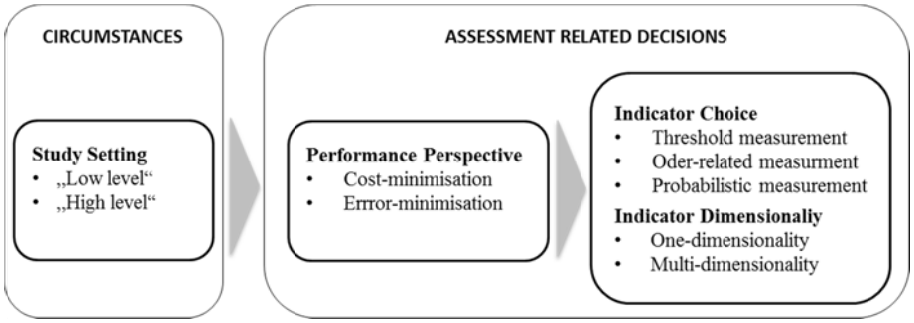


Fig. 2: Classification scheme - Causal decision chain

Study setting and performance perspective. The choice of a performance indicator depends on the information available on the classification circumstances, which are subsumed in the study setting. In the considered field of application, the study settings are characterised by two classes: "bad risk" and "good risk" credit applicants. Typically, the "bad risk" class is much smaller than the "good risk" class (e.g., Hand, 2001, p. 150), as the database included only data on applicants who had previously been granted credit. Further, the costs associated with granting credit to the rare class of "bad risk" applicants are higher than those associated with denying credit to "good risk" applicants (e.g., West, 2000, p. 1147). The degree of available or assumed information on these classification circumstances is differentiated into two study settings: "*low level*" and "*high level*".

"Low level" study setting*:* It should be noted that precise information on classification circumstances may be difficult to determine and is far from straightforward (e.g., Hand, 2001, p. 150). Therefore, the *"low level*" study setting is characterised by the vagueness of the available information. It may be understood that the misclassification of "bad risk" is associated with higher costs than those associated with the misclassification of "good risk" applicants. Additionally, it is expected that the class distribution between "good risk" and "bad risk" applicants will be unbalanced because only data from accepted credit applicants are available. Thus, the class distribution of the data basis is not necessarily identical to the data used when deploying the scorecards. This problem of sample selection bias is ubiquitous in the financial sector and is known as *reject inference* (e.g., Hand, 2006, p. 9). In this situation, the minimisation of the error of "bad risk" applicants being accepted, "good risk" applicants being rejected or both is important (e.g., Blöchlinger & Leippold, 2006, p. 852). Thus, the scorecards' discriminatory power is an important performance perspective. Nevertheless, adopting a cost-minimisation perspective is also conceivable. Researchers from other disciplines have introduced a cost perspective including the full range of possible class distributions and misclassification costs, as demonstrated by Cost Curves (Drummond & Holte, 2006). Thus, a vague level of available information could serve as a sufficient basis for a cost-minimisation performance perspective.

"High level" study setting: In this situation, information on some circumstances, including knowledge of cost ratios and the distribution between "good risk" and "bad risk" applicants in the population (i.e., prior probability), may be known or at least approximated. However, it should be noted that the available information on the classification circumstances underlies certain approximation insecurities: the expected loss depends on the exposure to default and the loss at a given default, where both terms are stochastic. Empirical studies demonstrate that the forecasting of these terms is a demanding task that can only be solved with great uncertainty (e.g., Loterman et al., 2011; Matuszyk et al., 2009). Forecasting opportunity costs may present an even greater challenge. Against this background, it is assumed that the costs associated with misclassification are not explicitly known; thus, only the likely ratios can be assumed (e.g., Adams & Hand, 1999; Hand, 2001). In addition, the approximation of class distributions is not a trivial task. An intuitive approach is to approximate the class distribution based on the available historical data. In application scoring, the sample distribution is biased because only previously accepted applicants are captured in the data basis (e.g., Hand, 2006, p. 9). Further, it is questionable how far across time periods these class distributions are stable. Customers' failure to repay their loans is frequently due to a lack of solvency caused, e.g., by sudden unemployment, which may

depend on changes in the economic environment. Thus, the distribution of the "good risk" and "bad risk" classes also depends on external factors. From a management perspective, one plausible assumption is to minimise the costs associated with classification to make cost-benefit decisions (e.g., Viaene & Dedene, 2005; Adams & Hand, 1999). Cost minimisation is one possible performance perspective. However, an error-minimisation perspective could also be adopted.

Performance perspective and indicator choice. Several indicators are available for the assessment of scorecard performance. These indicators can be roughly divided into three types (Caruana & Niculescu-Mizil, 2004) and linked to the performance perspective, as follows.

Threshold indicators depend on the choice of classification threshold. This is the value *t* such that objects are classified into the positive class when the predicted scores fall above *t* and into the negative class when their predicted scores fall below *t* (or vice versa). For a given set of applicants and a chosen threshold *t*, a two-by-two confusion matrix is given that presents the true or false classifications of the applicants (e.g., Fawcett, 2006, p. 862). Based on this matrix, various indicators can be calculated (please see appendix B).

		True Class: Positive	True Class: Negative
Predicted Class	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)
	Total	P = TP + FN	N = FP + FN

Fig. 3: Confusion Matrix

(TP = Correct classification of a positive object; FP = False classification of a negative object; FN = Correct classification of a negative object; FN = False classification of a positive object; P = Total number of positive objects in the sample; N = Total number of negative objects in the sample)

In the "*low level*" study setting, information on classification circumstances is either not available or not used in the scorecard evaluation. In this situation, an error-minimisation perspective can be chosen and operationalised using threshold indicators. Threshold indicators depend on the choice of a single classification threshold. This threshold can be set based on different decision rules (for threshold-setting rules, please see Banasik et al., 1996). Especially, if a "*high level*" study setting occurs, the cost-minimisation perspective will be relevant and expressed by a cost indicator (e.g., West, 2000, p. 1147):

$$Cost = C_{12}\pi_2 FPR + C_{21}\pi_1 FNR \tag{14}$$

This cost indicator assumes that the cost ratios that are connected with granting credit to a "bad risk" applicant (C_{12}) and denying credit to a "good risk" applicant (C_{21}) are known

(or can at least be reasonably approximated). In addition, the ratios of prior probabilities of "good risk" (π_1) and "bad risk" (π_2) in the application pool of credit scoring must be known. Moreover, the proportions of false positive (granting credit to "bad risk" applicants, *FPR*) and false negative (denying credit to "good risk" applicants, *FNR*) results are estimated. Alternatively, a cost minimisation threshold t can be chosen that categorises applicants into the "bad risk" or "good risk" class based on whether the score p for each applicant falls below or above a threshold t (e.g., Adams & Hand, 1999, p.1139).

$$p > t = \frac{C_{12}}{(C_{12} + C_{21})} \tag{15}$$

Based on this reasoning, these (and only these types of threshold indicators) can meet the requirements of a cost-minimisation perspective.

Ordering-related indicators use cases that have been ordered by their predicted scores and measure how well the scorecard ranks "negative risk" applicants above "positive risk" applicants (or vice versa). Thus, these measures do not depend on one single threshold but on the comparison of ordered cases. Instead, they represent a summary of the scorecard performance across all of the possible thresholds. In this study, the idea of ordering-related indicators (Caruana & Niculescu-Mizil, 2004) is extended by including additional indicators that conform to the fundamental principles underlying ordering-related indicators but do not summarise the analysed information in a single metric (e.g., Receiver Operating Characteristic curve (ROC)). Order-related indicators can be used to address an error-minimisation perspective because these indicators involve the independent determination of threshold settings and classification circumstances. It is also possible to connect the precise information of a given set of circumstances to order-related indicators for a cost-minimisation perspective. In fact, several different attempts have been made to utilise these types of indicators within the context of the cost-minimising perspective, e.g., through the creation of Iso-Performance lines in Receiver Operating Characteristic (ROC) spaces (Provost & Fawcett, 1997, Provost & Fawcett, 2001) or Cost Curves (Drummond & Holte, 2006).

Probabilistic indicators are entirely dependent on predicted scores and are not dependent on whether these scores fall above or below a threshold *t*. These indicators seek to ensure that the predicted scores for each individual case coincide with the true probability that the case in question represents a "bad risk" or "good risk" applicant. Based on this definition, these indicators are an expression of the discriminatory power. These indicators may be used to minimise errors but not to minimise costs.

Performance perspective and indicator dimensionality. All of the aforementioned considerations are fine in principle, but the operationalisation of the assessment perspective

and the choice of indicators are not straightforward tasks. Let us assume that a situation exists in which a large number of scorecards are benchmarked and an error-minimisation perspective is the primary objective. This scenario constitutes a typical study setting for research in the field of application scoring. One can choose a one-dimensional performance perspective that is expressed by a single scalar measurement that consolidates performance information into a single dimension. However, this approach might prove problematic for situations that cannot be readily represented by a single scalar value. For instance, error minimisation is frequently linked to a certain trade-off relationship between granting credit to a "bad risk" applicant and denying credit to a "good risk" applicant. Given this trade-off relationship, there are multiple indicators of importance in this hypothetical scenario. Thus, the modelling of this scenario requires a multi-dimensional performance perspective that must be expressed by several different indicators or by a graphical instrument that simultaneously visualises performance in two different performance dimensions (e.g., Drummond & Holte, 2006, pp. 96). Multi-dimensional performance perspectives can accurately evaluate trade-off situations. In general, however, the usage of graphical performance indicators might be challenging, given the large number of scorecards that are involved in many situations; thus, researchers and practitioners are likely to simply choose to use several indicators for their investigations.

2.3 The assessment-related criteria

A tremendous variety of possible combinations of assessment-related decisions can be made. Because this study seeks to analyse observed assessment-related decisions, this analysis focuses on the following three assessment criteria.

Consistency. Researchers must follow a consistent sequence of assessment-related decision steps. The choice of a performance perspective is dependent on the degree of information that is available with respect to the classification circumstances. Higher degrees of available and included information allow the chosen performance perspective to more closely approach an entrepreneur-relevant support dimension. Therefore, the assessment of scorecards should be operationalised by an indicator choice that reflects the available degree of information, the highest feasible chosen performance perspective and the multidimensional aspects of the performance in question.

Application Fit. The application fit is determined by the characteristics of the classification circumstances that are incorporated into the study setting. First, the chosen performance indicators should provide reliable decision support, irrespective of the skewness of the class distribution. It is widely accepted that the class distribution between "good risk"

and “bad risk” applicants is unbalanced and that the costs associated with granting credit to the rare class of “bad risk” applicants are higher than the costs of denying credit to “good risk” applicants (e.g., West, 2000, p. 1147). Second, the chosen performance indicator should consider the available degree of information. In this context, it is assumed that the costs associated with misclassification are not known or that likely but uncertain ratio estimates are provided. The latter issue is associated with estimation insecurities. Thus, the chosen performance indicator should account for both the available degree of information and potential parameter inaccuracies.

Transparency. A great variety of different performance indicators exist. Thus, it is necessary to provide transparent performance indicator documentation that includes the unambiguous definition, calculation and documentation of the parameter set that is required for the assessment of scorecard performances.

3. Research Analysis

3.1 Study settings and performance perspectives

This section provides an analysis of selected research studies. Assessment-related decisions are classified based on the classification scheme that was discussed in chapter 2. In *tab. 1*, an overview of all of the reviewed studies is provided (for further information, please see appendix A).

Accordingly, these studies choose the following performance perspectives based on the provided study settings:

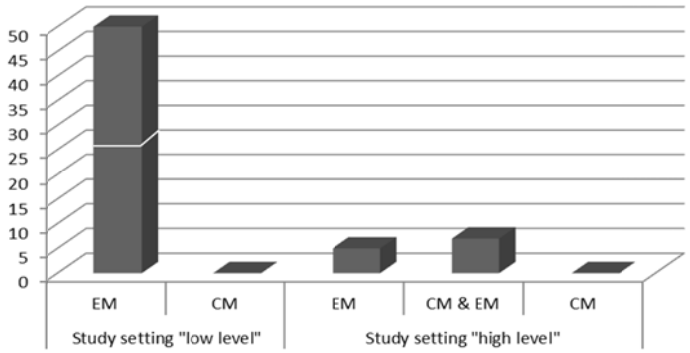


Fig. 4: Study setting and performance perspective
(Based on *tab. 1*)

Reference	Study Setting	Perspective	Performance Indicator
Abdou et al., 2008	HL	CM & EM	Cost indicator, HR/ER, ER-G, ER-B
Abdou, 2009a	HL	CM & EM	Cost indicator, HR/ER, HR-G, HR-B, ER-G, ER-B
Abdou, 2009b	HL	CM & EM	Cost indicator, HR/ER, HR-G, HR-B, ER-G, ER-B
Lee & Chen, 2005	HL	CM & EM	Cost indicator, HR/ER, HR-G, HR-B, ER-G, ER-B
West, 2000	HL	CM & EM	Cost indicator, HR/ER, ER-G, ER-B
Xiao et al., 2006	HL	CM & EM	Cost indicator, HR/ER, HR-G, HR-B
Tsai et al., 2009	HL	CM &EM	Cost indicator, HR/ER, HR-G, HR-B
Chuang & Lin, 2009	HL	EM	HR/ER, HR-G, HR-B, ER-G, ER-B
Hsieh & Hung, 2010	HL	EM	ER-Other (Gain-Chart)
Khashman, 2010	HL	EM	HR/ER
Martens et al., 2010	HL	EM	HR/ER
Yang, 2007	HL	EM	HR/ER, ER-G, ER-B
Antonakis & Sfakianakis, 2009	LL	EM	HR/ER, HR-G, HR-B, AUC, ROC, BRaA
Bellotti & Crook, 2009	LL	EM	AUC, ROC
Chen & Huang, 2003	LL	EM	HR/ER, HR-G, HR-B
Chen et al., 2009	LL	EM	HR/ER, HR-G, HR-B, ER-G, ER-B
Chen & Huang, 2011	LL	EM	HR/ER, ER-G, ER-B
Hoffmann et al., 2002	LL	EM	HR/ER
Holmes & Denison, 2003	LL	EM	HR/ER
Ince & Aktan, 2009	LL	EM	HR/ER, ER-G, ER-B
Lee et al., 2002	LL	EM	HR/ER, HR-G, HR-B, ER-G, ER-B
Lee et al., 2006	LL	EM	HR/ER, HR-G, HR-B, ER-G, ER-B
Li et al., 2006	LL	EM	HR/ER, HR-G, HR-B
Malhotra & Malhotra, 2002	LL	EM	HR/ER, HR-G, HR-B
Malhotra & Malhotra, 2003	LL	EM	HR/ER, HR-G, HR-B
Martens et al., 2007	LL	EM	HR/ER
Setiono et al., 2008	LL	EM	HR/ER
Šušteršič et al., 2009	LL	EM	HR/ER, ER-G, ER-B
Wang et al., 2005	LL	EM	HR/ER, HR-G, HR-B
Xu et al., 2009	LL	EM	HR/ER, ER-B
Yeh & Lien, 2009	LL	EM	HR/ER, AR, ER-Other (Gain-Chart)
Yobas et al., 2000	LL	EM	HR/ER
Yu et al., 2008a	LL	EM	HR/ER, HR-G, HR-B
Yu et al., 2008b	LL	EM	HR/ER, HR-G, HR-B
Yu et al., 2010	LL	EM	HR/ER, HR-G, HR-B
Zhang et al., 2009	LL	EM	HR/ER, HR-G, HR-B
Zuccaro, 2010	LL	EM	HR/ER, HR-G, HR-B
Baesens et al., 2003b	LL*	EM	HR/ER, HR-G, HR-B, AUC
Baesens et al., 2003a	LL*	EM	HR/ER
Chen & Li, 2010	LL*	EM	HR/ER, AUC, ROC
He et al., 2010	LL*	EM	HR/ER, HR-G, HR-B, KS
Hoffmann et al., 2007	LL*	EM	HR/ER
Hsieh, 2005	LL*	EM	HR/ER, ER-G, ER-B
Huang et al., 2006	LL*	EM	HR/ER
Huang et al., 2007	LL*	EM	HR/ER
Li et al., 2011	LL*	EM	HR/ER, HR-G, HR-B
Luo et al., 2009	LL*	EM	HR/ER
Nanni & Lumini, 2009	LL*	EM	HR/ER, ER-G, ER-B, AUC
Ong et al., 2005	LL*	EM	HR/ER
Peng et al., 2008	LL*	EM	HR/ER, HR-G, HR-B, KS, BRaA, ER-Other (CC, BRaR)
Ping & Yongheng, 2011	LL*	EM	HR/ER
Tsai & Wu, 2008	LL*	EM	HR/ER, HR-G, HR-B
Tsakonas & Dounias, 2007	LL*	EM	HR/ER
Wang & Huang, 2009	LL*	EM	HR/ER, AUC, BRaA, ER-Other (FM)
Yu et al., 2009b	LL*	EM	HR/ER, HR-G, HR-B, AUC
Yu et al., 2009a	LL*	EM	HR/ER, HR-G, HR-B
Yu et al., 2011	LL*	EM	HR/ER, HR-G, HR-B
Zhang et al., 2010	LL*	EM	HR/ER
Zhou et al., 2009b	LL*	EM	HR/ER, HR-G, HR-B
Zhou et al., 2009a	LL*	EM	AUC, ROC
Zhou et al., 2010	LL*	EM	HR/ER, HR-G, HR-B, AUC
Zhou et al., 2011	LL*	EM	HR/ER, HR-G, HR-B

Tab. 1: Study setting and performance perspective

(LL: "*Low level*" study setting; LL*: "*Low level*" study setting is extended, that is, for at least one dataset, cost, class ratios, or both are known or assumed; HL: "*High level*" study setting; EM: error-minimisation; CM: cost-minimisation; please appendix B)

As summarised in *tab. 1* and *fig. 4*, in 25 out of 62 reviewed studies, precise cost and class ratios are not provided for the datasets used. These studies choose the error-minimisation performance perspective. However, as comparative studies frequently use several datasets, it can be observed that in the other 25 studies, at least one dataset offers more precise information on the classification circumstances. Nevertheless, these studies assess their scorecards based on the error-minimisation perspective. In the remaining 12 studies, a higher degree of information is provided, or corresponding assumptions are made (e.g., cost ratios for German credit data are applied to Australian credit data; West, 2000, pp. 1147). Of these 12 studies, only 7 chose a cost-minimisation perspective. Interestingly, none of the studies focused on a pure cost-minimisation perspective; the cost-minimisation perspective is always combined with an additional overall error-minimisation perspective. In total, 55 studies chose the error-minimisation perspective, and 7 studies chose a combined-performance perspective.

3.2 Error minimisation, indicator choice and dimensionality

In the following section, *figs. 5* and *6* provide an overview of the performance indicator choice and dimensionality[1] for the error-minimisation performance perspective (for further information, please see appendices B and C):

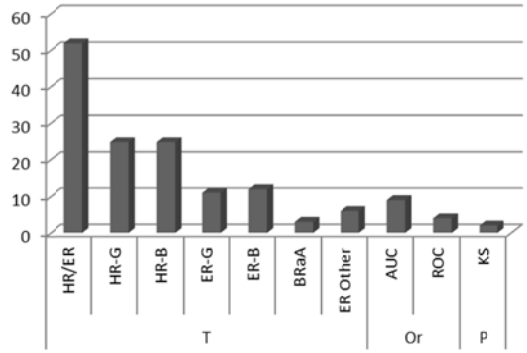


Fig. 5: Error minimisation - Choice indicator
(y-axis: number of studies, x-axis: performance indicators (T: threshold indicator, Or: order-related indicator, P: probabilistic indicator); please see appendix B)

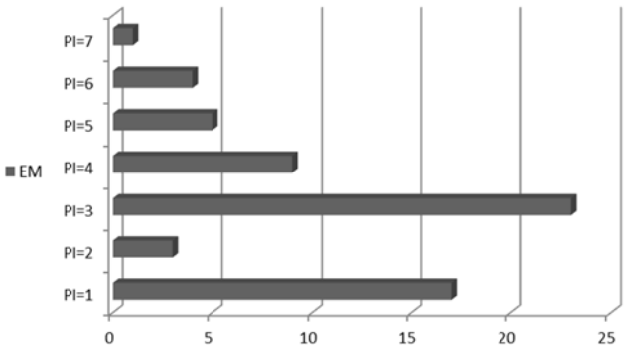


Fig. 6: Error minimisation - Dimensionality indicator
(y-axis: number of performance indicators (PI) used per study; x-axis: number of studies; please see appendix C)

Overview. As *fig. 5* demonstrates, a great variety of threshold indicators are used in the credit scoring literature. In total, 52 of 55 studies use the overall hit or error rate (HR/ER), which is calculated as the proportion of correct (incorrect) decisions to total decisions. The choice of the overall hit or error rate is followed by the choice of separately measured hit (HR) or error rates (ER) of "good risk" (G) and "bad risk" (B). Only a few studies have used order-related indicators. Among the studies that use order-related indicators, researchers have

[1] The number of performance indicators is the sum of reported performance indicators (text, tables or graphs) used in the comparison of scorecards.

mainly used the area under ROC curves (AUC) to compare scorecard performances, followed by the ROC curves themselves. The AUC metric is used as a substitution indicator for the overall hit or error rate (Bellotti & Crook, 2009; Zhou et al., 2009a) and as a complementary indicator (Antonakis & Sfakianakis, 2009; Baesens et al., 2003b; Chen & Li, 2010; Nanni & Lumini, 2009; Wang & Huang, 2009; Yu et al., 2009b; Zhou et al., 2010). Finally, probabilistic indicators are rarely used. The only identified probabilistic indicator is the Kolmogorov-Smirnov value (KS), which measures how far apart the distribution function of the scores of the "good risk" and "bad risk" groups are (He et al., 2010; Peng et al., 2008).

Consistency. Despite its theoretical unpopularity, 16 of 17 studies use a one-dimensional indicator choice in the form of the overall hit rate or error rate to operationalise the error-minimisation perspective.[2] This performance perspective is generally linked with the trade-off relationship between granting a "bad risk" applicant credit and denying a "good risk" applicant credit. However, it is widely accepted that applicants who may lead to substantial loss if accepted should be excluded. This is not achieved by the overall hit rate or error rate (e.g., Hand, 2001, p. 150) because this indicator is maximised by assigning everyone to the "good risk" class, as this class represents the majority of applicants captured in the data. Thus, one can hardly conclude that a scorecard is the "best" method based solely on the overall hit or error rate. The choice of this performance indicator does not properly reflect the need for multidimensional indicators. It is not a proper operationalisation of the performance perspective of error minimisation.

In contrast to the previously mentioned studies, the majority of the remaining studies account for the multidimensionality of the error perspective. The preferred combinations of performance indicators are based on three indicators, which are primarily linked to the combination of the overall hit or error rate and the hit rates of "good risk" and "bad risk", followed by the combination of the overall hit or error rate and the error rates for "good risk" and "bad risk" applicants. If researchers aim to assess scorecard predictions above these threshold indicators, the effort can additionally be realised by the AUC indicator, followed by the ROC curve. A total of 4 studies that use the AUC indicator without analysing the corresponding ROC curves have been identified. This method may be problematic because this indicator reflects an average performance perspective. It may be less likely that one scorecard will dominate the others for all the true-positive points across all the choices of false-positive rates. Therefore, this information is not captured by the AUC indicator, and it is advisable to analyse the AUC indicator in combination with the corresponding ROC curve.

[2] Only the study provided by Hsieh and Hung (2010) exhibits an exceptional character. This study only uses one performance indicator (Gain-Chart), which provides a two-dimensional performance perspective.

Furthermore, researchers tend to use a variety of performance indicators (between 4 and 7 performance indicators). The concept of "more is better" is not a useful guiding principle. Each indicator provides a different dimensionality of the scorecard performance, which could lead to several useful assessment dimensionalities. Unfortunately, it has been observed that researchers report several (types of) performance indicators but fail to interpret their findings (e.g., Peng et al., 2008; Wang & Huang, 2009). This procedure leads to a list of performance indicators, which frequently conclude that "the scorecard has competitive performance results", which provides no information about the usefulness of the scorecard under consideration in the application context of (application) credit scoring or its precise meaning in terms of error minimisation.

Application Fit. It has been shown that 52 of the 55 considered studies chose the overall hit or error rate. The choice of this performance indicator leads to the misleading assumption that no classification circumstances are included. Although no information may be available about the cost ratios and prior probabilities of the basic population, it is possible to obtain some types of information. As Drummond and Holte (2006) noted, there is a clear difference between two levels of information: the proportion of "bad risk" in the training and test datasets and the proportion of "bad risk" when the scorecard is deployed (put to use). The overall hit rate (overall error rate) assumes equally weighted class and cost ratios. This meets neither the actual nor the real distributions of the dataset(s) used. Therefore, the majority of the reviewed studies ignore the necessary application fit between the underlying study setting and the chosen performance indicator.

On the level of order-related indicators (i.e., ROC curve and AUC indicator), no conspicuous information regarding the application fit can be observed. The combination of ROC curves and the AUC indicator appears to be an especially plausible method of assessing scorecard performance, as these factors account for the multidimensionality of error-minimisation independent of the characteristics of the classification circumstances. On the level of probabilistic indicators, the choice of such indicators seldom appears plausible because performance indicators should support operational decisions (e.g., to whom to grant a loan; Hand, 2001, p. 151). As Hand (2009) noted, the motivation behind making this classification is to take an action. However, this is not provided by probabilistic indicators, as these indicators only express statistical dimensions.

Transparency. No conspicuous information regarding the choice and documentation of performance indicators is observed.

First, several indicators are differentially defined. For example, the indicator *precision* has been defined as the ratio of the misclassification rate of "bad risk" applicants to the total

of predicted "good risk" applicants (Wang & Huang, 2009). However, this does not correspond to conventional definitions of *precision*. For example, Fawcett, 2006 (p. 862) defines *precision* as the hit ratio of "positive risk" to total of predicted "good risk". It can be concluded that Wang and Huang (2009) (p. 5906) use the term *precision* but measure it as a *bad rate among accepts* (e.g., Hand, 2005, p. 1111). Moreover, the hit rate of "good risk", also referred to as *recall,* is measured as the error rate of misclassifying a "bad risk" instance in the "good risk" class (Wang & Huang, 2009). This calculation differs from the definition provided by Baesens et al. (2003b) (p.631). Baesens et al. (2003b) describes *recall (sensitivity)* as the hit rate of "good risk" among the positives. Therefore, it is quite misleading because indicator definitions are differentially used in the literature. Unfortunately, this is not an anomaly (error rates; Peng et al., 2008, p. 1022). It is particularly disconcerting that definitions of threshold indicators are often not documented clearly. Therefore, these results cannot be easily interpreted due to non-transparent indicator definitions.

Second, researchers frequently fail to explicitly report the threshold set (when it is necessary). This is quite surprising, as the determination of the threshold has a great influence on the indicator outcome. Only a few studies have documented decision rules to determine the classification threshold. As Wang et al. (2005) (p. 826) demonstrate, one method of doing so is to make the percentage of accepted applicants equal the percentage of "good risk" in the population. Because the percentage of "good risk" in the population is unknown, it is substituted with the percentage of "good risk" in the training dataset, and the corresponding threshold is chosen. Baesens et al. (2003b) (p. 634) describe the choice of a threshold based on marginal "good risk" and "bad risk" rates (for further information, please see Banasik et al., 1996, p.187). Alternatively, Tsai et al. (2009) (pp. 11687) choose a threshold that leads to equal hit rates of "good risk" and "bad risk" classification, allowing this threshold to equilibrate the predicting capability for the two types of classes ("fitted threshold"). Furthermore, it is observed that Šušteršič et al. (2009) (pp. 4741) present a sensitivity analysis of "good risk" and "bad risk" error rates for different thresholds to improve the best discriminatory power. Abdou (2009b) (p. 11416) presents a sensitivity analysis of the overall hit rate across different thresholds, and Antonakis and Sfakianakis (2009) choose different thresholds to compare the hit rates of "good risk" of accepted applicants (BRaA; please see appendix B). This makes it clear that no approach has been established to clearly set the threshold. For all these settings, the threshold decision depends on the subjective choice of the researcher. Moreover, methods that include prior probabilities in the form of class ratios (e.g., Wang et al., 2005) as decision support are linked with a particularly high estimation error.

This circumstance may lead to a biased threshold choice. Therefore, this choice of threshold indicators is not recommended.

3.3 Cost minimisation/error minimisation, indicator choice and dimensionality

Fig. 7 and *8* provide a closer look at the observed operationalisation of the cost- and error-minimisation perspectives and the dimensionality[3] of chosen performance indicators:

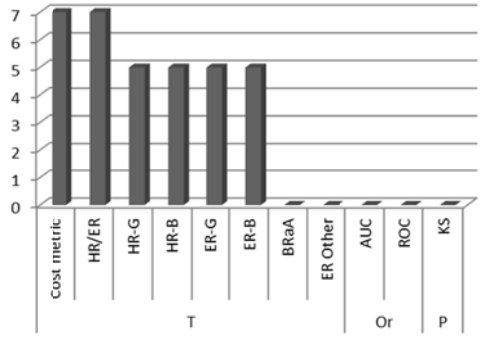


Fig. 7: Cost and error minimisation - Choice indicator
(y-axis: number of studies; x-axis: performance indicators)

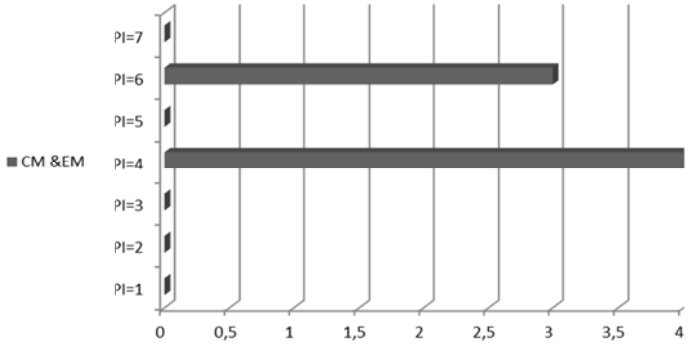


Fig. 8: Dimensionality indicator
(y-axis: number of performance indicators (PI); x-axis: number of studies)

Overview. For the performance perspective, the usage of threshold indicators seems to be widely accepted. All 7 studies calculate a cost indicator based on assumed cost and class ratios. This indicator is combined with the overall hit or error rate and, additionally, with separately measured hit rates of "good risk" or "bad risk" applicants, error rates of "good risk" or "bad risk" applicants, or both. Accordingly, these studies choose at least 4 performance indicators. Further, neither order-related performance indicators nor probabilistic indicators are included.

Consistency. As demonstrated in *figs. 7* and *8*, researchers combine the overall hit rate or error rate with the cost indicator. This is a surprising combination because these indicators reflect opposing assumptions. On the one hand, all the studies assume a cost ratio of 5:1 to calculate the cost indicator (please see *tab. 2*). These studies refer to the cost ratio originally provided by the German credit dataset, although the datasets used are not equal to the German credit dataset[4] (Abdou, 2009b; Abdou, 2009a; Abdou et al., 2008; Lee & Chen, 2005; Tsai et al., 2009). On the other hand, each study presents the overall hit or error rate, which assumes equally weighted cost and class ratios. In some ways, this assumption is contrary to the assumption of the cost indicator.

[3] The number of performance indicators is the sum of reported performance indicators (text, tables or graphs) used for the comparison of scorecards.
[4] Please see http://archive.ics.uci.edu/ml/datasets/Statlog+%28German+Credit+Data%29.

Reference	Data Source	PP	CD	CR
Abdou, 2009b	Egypt finance institute	Dataset	33% / 67%	5:1 7:1 10:1
Abdou, 2009a	Egypt finance institute	Dataset	26% / 74%	5:1
Abdou et al., 2008	Egypt finance institute	Dataset	26% / 74%	5:1
Lee & Chen, 2005	Taiwan finance institute	Dataset	10% / 90%	5:1
Tsai et al., 2009	Taiwan finance institute	Dataset	20% / 80%	5:1
West, 2000	Australian finance institute German finance institute	Market data + estimation error	56% / 44% 30% / 70%	5:1
Xiao et al., 2006	Australian finance institute German finance institute UK finance institute	Market data + estimation error	56% / 44% 30% / 70% 31% / 69%	5:1

Tab. 2: Cost minimisation perspective – data characteristics and cost ratios

(PP = Source for prior probability estimation; CD = Class distribution, "Good risk"/ "Bad risk" of dataset; CR = Cost ratio)

Furthermore, hit or error rates are presented and calculated (if necessary) on the basis of the threshold $\tau = 0.5$ (the exception is Tsai et al., 2009). Therefore, the loss perspective has not been operationalised by the cost-minimisation threshold. This is inherently contradictory; precise classification circumstances are assumed for the cost indicator, but the threshold setting is treated independently. Thus, it is obvious that the researchers have not systematically implemented a specific performance perspective.

Application fit. The choice of a cost indicator is problematic because this indicator assumes precise and constant information about the classification circumstances. This does not apply to the context of application scoring (please see section 2), as demonstrated in the following example. Real and precise market data are not documented or available in any of the reviewed studies. Accordingly, two different approximation procedures are observed. One method is to use the ratio of "good risk" and "bad risk" applicants in the empirical dataset as an approximation of the class distribution of the basic population (Abdou, 2009a; Abdou, 2009b; Abdou et al., 2008; Lee & Chen, 2005; Tsai et al., 2009[5]). The other possibility is demonstrated by West (2000) and Xiao et al. (2006). They estimate two scenarios based on the reported market default rates that are adjusted by the error rate of "bad risk" (West, 2000, p. 1147; Xiao et al., 2006, p. 430). As these examples show, the cost indicator should not be the preferred performance indicator. This applies to all the studies under consideration because the approximation of cost ratios and class ratios (prior probabilities) are related with a high level of uncertainty. Only Abdou (2009b) decided to use a sensitivity analysis (cost ratios from 5:1 to 10:1), as it is expected that higher cost ratios may be more appropriate to the environment of the Egyptian banking sector (Abdou, 2009b, p. 11406). Accordingly, the present operationalisation of the error-minimisation perspective does not meet the challenges

[5] Furthermore, the authors present a cost indicator based on prior probabilities that are set to be identical (Tsai et al., 2009, p. 11689).

of this application context. The need for a support decision indicator that can simplify the comparison of scorecards under imprecise and time-varying classification circumstances is obvious.

Transparency. Analogous to the pure error-minimisation perspective, studies focusing on a cost and error-minimisation perspective frequently refrain from explicitly documenting the threshold setting (if necessary).

4. Conclusions

This study carried out a literature review and classification of comparative studies in the field of application credit scoring. The literature review clearly notes how the selected studies assess scorecard performance in classification and prediction problems. More precisely, a classification framework is developed that demonstrates how assumptions of classification circumstances relate theoretically to different assessment decisions. Against this background, the selected studies are classified, and a variety of performance indicators is identified. Within a consideration of the introduced assessment criteria - *consistency*, *application fit* and *transparency* - this variety of assessment-related decisions is analysed critically. It is demonstrated that assessment-related decisions are not straightforward due to several factors.

Consistency. Researchers must follow a consistent chain of assessment-related decisions. The choice of performance perspective is dependent on the degree of available information on the classification circumstances. The higher the degree of available and included information is, the closer the chosen performance perspective can come to an entrepreneur-relevant support dimension. Thus, researchers should implement the highest feasible performance perspective. Nevertheless, this has not always been the case: for example, studies based on the "*high level*" study setting do not necessarily choose a cost-minimisation perspective. As, for example, Abdou et al. (2008) (p. 1287) stated, the cost perspective (i.e., an error indicator) is more subjective, while the error perspective (i.e., the overall error or hit rate) is more reliable. Researchers do not appear to exclusively trust the cost perspective based on the insecurities of parameter estimation. As a result, researchers cede additional information about the classification circumstances, leading to a less relevant decision-support dimension. Furthermore, it has been observed that the chosen performance perspective is frequently not explicitly implemented. It is unclear whether frequently used combinations of performance indicators provide managerial-relevant decision support for the choice of the "best" scorecard. Likewise, this point applies to the error- and cost-minimisation perspective. For instance, the combination of the overall hit or error rate and a cost indicator

is not straightforward because these indicators are inherently opposed. This combination may not lead to the desired decision support. It is recommended that researchers implement a specific performance perspective systematically.

Application fit. It is remarkable that frequently used performance indicators exhibit particular disadvantages due to assumptions and calculation restrictions. Threshold indicators depend on the user's subjective analysis in terms of the threshold setting. The overall hit or error rate underlies calculation restrictions that have been discussed critically (e.g., Adams & Hand, 1999; Hand, 2006). Classification circumstances in terms of cost and class ratios are linked with estimation uncertainties and errors; therefore, widely used performance indicators (i.e., cost indicators) cannot address these characteristics. Nevertheless, researchers appear to ignore the necessary application fit between the underlying characteristics of classification circumstances and the chosen performance indicators. Against this background, future research should address this challenge and verify the applicability of performance indicators from other fields (e.g., Cost Curves: Drummond & Holte, 2006; H-measure: Hand, 2009).

Transparency. It would be reasonable to guess that there are no uncertainties regarding the definitions and calculations of well-established and frequently used indicators. Surprisingly, this is not always the case for threshold performance indicators, which can be quite misleading. Consistency in the language and definitions utilised would greatly facilitate communication and research in this area.

In conclusion, the selected studies are characterised by a tremendous variety of scorecard types developed for decision support. It may be appropriate to conclude that researchers pay a great deal of attention to scorecard development. Nevertheless, it must be kept in mind that when scorecards are implemented in the credit-scoring application context, a single model's claims of superior performance may be misleading in the absence of a straightforward scorecard assessment.

References

Abdou, H. A. (2009a). An evaluation of alternative scoring models in private banking. *Journal of Risk Finance*, *10*(1), 38–53.

Abdou, H. A. (2009b). Genetic programming for credit scoring: The case of Egyptian public sector banks. *Expert Systems with Applications*, *36*(9), 11402–11417.

Abdou, H. A., Pointon, J., & El-Masry, A. (2008). Neural nets versus conventional techniques in credit scoring in Egyptian banking. *Expert Systems with Applications*, *35*(3), 1275–1292.

Adams, N. M., & Hand, D. J. (1999). Comparing classifiers when the misallocation costs are uncertain. *Pattern Recognition*, *32*(7), 1139–1147.

Antonakis, A. C., & Sfakianakis, M. E. (2009). Assessing naive Bayes as a method for screening credit applicants. *Journal of Applied Statistics*, *36*(5), 537–545.

Baesens, B., Setiono, R., Mues, C., & Vanthienen, J. (2003a). Using neural network rule extraction and decision tables for credit-risk evaluation. *Management Science*, *49*(3), 312–329.

Baesens, B., van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003b). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, *54*(6), 627–635.

Banasik, J., Crook, J. N., & Thomas, L. C. (1996). Does scoring a subpopulation make a difference? *International Review of Retail, Distribution and Consumer Research*, *6*(2), 180–195.

Bellotti, T., & Crook, J. (2009). Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications*, *36*(2), 3302–3308.

Blöchlinger, A., & Leippold, M. (2006). Economic benefit of powerful credit scoring. *Journal of Banking & Finance*, *30*(3), 851–873.

Caruana, R., & Niculescu-Mizil, A. (2004). Data mining in metric space: an empirical analysis of supervised learning performance criteria. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 69–78). Seattle, WA, USA: ACM.

Chen, F.-L., & Li, F.-C. (2010). Combination of feature selection approaches with SVM in credit scoring. *Expert Systems with Applications*, *37*(7), 4902–4909.

Chen, M.-C., & Huang, S.-H. (2003). Credit scoring and rejected instances reassigning through evolutionary computation techniques. *Expert Systems with Applications*, *24*(4), 433–441.

Chen, S., & Huang, M. (2011). Constructing credit auditing and control & management model with data mining technique. *Expert Systems with Applications*, *38*(5), 5359–5365.

Chen, W., Ma, C., & Ma, L. (2009). Mining the customer credit using hybrid support vector machine technique. *Expert Systems with Applications*, *36*(4), 7611–7616.

Chuang, C.-L., & Lin, R.-H. (2009). Constructing a reassigning credit scoring model. *Expert Systems with Applications*, *36*(2), 1685–1694.

Drummond, C., & Holte, R. C. (2006). Cost curves: An improved method for visualizing. *Machine Learning*, *65*(1), 95–130.

Fawcett, T. (2006). An introduction to ROC analysis: ROC Analysis in pattern recognition. *Pattern Recognition Letters*, *27*(8), 861–874.

Hand, D. J. (2001). Modeling consumer credit risk. *IMA Journal of Management Mathematics*, *12*(2), 139–155.

Hand, D. J. (2005). Good practice in retail credit scorecard assessment. *Journal of the Operational Research Society*, *56*(9), 1109–1117.

Hand, D. J. (2006). Classifier technology and the illusion of progress. *Statistical Science*, *21*(1), 1–15.

Hand, D. J. (2009). Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning*, *77*(1), 103–123.

He, J., Zhang, Y., Shi, Y., & Huang, G. (2010). Domain-driven classification based on multiple criteria and multiple constraint-level programming for intelligent credit scoring. *IEEE Transactions on Knowledge and Data Engineering*, *22*(6), 826–838.

Henley, W. E., & Hand, D. J. (1997). Construction of a k-nearest-neighbour credit-scoring system. *IMA Journal of Mathematics Applied in Business & Industry*, *8*(4), 305–321.

Hoffmann, F., Baesens, B., Martens, J., Put, F., & Vanthienen, J. (2002). Comparing a genetic fuzzy and a neurofuzzy classifier for credit scoring. *International Journal of Intelligent Systems*, *17*(11), 1067–1083.

Hoffmann, F., Baesens, B., Mues, C., van Gestel, T., & Vanthienen, J. (2007). Inferring descriptive and approximate fuzzy rules for credit scoring using evolutionary algorithms. *European Journal of Operational Research*, *177*(1), 540–555.

Holmes, C. C., & Denison, D. G. T. (2003). Classification with Bayesian MARS. *Machine Learning*, *50*(1-2), 159–173.

Hsieh, N.-C. (2005). Hybrid mining approach in the design of credit scoring models. *Expert Systems with Applications*, *28*(4), 655–665.

Hsieh, N.-C., & Hung, L.-P. (2010). A data driven ensemble classifier for credit scoring analysis. *Expert Systems with Applications*, *37*(1), 534–545.

Huang, C.-L., Chen, M.-C., & Wang, C.-J. (2007). Credit scoring with a data mining approach based on support vector machines. *Expert Systems with Applications*, *33*(4), 847–856.

Huang, J.-J., Tzeng, G.-H., & Ong, C.-S. (2006). Two-stage genetic programming (2SGP) for the credit scoring model. *Applied Mathematics and Computation*, *174*(2), 1039–1053.

Ince, H., & Aktan, B. (2009). A comparison of data mining techniques for credit scoring in banking: A managerial perspective. *Journal of Business Economics and Management*, *10*(3), 233–240.

Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. *Expert Systems with Applications*, *37*(9), 6233–6239.

Lee, T.-S., & Chen, I.-F. (2005). A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, *28*(4), 743–752.

Lee, T.-S., Chiu, C.-C., Chou, Y.-C., & Lu, C.-J. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis*, *50*(4), 1113–1130.

Lee, T.-S., Chiu, C.-C., Lu, C.-J., & Chen, I.-F. (2002). Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications*, *23*(3), 245–254.

Li, J., Wei, L., Li, G., & Xu, W. (2011). An evolution strategy-based multiple kernels multi-criteria programming approach: The case of credit decision making. *Decision Support Systems*, *51*(2), 292–298.

Li, S.-T., Shiue, W., & Huang, M.-H. (2006). The evaluation of consumer loans using support vector machines. *Expert Systems with Applications*, *30*(4), 772–782.

Loterman, G., Brown, I., Martens, D., Mues, C., & Baesens, B. (2011). Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting*, *28*(1), 161–170.

Luo, S.-T., Cheng, B.-W., & Hsieh, C.-H. (2009). Prediction model building with clustering-launched classification and support vector machines in credit scoring. *Expert Systems with Applications*, *36*(4), 7562–7566.

Malhotra, R., & Malhotra, D. K. (2002). Differentiating between good credits and bad credits using neuro-fuzzy systems. *European Journal of Operational Research*, *136*(1), 190–211.

Malhotra, R., & Malhotra, D. K. (2003). Evaluating consumer loans using neural networks. *Omega - The international Journal of Management Science*, *31*(2), 83–96.

Martens, D., Baesens, B., van Gestel, T., & Vanthienen, J. (2007). Comprehensible credit scoring models using rule extraction from support vector machines. *European Journal of Operational Research*, *183*(1), 1466–1476.

Martens, D., van Gestel, T., Backer, M. de, Haesen, R., Vanthienen, J., & Baesens, B. (2010). Credit rating prediction using ant colony optimization. *Journal of the Operational Research Society*, *61*(4), 561–573.

Matuszyk, A., Mues, C., & Thomas, L. C. (2009). Modelling LGD for unsecured personal loans: Decision tree approach. *Journal of the Operational Research Society*, *61*(3), 393–398.

Nanni, L., & Lumini, A. (2009). An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, *36*(2), 3028–3033.

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature: On quantitative methods for detection of financial fraud. *Decision Support Systems*, *50*(3), 559–569.

Ong, C.-S., Huang, J.-J., & Tzeng, G.-H. (2005). Building credit scoring models using genetic programming. *Expert Systems with Applications*, *29*(1), 41–47.

Peng, Y., Kou, G., Shi, Y., & Chen, Z. (2008). A multi-criteria convex quadratic programming model for credit data analysis. *Decision Support Systems*, *44*(4), 1016–1030.

Ping, Y., & Yongheng, L. (2011). Neighborhood rough set and SVM based hybrid credit scoring classifier. *Expert Systems with Applications*, *38*(9), 11300–11304.

Provost, F., & Fawcett, T. (1997). Analysis and visuatization of classifier performance: Comparison under imprecise class and cost distributions. In *Proceedings of the third International Conference on Knowledge Discovery and Data Mining (KDD-97)* (pp. 43–48).

Provost, F., & Fawcett, T. (2001). Robust classification for imprecise environments. *Machine Learning*, *42*(3), 203–231.

Setiono, R., Baesens, B., & Mues, C. (2008). Recursive neural network rule extraction for data with mixed attributes. *IEEE Transactions on Neural Networks*, *19*(2), 299–307.

Šušteršič, M., Mramor, D., & Zupan, J. (2009). Consumer credit scoring models with limited data. *Expert Systems with Applications*, *36*(3), 4736–4744.

Thomas, L. C., Oliver, R. W., & Hand, D. J. (2005). A survey of the issues in consumer credit modelling research. *Journal of the Operational Research Society*, *56*(9), 1006–1015.

Tsai, C.-F., & Wu, J.-W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, *34*(4), 2639–2649.

Tsai, M.-C., Lin, S.-P., Cheng, C.-C., & Lin, Y.-P. (2009). The consumer loan default predicting model - An application of DEA-DA and neural network. *Expert Systems with Applications*, *36*(9), 11682–11690.

Tsakonas, A., & Dounias, G. (2007). Evolving neural-symbolic systems guided by adaptive training schemes: Applications in finance. *Applied Artificial Intelligence*, *21*(7), 681–706.

Viaene, S., & Dedene, G. (2005). Cost-sensitive learning and decision making revisited: Metaheuristics and worst-case guarantee algorithms: Relations, provable properties and applications. *European Journal of Operational Research*, *166*(1), 212–220.

Wang, C.-M., & Huang, Y.-F. (2009). Evolutionary-based feature selection approaches with new criteria for data mining: A case study of credit approval data. *Expert Systems with Applications*, *36*(3), 5900–5908.

Wang, Y., Wang, S., & Lai, K. K. (2005). A new fuzzy support vector machine to evaluate credit risk. *IEEE Transactions on Fuzzy Systems*, *13*(6), 820–831.

West, D. (2000). Neural network credit scoring models. *Computers & Operations Research*, *27*(11-12), 1131–1152.

Xiao, W., Zhao, Q., & Fei, Q. (2006). A comparative study of data mining methods in consumer loans credit scoring management. *Journal of Systems Science and Systems Engineering*, *15*(4), 419–435.

Xu, X., Zhou, C., & Wang, Z. (2009). Credit scoring algorithm based on link analysis ranking with support vector machine. *Expert Systems with Applications*, *36*(2), 2625–2632.

Yang, Y. (2007). Adaptive credit scoring with kernel learning methods. *European Journal of Operational Research*, *183*(3), 1521–1536.

Yeh, I.-C., & Lien, C.-h. (2009). The comparison of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, *36*(2), 2473–2480.

Yobas, M. B., Crook, J. N., & Ross, P. (2000). Credit scoring using neural and evolutionary techniques. *IMA Journal of Mathematics Applied in Business & Industry*, *11*(2), 111–125.

Yu, L., Wang, S., & Cao, J. (2009a). A modified least squares support vector machine classifier with application to credit risk analysis. *International Journal of Information Technology & Decision Making*, *8*(4), 697–710.

Yu, L., Wang, S., & Lai, K. K. (2008a). Credit risk assessment with a multistage neural network ensemble learning approach. *Expert Systems with Applications*, *34*(2), 1434–1444.

Yu, L., Wang, S., & Lai, K. K. (2009b). An intelligent-agent-based fuzzy group decision making model for financial multicriteria decision support: The case of credit scoring. *European Journal of Operational Research*, *195*(3), 942–959.

Yu, L., Wang, S., Wen, F., Lai, K. K., & He, S. (2008b). Designing a hybrid intelligent mining system for credit risk evaluation. *Journal of Systems Science & Complexity*, *21*(4), 527–539.

Yu, L., Yao, X., Wang, S., & Lai, K. K. (2011). Credit risk evaluation using a weighted least squares SVM classifier with design of experiment for parameter selection. *Expert Systems with Applications*, *38*(12), 15392–15399.

Yu, L., Yue, W., Wang, S., & Lai, K. K. (2010). Support vector machine based multiagent ensemble learning for credit risk evaluation. *Expert Systems with Applications*, *37*(2), 1351–1360.

Zhang, D., Zhou, X., Leung, S. C. H., & Zheng, J. (2010). Vertical bagging decision trees model for credit scoring. *Expert Systems with Applications*, *37*(12), 7838–7843.

Zhang, J., Shi, Y., & Zhang, P. (2009). Several multi-criteria programming methods for classification. *Computers & Operations Research*, *36*(3), 823–836.

Zhou, L., Lai, K. K., & Yen, J. (2009a). Credit scoring models with AUC maximization based on weighted SVM. *International Journal of Information Technology & Decision Making*, *8*(4), 677–696.

Zhou, L., Lai, K. K., & Yu, L. (2009b). Credit scoring using support vector machines with direct search for parameters selection. *Soft Computing*, *13*(2), 149–155.

Zhou, L., Lai, K. K., & Yu, L. (2010). Least squares support vector machines ensemble models for credit scoring. *Expert Systems with Applications*, *37*(1), 127–133.

Zhou, X., Jiang, W., Shi, Y., & Tian, Y. (2011). Credit risk evaluation with kernel-based affine subspace nearest points learning method. *Expert Systems with Applications*, *38*(4), 4272–4279.

Zuccaro, C. (2010). Classification and prediction in customer scoring. *Journal of Modelling in Management*, *5*(1), 38–53.

Appendix A

Study distribution across journals:

Journal	No. studies
Expert Systems with Applications	31
European Journal of Operational Research	5
Computers & Operations Research	2
Decision Support Systems	2
International Journal of Information Technology & Decision Making	2
Journal of the Operational Research Society	2
Applied Artificial Intelligence	1
Applied Mathematics and Computation	1
Computational Statistics & Data Analysis	1
IEEE Transactions on Fuzzy Systems	1
IEEE Transactions on Knowledge and Data Engineering	1
IEEE Transactions on Neural Networks	1
IMA Journal of Mathematics Applied in Business and Industry	1
International Journal of Intelligent Systems	1
Journal of Applied Statistics	1
Journal of Business Economics and Management	1
Journal of Modeling in Management	1
Journal of Systems Science and Complexity	1
Journal of Systems Science and Systems Engineering	1
Machine Learning	1
Management Science	1
Omega	1
Software Computation	1
The Journal of Risk Finance	1

Appendix B

Overview of short cuts in *tab. 1*, *figs. 3, 5* and *7* and performance indicator calculations:

Shortcut	Performance indicator	Calculation
HR/ER	Overall hit rate, overall error rate	HR = (TP + TN) / (P +N) ER = (FP + FN) / (P+N)
HR-B	Hit rate of "bad risk"	HR-B = TN / N
HR-G	Hit rate of "good risk"	HR-G = TP / P
ER-B	Error rate of "bad risk"	ER-B = FP / N
ER-G	Error rate of "good risk"	ER-G = FN / P
BRaA	Bad rate among accepts	BRaA = FP / (TP + FP)
ER-Other	Other error indicators, including the F1 indicator (FM), Correlation coefficient (CC), Gain chart (GC), and Area of gain chart (AC), Bad rate among rejected (BRaR)	-
AUC	Area under the Receiver Characteristic Curve	-
ROC	Receiver Characteristic Curve	-
KS	Kolmogorov-Smirnov statistic	-

Appendix C

Classification of studies among the number of utilised performance indicators:

Dimensionality	No. PI	References
One-dimensionality	**1**	Baesens et al., 2003a; Hoffmann et al., 2002; Hoffmann et al., 2007; Holmes & Denison, 2003; Hsieh & Hung, 2010; Huang et al., 2007; Huang et al., 2006; Khashman, 2010; Luo et al., 2009; Martens et al., 2007; Martens et al., 2010; Ong et al., 2005; Ping & Yongheng, 2011; Setiono et al., 2008; Tsakonas & Dounias, 2007; Yobas et al., 2000; Zhang et al., 2010
Multi-dimensionality	**2**	Bellotti & Crook, 2009; Xu et al., 2009; Zhou et al., 2009a
	3	Chen & Huang, 2003; Chen & Huang, 2011; Chen & Li, 2010; Hsieh, 2005; Ince & Aktan, 2009; Li et al., 2006; Li et al., 2011; Malhotra & Malhotra, 2002; Malhotra & Malhotra, 2003; Šušteršič et al., 2009; Tsai & Wu, 2008; Wang et al., 2005; Yang, 2007; Yeh & Lien, 2009; Yu et al., 2009a; Yu et al., 2008a; Yu et al., 2008b; Yu et al., 2011; Yu et al., 2010; Zhang et al., 2009; Zhou et al., 2009b; Zhou et al., 2011; Zuccaro, 2010
	4	Abdou et al., 2008*; Baesens et al., 2003b; He et al., 2010; Nanni & Lumini, 2009; Tsai et al., 2009*; West, 2000*; Xiao et al., 2006*; Yu et al., 2009b; Zhou et al., 2010
	5	Chen et al., 2009; Chuang & Lin, 2009; Lee et al., 2006; Lee et al., 2002; Wang & Huang, 2009
	6	Abdou, 2009a*; Abdou, 2009b*; Antonakis & Sfakianakis, 2009; Lee & Chen, 2005*
	7	Peng et al., 2008

*) Research studies that focus on a cost and error minimisation perspective (PI = performance indicator)

III. Bewertung und Auswahl von Scorecards im Kreditwesen:

Eine Analyse zur Eignung von *Kosten-Kurven*

Authors: **Nicole Martin**

Stefan Lessmann

Year: *2012*

Bibliographic Status: *Working Paper,*

Submitted to

Zeitschrift für betriebswirtschaftliche

Forschung

Bewertung und Auswahl von Scorecards im Kreditwesen:

Eine Analyse zur Eignung von *Kosten-Kurven*

Nicole Martin*

Stefan Lessmann**

*Dipl.-Kffr. Nicole Martin,
Lehrstuhl für BWL - Marketing und Branding, Institut für Marketing und Medien, Universität Hamburg, Welckerstraße 8, 20354 Hamburg, Deutschland,
nicole.martin@wiso.uni-hamburg.de

**Dr. Stefan Lessmann,
Institut für Wirtschaftsinformatik, Universität Hamburg,
Von-Melle-Park 5, 20146 Hamburg, Deutschland,
lessmann@econ.uni-hamburg.de

Zusammenfassung:

Scorecards werden im Kreditwesen routinemäßig eingesetzt, um Entscheidungsprozesse im Marketing und im Risikomanagement zu unterstützen. Dem Einsatz einer Scorecard geht ein Auswahlprozess voraus, in dessen Rahmen alternative Modelle entwickelt und verglichen werden. Der vorliegende Beitrag behandelt die Frage, wie diese Scorecardbewertung bzw. Scorecardauswahl erfolgen sollte. Hierzu wird ein Kriterienkatalog entwickelt, der die spezifischen Anforderungen des Kreditwesens zusammenfasst. Auf dieser Basis werden gebräuchliche Instrumente zur Scorecardauswahl und Scorecardbewertung analysiert und deren Schwächen offenbart. Mit den *Kosten-Kurven* wird ein neues Bewertungsinstrument für das Kreditwesen vorgestellt und empirisch verdeutlicht, welche Vorteile sich aus seinem Einsatz ergeben. Eine wesentliche Implikation des Beitrags ist, dass *Kosten-Kurven* eine ökonomisch motivierte Scorecardbewertung ermöglichen und damit zu einer höheren Entscheidungsqualität in Scorecard-gestützten Geschäftsprozessen beitragen.

JEL-Classification: M 10

Keywords: Kreditwirtschaft, Risikomanagement, Entscheidungsunterstützung, Klassifikation

Bewertung und Auswahl von Scorecards im Kreditwesen:
Eine Analyse zur Eignung von *Kosten-Kurven*

1 Einleitung

Scorecards sind ein universelles Instrument zur Unterstützung betrieblicher Entscheidungen. In der Finanzdienstleistungsindustrie ist ihr Einsatz besonders verbreitet. Sie werden zum Beispiel in kunden- und marketingbezogenen Geschäftsprozessen genutzt, um auf Basis von Vergangenheitsdaten das Verhalten von Kunden zu antizipieren[1]. Ein klassisches Beispiel ist die Bewertung von Kreditanträgen, um über die Vergabe bzw. Verweigerung eines Darlehens zu entscheiden. Daneben existieren zahlreiche weitere Anwendungen, die sich zum Beispiel mit Fragen der Akzeptanz (würde der Kunde ein bestimmtes Angebot annehmen), der Nutzung (wird ein Kreditprodukt kontinuierlich genutzt werden) beziehungsweise der Weiternutzung (wird das Produkt auch nach Ablauf eines Einführungsangebots weiter verwendet) sowie des Ausfallmanagements (wie sollte mit säumigen Kunden umgegangen werden) beschäftigen[2].

Dem Einsatz einer Scorecard geht ein Auswahlprozess voraus, in dem ein Analyst mehrere alternative Scorecards erstellt, vergleicht und bewertet. Die Wahl der „richtigen" Alternative ist ökonomisch relevant, da sie die Vorhersagequalität der Scorecard unmittelbar beeinflusst[3]. Aufgrund der intensiven Nutzung von Scorecards sowie Skaleneffekten durch die große Menge an Kundenbeziehungen sind gerade im Finanzdienstleistungsbereich erhebliche Gewinnsteigerungen möglich, wenn die Vorhersagequalität von Scorecards verbessert wird[4].

Der vorliegende Beitrag verfolgt vor diesem Hintergrund folgende Ziele: (1) Es sollen die spezifischen Anforderungen der Scorecardbewertung im Finanzdienstleistungsbereich herausgearbeitet und in einem Kriterienkatalog zusammengefasst werden. (2) Es soll verdeutlicht werden, dass derzeit übliche Vorgehensweisen zur Scorecardbewertung und Scorecardauswahl nur bedingt für Anwendungen im Finanzdienstleistungsbereich geeignet sind. (3) Es soll mit den *Kosten-Kurven*[5] ein neues Bewertungsinstrument für das Kreditwesen vorgestellt werden, dass die zuvor entwickelten Anforderungen besser erfüllt und damit den Prozess der Scorecardauswahl besser unterstützt. (4) Es soll empirisch verdeutlicht werden, wie *Kosten-Kurven* im Kreditwesen eingesetzt werden können und welche ökonomischen Vorteile sich aus ihrem Einsatz ergeben.

---

[1] Vgl. *Thomas* (2000).
[2] Vgl. *Thomas/Oliver/Hand* (2005).
[3] Vgl. *Hand* (2005).
[4] Vgl. *Reichheld/Sasser* (1990); *Baesens et al.* (2003); *Burez/Van den Poel* (2007); *Larivière/Van den Poel* (2005).
[5] Vgl. *Drummond/Holte* (2000); *Drummond/Holte* (2006).

Zu diesem Zweck werden zunächst in Kapitel 2 entscheidungsrelevante Bewertungskriterien für die Auswahl einer Scorecard hergeleitet und systematisiert. Darauf aufbauend erfolgt eine Kategorisierung aktuell gebräuchlicher Gütemaße. Anschließend werden in Kapitel 3 die Grundlagen des *Kosten-Kurven*-Prinzips vor dem Hintergrund eines Einsatzes im Kreditwesen erläutert und die angenommenen Vorteile von *Kosten-Kurven* anhand einer Simulationsstudie empirisch validiert. Die Arbeit wird mit einer Zusammenfassung der wesentlichen Erkenntnisse und einem Ausblick auf zukünftige Forschungsaktivitäten in Kapitel 4 beschlossen.

2 Bewertung von Scorecards im Kreditwesen

Scorecards dienen allgemein der Prognose zukünftigen Kundenverhaltens. Um zum Beispiel zu entscheiden, ob ein Darlehen bewilligt werden soll, verdichtet eine Scorecard sämtliche Angaben des Kreditantrags (Kredithöhe, -zweck/-laufzeit, demografische Daten des Antragstellers, etc.) mittels einer mathematischen Vorschrift auf einen nummerischen Wert, den sogenannten *Credit-Score*[6]. Je höher der Wert ausgeprägt ist, desto größer ist die Wahrscheinlichkeit, dass der betreffende Kunde ein bestimmtes Verhalten zeigen wird, zum Beispiel den Kredit ordnungsgemäß zurückzahlen würde. Um eine Entscheidung über die weitere Behandlung des potenziellen Kunden zu fällen, wird der Score mit einem Schwellenwert τ verglichen. Ist dieser übertroffen, wird eine bestimmte Aktion ausgelöst (zum Beispiel Bewilligung des Antrags).

Im vorliegenden Beitrag wird die Eignung einer Scorecard durch die Genauigkeit (auch als Güte bezeichnet) ihrer Prognosen bewertet. Speziell im Finanzdienstleistungsbereich ist weiterhin die Verständlichkeit von Scorecards von großer Bedeutung und zum Teil gesetzlich vorgeschrieben[7]. Zur Auswahl einer Scorecard aus einer Menge prinzipiell opportuner Alternativen sind entsprechende Kriterien aber weniger geeignet, da sie in der Regel keine quantitative Bewertung erlauben und ein objektiver Vergleich damit kaum möglich ist.

Die Prognosegüte einer Scorecard wird durch einen Vergleich ihrer Vorhersagen (*Credit-Scores*) mit dem tatsächlich beobachteten Kundenverhalten gemessen[8]. Es ist allgemein akzeptiert, dass eine verbesserte Prognoseleistung zu substanziellen Gewinnsteigerungen führen kann[9]. Allerdings finden sich in der Literatur viele alternative Gütemaße, die verschiedene Facetten der Prognosegenauigkeit betonen. Speziell im Finanzdienst-

[6] Vgl. *Hartmann-Wendels/Pfingsten/Weber* (2010).
[7] Vgl. z. B. *Hofmann* (2007); *Martens et al.* (2007).
[8] Für die Antragsbewertung wird beispielsweise auf seinerzeit gewährte Anträge inkl. deren Charakteristika sowie das tatsächlich beobachtete Zahlungsverhalten zurückgegriffen.
[9] Vgl. z. B. *Reichheld/Sasser* (1990); *Van den Poel/Lariviere* (2004).

leistungsbereich bietet es sich an, alternative Scorecards gemäß ökonomischer Kriterien zu bewerten[10].

2.1 Rahmenbedingungen einer ökonomischen Bewertung

Eine ökonomische Bewertung alternativer Scorecards ist einer rein statistischen Betrachtung vorzuziehen. Sie stellt dem Management entscheidungsrelevante Informationen zur Verfügung und liefert inhaltlich nachvollziehbare Argumente für die Wahl einer Handlungsoption. Die Umsetzung einer ökonomischen Scorecardbewertung gestaltet sich in der Kreditwirtschaft jedoch als schwierig, was nachfolgend am Beispiel der Kreditantragsbewertung verdeutlicht werden soll. Dabei wird vom Ziel der Kostenminimierung ausgegangen, da dies in der Literatur die übliche Vorgehensweise ist.[11] Kunden, die einen Kredit ordnungsgemäß bedienen würden, werden nachfolgend als „gute" Kunden bzw. Mitglieder der Gruppe „positiv" bezeichnet. Analog wird von „schlechten" Kunden/Kreditrisiken bzw. Mitgliedern der Gruppe „negativ" gesprochen[12].

Es sei $p(+ \mid x)$ die durch die Scorecard geschätzte *a posteriori*-Wahrscheinlichkeit, dass ein Kunde *x* der Gruppe positiv angehört. Gemäß der *Bayes'schen* Entscheidungstheorie gilt dann folgende Regel für die Zuordnung eines Kunden zur positiven Klasse (entspricht der Gewährung eines Kredits)[13]:

$$p(+ \mid x) > \tau^* = \frac{C(+ \mid -)}{C(+ \mid -) + C(- \mid +)} \tag{1}$$

Dabei beschreibt τ^* den kostenminimalen Schwellenwert. $C(-|+)$ und $C(+|-)$ repräsentieren die sogenannten Fehlerkosten. Fehlerkosten umfassen sowohl die Konsequenzen einer Fehlentscheidung, wenn einem potenziell „schlechten" Kreditnehmer ein Darlehen gewährt wird, woraus ein Verlust in Höhe des nicht zurückgezahlten Betrags entsteht. Diese Kosten werden in (1) durch $C(+|-)$ repräsentiert; falsche Zuordnung eines tatsächlich negativen Objekts zur Gruppe positiv. Zudem entstehen Kosten wenn ein Kredit verweigert wird, obgleich dieser ordnungsgemäß bedient worden wäre. Hieraus folgen Opportunitätskosten in Höhe der entgangenen Zinsgewinne[14]. Dies werden in (1) mit $C(-|+)$ repräsentiert; falsche Zuordnung eines tatsächlich positiven Objekts zur Gruppe negativ.

[10] Vgl. *Hand* (2005).
[11] Vgl. *Viaene/Dedene* (2004).
[12] In der Literatur ist die Bezeichnung „positive"/"negative" Klasse üblich. Sie wurde daher übernommen. Methodisch ist es nicht von Belang, ob die im Sinne der Anwendung guten Kunden (geringes Kreditausfallrisiko) mit + oder – kodiert werden.
[13] Vgl. *Adams/Hand* (1999).
[14] Vgl. z. B. *West* (2000).

Unter Verwendung des Bayes'schen Theorems lässt sich (1) wie folgt umformen[15]:

$$\frac{p(x|+)}{p(x|-)} > \frac{C(+|-)\cdot p(-)}{C(-|+)\cdot p(+)} \qquad (2)$$

Die linke Seite von *Gleichung (2)* repräsentiert das sogenannte Likelihood-Ratio. Die rechte Seite gibt erneut den kostenminimalen Schwellenwert τ^* an, allerdings bezogen auf das Likelihood-Ratio. Dabei repräsentieren $p(+)$ und $p(-)$ die *a priori*-Wahrscheinlichkeiten beider Klassen, das heißt die in einer (Trainings-) Datenbasis vorliegenden Klassenverteilung. Diese repräsentiert die Zugehörigkeit der Kunden zu der Gruppe der „guten" Kreditnehmer (in Form von Zahlern) beziehungsweise zu der Gruppe der „schlechten" Kreditnehmer (in Form von Nicht-Zahlern)[16].

Gleichung (2) verdeutlicht, dass eine Entscheidung auf Basis der probabilistischen Scorecardprognosen unter der Zielsetzung der Kostenminimierung exakte Kenntnisse über zwei zentrale Anwendungsparameter voraussetzt: die mit falschen Prognosen beziehungsweise Entscheidungen verbundenen Kosten (Fehlerkosten) und die relative Verteilung der beiden Klassen in der Grundgesamtheit (*a priori*-Wahrscheinlichkeiten).

2.2 Anforderungen an Gütemaße

Die Prinzipien einer ökonomischen Scorecard-Bewertung sind zwar prinzipiell bekannt[17], konnten sich bisher jedoch in der praktischen Realität nicht durchsetzen[18]. Die Autoren führen diesen Umstand auf die Schwierigkeiten zurück, die mit der Implementierung einer entsprechenden Bewertungsphilosophie verbunden sind. Anforderungen, die sich aus solch einer Bewertungsphilosophie und den damit verknüpften Anwendungsparametern ergeben, lassen sich durch folgende Bewertungskriterien zusammenfassen: *Schiefe Verteilungsstrukturen*, *Konkretisierungsgrad, Instationarität* und *Informationsgrad.*

Schiefe Verteilungsstrukturen: Auf der Seite der Anwendungsparameter wird üblicherweise von asymmetrisch verteilten Fehlerkosten ausgegangen, da die Folgen eines Kreditausfalls wesentlich schwerer wiegen als die Verweigerung eines Darlehens an einen potenziell „guten" Kreditnehmer[19]. Auch die (Trainings-) Datenbasis ist durch eine schiefe Verteilungsstruktur gekennzeichnet. Typischerweise lässt sich eine ungleich verteilte Zugehörigkeit der Kunden zu der Gruppe der „guten" Kreditnehmer (in Form von Zahlern) beziehungsweise zu der Gruppe der „schlechten" Kreditnehmer (in Form von Nicht-Zahlern)

[15] Vgl. *Duda/Hart/Stork* (2001).
[16] Vgl. *Thomas/Oliver/Hand* (2005).
[17] Vgl. *Hartmann-Wendels/Pfingsten/Weber* (2010).
[18] Vgl. *Baesens/Gestel* (2009); *Hand* (2005); *Thomas* (2000).
[19] Vgl. *Thomas/Edelman/Crook* (2002).

feststellen. Da dem Unternehmen lediglich Daten zu genehmigten Krediten und der korrespondierenden Historie vorliegen, beinhaltet die Datenbasis deshalb weniger Nicht-Zahler als Zahler[20]. Für die Realisierung einer ökonomischen Bewertungsperspektive ist es daher notwendig, dass angewandte Gütemaße eine gewisse Robustheit gegenüber schiefen Verteilungsstrukturen der Anwendungsparameter aufweisen, sodass asymmetrische Verteilungsstrukturen nicht zu verzerrten Handlungsempfehlungen führen[21].

Konkretisierungsgrad: Anwendungsparameter in Form von Fehlerkosten können in praktischen Anwendungen oftmals nicht genau quantifiziert werden. Allerdings ist davon auszugehen, dass tendenzielle Aussagen, das heißt ein Fehler wiegt schwerer als der andere (z. B. etwa 3 bis 5 Mal so schwer), möglich sind[22]. Demnach ist es sicherlich häufiger der Fall, dass Erwartungen über ein Kostenverhältnis, in dem sich die Fehlerkosten bewegen dürften, vorliegen. Dennoch mag in seltenen Fällen lediglich bekannt sein, dass ein Fehlertyp schwerer wiegt als ein anderer. Darüber hinaus bleibt dem Unternehmen die wahre Klassenverteilung aller Antragssteller zumeist verborgen, da nur erschwerte Rückschlüsse über das Zahlungsverhalten abgelehnter Kredite möglich sind. Ein weiterer zentraler Aspekt, der den Konkretisierungsgrad maßgeblich prägt, ist der Umstand, dass die Analyse und Bewertung der Klassifikationsleistung auf einem Beispieldatensatz und damit auf einer Klassenverteilung beruht, die die spätere Anwendungssituation nur unzureichend wiedergibt[23]. Hieraus resultiert eine für die Anwendung so typische Situation, dass oftmals nur Tendenzaussagen über Verteilungsstrukturen vorliegen dürften. Für eine Realisierung einer ökonomisch relevanten Bewertungsperspektive ist es daher unabdingbar, dass eingesetzte Gütemaße unterschiedliche Konkretisierungsgrade verarbeiten. Nur hierdurch kann gewährleistet werden, dass alle zur Verfügung stehenden Informationen in die Gütebewertung einfließen.

Instationarität: Weiterhin ist fraglich, inwieweit Klassenverteilungen zeitlich stabil sind. Das Nicht-Zurückzahlen eines Kredits ist oftmals Resultat einer mangelnden Zahlungsfähigkeit des Kreditnehmers, zum Beispiel durch plötzliche Arbeitslosigkeit, Kurzarbeit, die wiederum aus Änderungen im ökonomischen Umfeld folgt. Ähnlich verhält es sich mit angenommenen Kostenverteilungen, die auf einer klassenbasierten Perspektive beruhen und aggregiert potenzielle Kreditausfälle sowie Opportunitätskosten betrachten. Demnach kann

[20] Vgl. *Banasik/Crook/Thomas* (2003).
[21] Eine analoge Argumentation kann z. B. *Provost/Fawcett/Kohavi* (1998) entnommen werden.
[22] Vgl. *Adams/Hand* (1999); *Hand* (2001).
[23] Vgl. *Provost/Fawcett/Kohavi* (1998).

angenommen werden, dass die Verteilung „guter“ und „schlechter“ Kunden als auch die der Kosten von externen Faktoren beeinflusst wird und zeitlich variiert. Hieraus kann die Anforderung abgeleitet werden, dass bei der Bewertung von Klassifikationsprognosen solche Gütemaße ihren Einsatz finden sollten, die aufzeigen, inwieweit eine Überlegenheit einer Scorecard auch bei Veränderungen von Anwendungsparametern stabil ist.

Informationsgrad: Die Quantifizierung von Kosten- und Klassenverteilungen repräsentiert ein eigenständiges Schätzproblem, das mit erheblicher Unsicherheit behaftet ist. Der erwartete Verlust eines Kreditgeschäfts hängt neben der Ausfallwahrscheinlichkeit von der erwarteten Forderungshöhe zum Zeitpunkt des Ausfalls und der Verlustquote bei Ausfall ab. Beide Größen sind selbst zufallsbehaftet[24]. Empirische Untersuchungen verdeutlichen, dass die Vorhersage eben dieser Größen eine anspruchsvolle Prognoseaufgabe repräsentiert, die nur mit begrenzter Genauigkeit gelöst werden kann[25]. Die Schätzung der Opportunitätskosten nicht gewährter Kredite ist vermeintlich noch schwieriger, da hier die Profitabilität abgelehnter Kunden, für die per Definition keine weiteren Informationen vorliegen, geschätzt werden muss. Darüber hinaus ist die Kenntnis über die wahre Klassenverteilung aller Antragssteller mit hoher Ungenauigkeit verbunden, da Rückschlüsse über das Zahlungsverhalten abgelehnter Kredite erschwert möglich sind. Basierend auf einer Abweichung zwischen den verwendeten (historischen) Beispieldaten und den eigentlichen Anwendungsdaten ist die Approximation der wahren Klassenverteilung zu einem Zeitpunkt mit Unsicherheit behaftet. In Anbetracht dieser Schwierigkeiten ist davon auszugehen, dass alle relevanten Anwendungsparameter gewissen Schätzungenauigkeiten unterliegen. Hieraus resultiert die Notwendigkeit, dass ein Bewertungsinstrument darüber Aufschluss geben sollte, ob die Prognoseleistung einer Scorecard zu relevanten Prognoseverbesserungen führen kann. Relevante Prognoseleistungen wären dadurch determiniert, dass Prognoseverbesserungen in solchen Verteilungsspannen auftreten, die für den Anwendungskontext als wahrscheinlich gelten. Die Relevanz dieser Anforderung zeigt sich insbesondere dann, wenn eine Scorecard zu bestimmten Rahmenbedingungen eine andere dominiert, diese jedoch bei einer anderen Konstellation unterlegen ist.

Im Folgenden soll nun ein kurzer Überblick über weit verbreitete Bewertungsansätze für Scorecards gegeben werden, die im Anschluss anhand der abgeleiteten Bewertungskriterien klassifiziert werden. Wie diese Klassifizierung zeigt, sind beide Bewertungsansätze nur

[24] Vgl. *Hartmann-Wendels/Pfingsten/Weber* (2010).
[25] Vgl. *Loterman et al.* (2011); *Matuszyk/Mues/Thomas* (2009).

bedingt für Credit-Scoring-Anwendungen geeignet, was die Suche nach alternativen Instrumenten, wie den hier (Kapitel 3) betrachteten *Kosten-Kurven*, motiviert.

2.3 Eignung gebräuchlicher Gütemaße

Bewertungsansätze für Scorecards können in zwei Gruppen eingeteilt werden. Einerseits kann die Güte einer diskreten Einteilung von Kunden in die Klassen „gut"/"schlecht" mittels einer Kontingenztabelle bewertet werden. Andererseits kann die Güte einer Scorecard direkt auf Basis der numerischen Scores erfolgen. Als Repräsentant der zweiten Gruppe wird im vorliegenden Beitrag die *Receiver-Operating-Characteristics* (ROC) Analyse[26] betrachtet, die im Credit-Scoring weit verbreitet ist[27].

2.3.1 Kontingenztabellen zur Scorecard-Bewertung

Kontingenztabellen basieren auf einer diskreten Einteilung von Kunden in bekannte Klassen. Es sei $p(+\mid x)$ die durch eine Scorecard geschätzte *a posteriori*-Wahrscheinlichkeit, dass ein Testfall *x* der Gruppe „positiv" angehört, mit analoger Bedeutung für $p(-\mid x)$. Eine diskrete Klassenprognose $\hat{Y}(x) \in \{+,-\}$ ergibt sich dann unter Verwendung des Schwellenwertes τ, gemäß (3):

$$\hat{Y}(x) = \begin{cases} + & \text{wenn } p(+\mid x) \geq \tau \\ - & \text{wenn } p(+\mid x) < \tau \end{cases} \tag{3}$$

Eine Kontingenztabelle stellt die möglichen Kombinationen geschätzter und tatsächlicher Klassenzugehörigkeiten in einem Vier-Felder-Schema dar (vgl. *Tabelle 1*):

Tabelle 1: Kontingenztabelle einer binären Klassifikation in die Gruppen + und -

		Scorecard-Prognose (diskretisiert)		
		+	-	
Tatsächliche Klasse	+	*TP*	*FN*	$n^+ = TP + FN$
	-	*FP*	*TN*	$n^- = FP + TN$

Ist ein Antrag der positiven Klasse zugehörig und liegt der prognostizierte Score oberhalb des Schwellenwertes, so liegt ein richtig-positiver (*TP*) Zustand vor. Gilt für diesen Antrag dagegen $p(+\mid x) < \tau$, liegt ein falsch-negativer (*FN*) Zustand vor. Die dazugehörigen Raten *TPR* und *FNR* ergeben sich durch Division mit der Gesamtzahl positiver Anträge (n^+) in der Datenmenge. Der Umgang mit „negativen" Anträgen ist analog und liefert die Anzahl der

[26] Vgl. *Swets* (1988).
[27] Vgl. *Basel Committee on Banking Supervision* (2005).

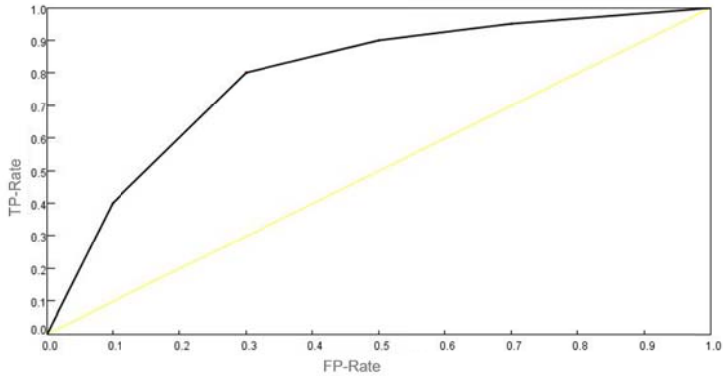
falsch-positiven (*FP*) und der richtig-negativen (*TN*) Zuordnungen sowie die dazugehörigen Raten *FPR* und *TNR*.

Auf Basis der Kontingenztabelle können diverse Gütemaße berechnet werden, die folglich von der Wahl des Schwellenwertes τ abhängen[28]. Das bekannteste statistische Maß ist die Trefferrate, welche sich als Verhältnis richtig zugeordneter Fälle (*TP* + *TN*) zur Gesamtzahl der Testdaten ($n^+ + n^-$) ergibt. Gleichermaßen kann eine ökonomische Bewertung erfolgen, indem die vier Zustände mit den Kosten/Erträgen der durch die Klassifikation ausgelösten Aktionen gewichtet werden[29] oder ein kostenminimierender Schwellenwert τ^* angesetzt wird (Kapitel 2.1).

2.1.2 ROC-Analyse zur Scorecard-Bewertung

Die *Receiver-Operating-Characteristics* (ROC) Analyse[30] vermeidet die Abhängigkeit von einem einzigen Schwellenwert. Stattdessen werden alle möglichen Schwellenwerte in die Betrachtung einbezogen und jeweils ermittelt, welche Klassifikationsleistung – in Form von *TPR*/*FPR*-Paaren – sich bei einem bestimmten Schwellenwert ergeben würde. Hierdurch wird der Trade-Off zwischen der *FPR* und der *TPR* ohne Bezug zu Klassenzugehörigkeits- und Kostenverteilungen analysiert[31], sodass keine Annahmen über die Anwendungsparameter der Scorecard getroffen werden müssen. Die grafische Darstellung der *TPR*/*FPR*-Paare im zweidimensionalen Raum liefert die ROC-Kurve (vgl. *Abbildung 1*).

Abbildung 1: Beispielhafte Darstellung einer ROC-Kurve



Ein Punkt im ROC-Raum dominiert einen anderen, wenn dieser „nordwestlich“ liegt, das heißt eine höhere *TPR* und eine niedrigere *FPR* aufweist. Eine rein zufällige Klassifikation wird durch die 45°-Linie repräsentiert. Eine Scorecard sollte folglich Ergebnisse oberhalb dieses naiven Benchmarks liefern. Diese Bewertungsmethode von Scorecards unterstellt, dass

[28] Für eine umfangreiche Darstellung alternativer Gütemaße sei auf *Hastie/Tibshirani/Friedman* (2009) verwiesen.
[29] Vgl. *Viaene/Dedene* (2004).
[30] Vgl. *Swets* (1988).
[31] Vgl. *Provost/Fawcett* (2001).

keinerlei Informationen über Klassen- und Kostenverteilungen bekannt sind[32], weshalb Annahmen über explizite Verteilungszustände keine notwendige Anwendungsvoraussetzung darstellen. Hierdurch ist eine ökonomische Bewertungsperspektive der Kostenbetrachtung nicht für ROC-Kurven vorgesehen. Erst durch eine nachträgliche Einbindung von konkreten Informationen zu Kosten- und Klassenverteilungen kann mithilfe von Iso-Performanz-Linien eine ökonomische Perspektive eingebunden werden[33].

2.1.3 Eignungsprüfung von Kontingenztabellen und ROC-Analyse

Die Eignungsprüfung eines adäquaten Prognosemaßes sollte an zentrale Entscheidungskriterien gekoppelt sein, die im Wesentlichen den Abgleich zwischen einer ökonomisch-relevanten Bewertungsperspektive, vorzufindenden Informationen zu Anwendungsparametern und dem jeweiligen Prognosemaße durchführt. Basierend auf den obigen Erläuterungen lässt sich eine Klassifizierung der vorgestellten Bewertungsansätze wie folgt durchführen (vgl. *Tabelle 2)*.

Tabelle 2: Eignungsprüfung gängiger Bewertungsansätze unter einer kostenminimierenden Bewertungsperspektive

	Kontingenztabelle	ROC-Kurve
Schiefe Verteilungsstruktur	✗/✓	✓
Konkretisierungsgrad	✗	✗
Instationarität	✗	✗
Informationsgrad	✗	✗

Kontingenztabelle zur Scorecard-Bewertung: Eine vielfach auftauchende Problematik liegt in der spezifischen Auswahl eines auf Kontingenztabellen basierenden Gütemaßes und dessen Berechnungsvoraussetzungen. Explizit sei in diesem Zusammenhang auf das weit verbreitete Prognosemaß der Treffergenauigkeit verwiesen, das auch im Zuge einer ökonomischen Scorecard-Bewertung eingesetzt wird[34]. Hierbei wird implizit unterstellt, dass eine Maximierung dieser Metrik einer ökonomischen Scorecard-Bewertung dienlich ist[35]. Es ist jedoch zu berücksichtigten, dass solche eine Maximierung der Treffergenauigkeit erst dann erreicht wird, wenn alle „guten" Kreditnehmer korrekt klassifiziert werden. Diese Metrik impliziert eine symmetrische Verteilungsstruktur, obgleich die Klasse der „guten"

[32] Vgl. *Fawcett* (2006).
[33] Vgl. *Provost/Fawcett* (2001).
[34] Vgl. z. B. *Abdou* (2009); *Abdou/Pointon/El-Masry* (2008); *Lee/Chen* (2005); *West* (2000).
[35] Vgl. *Hand* (2001).

Antragssteller in den Testdaten prinzipiell überwiegt[36]. Hieraus ergibt sich eine offensichtlich mangelnde Passung zwischen vorausgesetzten und gegebenen Ausprägungen der Anwendungsparameter und der Bewertungszielsetzung. Wird jedoch eine entsprechende Fehlerrate (Klassifikationsfehler im Verhältnis zu Klassifikationsentscheidungen) herangezogen und diese um anfallende Fehlerkosten erweitert[37], so kann eine Berücksichtigung einer asymmetrischen Verteilungsstruktur erfolgen. Eine Erfüllung dieses Kriteriums liefert auch der kostenminimierende Schwellenwert $\tau*$ (siehe Kapitel 2.1). Einschränkend dazu sollte jedoch postuliert werden, dass sich in der Literaturlandschaft nur wenige Beispiele finden lassen, die einen Einsatz der gewichteten Fehlerraten unter der Berücksichtigung von Fehlerkosten oder den kostenminimierenden Schwellenwert vornehmen[38].

Prinzipiell setzt eine ökonomische Bewertung auf Kontingenztabellen basierende Gütemaße voraus, dass präzise, stationäre und sichere Kenntnisse über Parameterverteilung vorliegen. Wie bereits bei der Einführung der Bewertungskriterien diskutiert wurde, werden oftmals diese Voraussetzungen nicht erfüllt. Dies führt bei der Anwendung von auf Kontingenztabellen basierenden Gütemaßen zu nennenswerten Einschränkungen. So folgt aus einem geringen Konkretisierungsgrad der Anwendungsparameter, dass oftmals keine explizite ökonomische Bewertungsperspektive eingenommen werden kann, wie am Beispiel der mit Kosten gewichteten Fehlerrate deutlich wird. Zudem resultiert aus den typischerweise vorliegenden Rahmenbedingungen, dass durch Unsicherheit oder zeitliche Veränderungen der Anwendungsparameter eine endliche Anzahl an Parameterkonstellationen einbezogen werden müsste, wodurch eine endliche Anzahl an wiederholten Berechnungen der Gütemaße vorzunehmen wäre. Dies verdeutlicht die massiv eingeschränkte Anwendungsflexibilität. Ferner ist nicht explizit ableitbar, unter welchen Parameterverteilungszuständen eine Scorecard-Performanz eine andere dominiert. Der Informationsgrad dieser Metriken ist deshalb grundsätzlich stark begrenzt. Somit ist eine kostenorientierte Scorecard-Bewertung auf Kontingenztabellen basierenden Metriken zwar theoretisch durchführbar, jedoch ist diese erheblichen Verzerrungen unterlegen, die die Ergebnisse der Bewertung beeinflussen und damit zu falschen Entscheidungen führen können.

ROC-Analyse zur Scorecard-Bewertung: ROC-Kurven sind unabhängig von konkreten Informationen über Kosten- und Klasseninformationen, wodurch eine ökonomische Bewertungsperspektive der Kostenbetrachtung nicht für ROC-Kurven vorgesehen ist. Erst

[36] Vgl. z. B. *Hand* (2001).
[37] Vgl. *Viaene/Dedene* (2004).
[38] Vgl. z. B. *Abdou* (2009); *Abdou/Pointon/El-Masry* (2008); *Lee/Chen* (2005); *West* (2000).

durch eine nachträgliche Einbindung von konkreten Informationen zu Kosten- und Klassenverteilungen kann mithilfe von Iso-Performanz-Linien eine ökonomische Perspektive eingebunden werden[39].

Eine Operationalisierung dieser Bewertungsperspektive lässt zwar schiefe Verteilungsstrukturen zu, ist allerdings zentralen Schwächen unterlegen. Auch Iso-Performanz-Linien setzten präzise, stationäre und sichere Parameterschätzungen von Kosten- und Klassenverteilungen voraus, für die in analoger Weise zu Kontingenztabellen basierende Gütemaßen dieselben Limitationen gelten. Um Tendenzaussagen abbilden zu können, dass beispielsweise die Kosten einer fälschlichen Kreditgewährung an einen „schlechten" Kunden dem Drei- bis Fünffachen der Opportunitätskosten des umgekehrten Fehlers entsprechen, müsste eine Menge von Iso-Performanz-Linien in den ROC-Raum eingezeichnet werden. Die Darstellung würde hierdurch unübersichtlich. Ferner gilt, dass auch dieses Instrumentarium keinerlei Informationen zur Verfügung stellt, unter welchen Anwendungs-parameterverteilungen eine etwaige Dominanz in der Prognosegüte erzielt wird. Somit ist es grundsätzlich problematisch, dass wichtige ökonomische Informationen wie Fehlerkosten nur durch Iso-Performanz-Linien im ROC-Raum repräsentiert werden, dessen Dimensionen selbst (*TPR*, *FPR*) statistischer Natur und für Entscheidungsträger wenig greifbar sind.

Eine Scorecard-Bewertung gemäß ökonomischer Kriterien würde besser unterstützt werden, wenn der Entscheidungsraum beziehungsweise -rahmen unmittelbar ökonomisch interpretierbar wäre und eine flexible Berücksichtigung unscharfer Informationen über die Anwendungsparameter der Scorecard zuließe. Nichtsdestotrotz liegen zentrale Vorteile für das Kreditwesen in der Robustheit von ROC-Kurven gegenüber Klassen- und Kostenverteilungen vor und können damit typischen Eigenschaften von Anwendungsparametern begegnen.

3 *Kosten-Kurven*

Die vorangegangene Diskussion verdeutlicht, dass die in Kapitel 2 hergeleiteten Eignungskriterien des hier betrachteten Scorecard-Wahlproblems durch gängige Bewertungsinstrumente nur bedingt erfüllt werden. Speziell die Umsetzung einer ökonomischen Scorecard-Bewertung gestaltet sich schwierig. Daher soll in Kapitel 3.1 das Instrument der *Kosten-Kurven* als alternatives Bewertungsinstrument vorgestellt und in Kapitel 3.2 anhand einer Simulationsstudie und den vorgestellten Eignungskriterien auf deren Anwendungseignung für die Antragsbewertung geprüft werden.

[39] Vgl. *Provost/Fawcett* (2001).

3.1 Grundlagen von *Kosten-Kurven*

Kosten-Kurven[40] wurden im maschinellen Lernen entwickelt und stellen ein grafisches Bewertungsinstrument für Scorecards dar, dessen Prognosen für die Ableitung binärer Handlungsoptionen genutzt werden könnten (z. B. Genehmigung eines Kredites oder Verwehrung eines Kredites). Zur Evaluation wird ein sogenannter Kosten-Raum aufgespannt, dessen Dimensionen ökonomisch interpretierbar sind und der gleich gewichtete oder auch ungleich gewichtete Kosten- und Klassenverteilungen unterstützt.

3.1.1 Kosten-Raum

Der Kosten-Raum basiert auf den erwarteten Kosten einer Scorecard, die sich wie folgt darstellen lassen[41].

$$E(Kosten) = FNR \cdot p(+) \cdot C(-|+) + FPR \cdot p(-) \cdot C(+|-) \tag{4}$$

Die erwarteten Kosten setzten sich gemäß (4) aus der Summe der Fehlerraten (*FNR*; *FPR*) zusammen, die durch die *a priori*-Wahrscheinlichkeiten (*p(+); p(-)*) und die Fehlerkosten ($C(-|+); C(+|-)$) gewichtet werden. Es wird davon ausgegangen, dass es sich bei der „positiven" Klasse um die ökonomisch primär relevanten Fälle handelt, zum Beispiel die „schlechten" Risiken in der Antragsbewertung. Zum Zweck einer Normierung kann zusätzlich das Maximum der erwarteten Kosten bestimmt werden. Dieses ergibt sich, wenn alle Objekte falsch klassifiziert werden, sodass $FPR = 1$ und $FNR = 1$ gilt. Durch Einsetzen in (4) ergibt sich[42]:

$$Max(E(Kosten)) = p(+) \cdot C(-|+) + p(-) \cdot C(+|-) \tag{5}$$

Die Division der erwarteten Kosten durch die maximal erwarteten Kosten führt zu den normalisierten erwarteten Kosten:

$$Norm\ (E(Kosten)) = \frac{FNR \cdot p(+) \cdot C(-|+) + FPR \cdot p(-) \cdot C(+|-)}{p(+) \cdot C(-|+) + p(-) \cdot C(+|-)} \tag{6}$$

Eine Vereinfachung der Gleichung kann erzielt werden, indem die Spanne der gewichteten Kosten der „positiven" beziehungsweise „negativen" Klasse im Verhältnis zu der Summe der maximal erwarteten Kosten zusammengefasst wird:

$$PC(+) = \frac{p(+) \cdot C(-|+)}{p(+) \cdot C(-|+) + p(-) \cdot C(+|-)} \qquad PC(-) = \frac{p(-) \cdot C(+|-)}{p(+) \cdot C(-|+) + p(-) \cdot C(+|-)} \tag{7}$$

Basierend auf dieser Zusammenfassung ergeben sich die normalisierten erwarteten Kosten wie folgt:

[40] Vgl. *Drummond/Holte* (2000); *Drummond/Holte* (2006).
[41] Vgl. *Fawcett* (2006); *Drummond/Holte* (2006).
[42] Vgl. *Drummond/Holte* (2000); *Drummond/Holte* (2006).

$$Norm\,(E(Kosten)) = FNR \cdot PC(+) + FPR \cdot PC(-) \quad (8)$$

Die Beziehung $PC(+) + PC(-) = 1$ kann nun genutzt werden, um (8) weiter zu vereinfachen:

$$Norm\,(E(Kosten)) = (FNR - FPR) \cdot PC(+) + FPR \quad (9)$$

Der Kosten-Raum wird anhand der Gleichungen (7) und (9) aufgespannt[43]:

x-Achse: $$PC(+) = \frac{p(+) \cdot C(-|+)}{p(+) \cdot C(-|+) + p(-) \cdot C(+|-)}$$

y-Achse: $$Norm\,(E(Kosten)) = (FNR - FPR) \cdot PC(+) + FPR$$

Die x-Achse des Kosten-Raumes in Form von *PC*(+) lenkt den Fokus der Betrachtung auf ökonomisch relevante Fälle. Dies wird erreicht, indem die Spanne erwarteter (gewichteter) Kosten, die aus der Fehlklassifikation von „schlechten" Risiken resultieren, im Verhältnis zu der Summe der maximal erwarteten (gewichteten) Kosten abgebildet wird. Anhand der Normierung wird bewirkt, dass die Spanne möglicher *PC*(+)-Werte zwischen 0 und 1 liegt[44].

3.1.2 Kosten-Gerade und Kosten-Kurve

In dem aufgespannten Kosten-Raum wird eine sogenannte *Kosten-Gerade* gezeichnet, die auf einem (*FPR*, *FNR*)-Paar und damit auf der Diskretisierung einer Klassifikationsmatrix beruht. Eine Kosten-Gerade bildet damit die Klassifikationsgüte einer Scorecard bei gegebenem Schwellenwert ab. Dagegen repräsentiert die *Kosten-Kurve* die Prognosegüte der Scorecard zu allen möglichen Schwellenwerten. Genau wie bei der ROC-Analyse werden die probabilistischen Scorecard-Prognosen mit jedem Schwellenwert diskretisiert und die relevanten Fehlerraten *FPR* und *FNR* berechnet. Es sei beispielsweise davon ausgegangen, dass eine Scorecard bei sieben verschiedenen Schwellenwerten die folgenden diskreten Ergebnisse liefert (vgl. *Tabelle 3*):

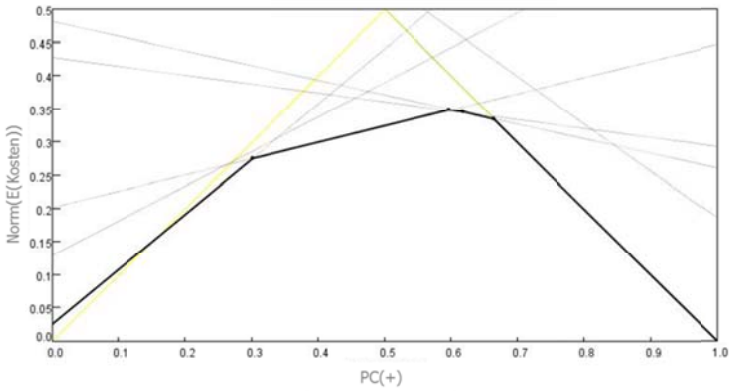
Tabelle 3: *FPR* und *FNR* einer Scorecard zu sieben Schwellenwerten

FPR	***FNR***
0,0240	*0,8587*
0,1298	*0,6522*
0,2019	*0,4457*
0,4279	*0,2935*
0,4808	*0,2609*
0,8990	*0,1848*
1,0000	*0,0000*

[43] Vgl. *Drummond/Holte* (2000); *Drummond/Holte* (2006).
[44] Vgl. *Drummond/Holte* (2000); *Drummond/Holte* (2006).

Die zu *Tabelle 3* gehörigen Kosten-Geraden ergeben sich, indem gemäß (9) für jede Zeile eine Gerade mit der *FPR* als Achsenabschnitt und der Differenz von *FNR* und *FPR* als Steigung entlang aller *PC*(+)-Werte gezogen wird. Jede diskretisierte Scorecard-Prognose lässt sich hierdurch in eine Kosten-Gerade überführen, welche die Klassifikationsleistung der Scorecard über die Spanne aller Kosten- und Klassenverteilungen (*PC*(+)-Werte) hinweg widerspiegelt. Eine beispielhafte Darstellung des Kostenraums für die Scorecard von *Tabelle 3* zeigt *Abbildung 2*.

Abbildung 2: Beispielhafte Darstellung einer *Kosten-Kurve* im Kosten-Raum



Wie anhand von *Abbildung 2* zu sehen ist, liefert die Scorecard eine endliche Anzahl an Klassifikatoren, deren Kosten-Geraden im Kosten-Raum abgebildet werden (graue Linien). Zur Bewertung der Güte der Scorecard, das heißt aller resultierenden Kontingenztabellen, wird die *Kosten-Kurve* als *untere Hüllkurve* aller Kosten-Geraden gebildet (schwarze Linie). Sie verläuft ebenfalls über den gesamten *PC*(+)-Bereich, wobei für jeden x-Wert über alle Kosten-Geraden der geringste, erreichbare y-Wert ermittelt wird.

Alternative Scorecards können anhand ihrer *Kosten-Kurven* verglichen werden. Dabei dominiert eine Scorecard eine andere, wenn ihre *Kosten-Kurve* stringent unterhalb der *Kosten-Kurve* der alternativen Scorecard verläuft, das heißt unter allen denkbaren Ausprägungen der Anwendungsparameter geringere (normalisierte erwartete) Kosten aufweist[45].

3.2 Simulation zur Anwendung von *Kosten-Kurven* im Kreditwesen

Kosten-Kurven weisen zentrale Vorteile auf, die sie besonders für die Bewertung von Scorecards im Kreditwesen empfehlen. Diese Eignung soll im Folgenden anhand einer Simulationsstudie und den in Kapitel 2.2 hergeleiteten Eignungskriterien erläutert werden.

[45] Vgl. *Drummond/Holte* (2000); *Drummond/Holte* (2006).

3.2.1 Grundlagen der Simulationsstudie

Die empirische Untersuchung basiert auf dem Datensatz *German Credit* der *UCI Machine Learning Library*[46]. Dieser wird häufig zur Validierung neuer Klassifikationsverfahren für die Kreditwürdigkeitsprüfung herangezogen[47]. Der Datensatz repräsentiert ein Entscheidungsproblem aus dem Bereich des Antrags-Scorings. Es sollen 1000 Antragsteller in die Klassen hohes/geringes Kreditrisiko eingeteilt werden. Für die Erstellung einer Scorecard stehen zwanzig erklärende Variablen zur Verfügung, die sich unter anderem auf die Kredithöhe, den Kreditzweck sowie die Vermögenssituation der Antragsteller beziehen[48].

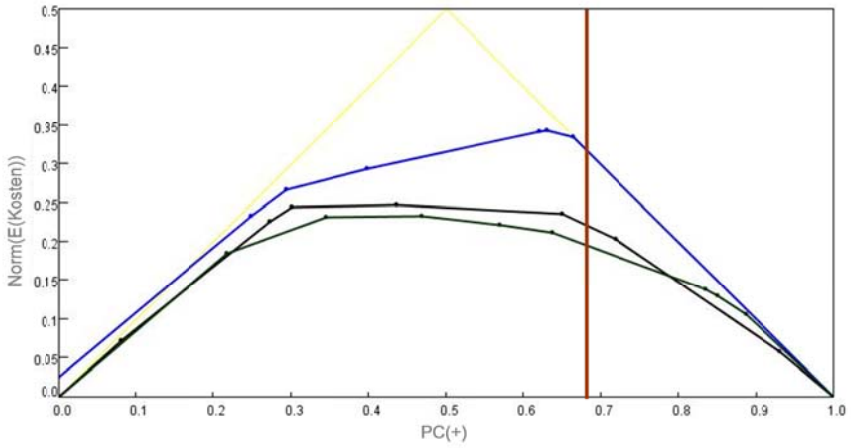
Die Datenmenge wurde im Verhältnis 70:30 in eine Trainings- und Testmenge zur Scorecard-Erstellung und -bewertung aufgeteilt. Um das Prozedere beim Einsatz von *Kosten-Kurven* für die Scorecard-Bewertung beziehungsweise -Auswahl zu illustrieren, wurden beispielhaft drei Klassifikationsverfahren ausgewählt. Es handelt sich dabei um die Logistische Regression, einen C 4.5 Entscheidungsbaum sowie einen Random-Forest-Klassifikator[49]. Die Verfahrenswahl ist für den vorliegenden Beitrag von nachrangiger Bedeutung, da lediglich der Vergleich alternativer Scorecards demonstriert werden soll. In der Praxis könnten die Alternativen zum Beispiel auch aus der Nutzung eines einzigen Klassifikationsverfahrens und verschiedener Merkmalsmengen resultieren.

3.2.2 Eignungsprüfung von Kosten-Kurven

Im Folgenden werden die drei resultierenden Scorecards herangezogen, um Prognosen für die Testfälle zu erzeugen. Anhand dieser Vorhersagen soll mittels *Kosten-Kurven* die für die gegebenen Ausprägungen der Anwendungsparameter bestgeeignete Scorecard identifiziert werden. Die Diskussion dieser korrespondierenden *Kosten-Kurven* soll anhand der eingeführten Eignungskriterien *schiefe Verteilungsstrukturen, Konkretisierungsgrad, Instationarität* und *Informationsgrad* erfolgen. *Abbildung 3* gibt einen Überblick über die in einem Kosten-Raum abgebildeten Scorecards beziehungsweise deren *Kosten-Kurven*:

[46] Vgl. http://archive.ics.uci.edu/ml/datasets/Statlog+%28German+Credit+Data%29.
[47] Vgl. z. B. *Khashman* (2010); *Sinha/May* (2004); *Zhang et al.* (2010).
[48] Vgl. *Baesens et al.* (2003).
[49] Vgl. *Hastie/Tibshirani/Friedman* (2009)

Abbildung 3: Gegenüberstellung der *Kosten-Kurven* alternativer Scorecards (blau: C 4.5 Entscheidungsbaum; schwarz: Logistische Regression; grün: Random-Forest)



Schiefe Verteilungsstrukturen: Es sei zunächst von konkreten Kostenverhältnissen in Höhe von 5:1 ausgegangen[50], wonach die falsche Klassifikation von „schlechten" Risiken um ein 5-faches teurer ist als die Fehlklassifikation eines „guten" Kreditnehmers. Zudem sei eine *a priori*-Wahrscheinlichkeit von 30% für „schlechte" Kreditnehmer unterstellt, sodass beide Anwendungsparameter asymmetrische Verteilungsstrukturen aufweisen. Damit ergibt sich gemäß (9) ein *PC(+)*-Wert in Höhe von 0,6818, der als rote vertikale Linie im Kosten-Raum eingezeichnet ist (vgl. *Abbildung 3*). Die entsprechenden normalisierten erwarteten Kosten für die einzelnen Scorecards ergeben sich durch den Schnittpunkt der vertikalen Linie und den einzelnen *Kosten-Kurven*. Rechnerisch erhält man die normalisierten erwarteten Kosten wie folgt: Gemäß (4) resultieren erwartete Kosten in Höhe von 0,4305 (Random Forest), 0,4848 (Logistische Regression) und 0,7278 (C 4.5 Entscheidungsbaum). Die maximal erwarteten Kosten ergeben nach (5) den Wert 2,2, woraus sich schlussendlich normalisierte erwartete Kosten in Höhe von 0,1957 (Random Forest), 0,2204 (Logistische Regression) und 0,3308 (C 4.5 Entscheidungsbaum) ergeben. Diese Werte würden bedeuten, dass die in der Anwendung erwarteten Fehlerkosten durch den Einsatz der Scorecards gegenüber dem Worst-Case um 1-0,1957= 80,43% (Random Forest,), 1-2204 = 77,96% (Logistische Regression) und 1-0,3308= 66,92% (C 4.5 Entscheidungsbaum) reduziert werden. Eine explizite Bezugnahme zum Worst-Case Szenario kann im Kreditwesen als Form einer konservativen Risikobewertung interpretiert werden. Im Übrigen ist ein wesentlicher Vorteil von *Kosten-Kurven* eben darin zu sehen, dass die Leitungsunterschiede zwischen alternativen Scorecards unmittelbar in ökonomischen Größen gemessen werden. *Kosten-Kurven* können damit die asymmetrischen Verteilungsstrukturen der Anwendungsparameter adäquat in die Scorecard-

[50]Zum besseren Verständnis dieses Wertes sei an die Annahme erinnert, dass richtige Prognosen mit Kosten von Null verbunden sind – die zu richtigen/falschen Entscheidungen korrespondierende Ertrags-/Kosten-Matrix hierzu gegebenenfalls transformiert wurde (vgl. *Viaene/Dedene* (2004)) – und eine Kostenminimierung (durch Maximierung des Abstands zum Worst-Case) somit ökonomisch sinnvoll ist.

Bewertung einbeziehen und stellen damit ein robustes Gütemaß für den vorliegenden Anwendungskontext dar.

Konkretisierungsgrad: Es sei zunächst von dem Beispiel ausgegangen, dass die in den Beispieldaten vorliegende Klassenverteilung nicht der Verteilungsstruktur der eigentlichen Anwendungssituation entspricht und keinerlei Informationen über Fehlerkosten berücksichtigt werden sollen. In diesem Fall werden die *Kosten-Kurven* der drei Scorecards über den gesamten *PC*(+)-Bereich betrachtet. Wie *Abbildung 3* zeigt, dominieren die Random-Forest-Scorecard und die Logistische Regression den C 4.5 Entscheidungsbaum über den gesamten Achsenverlauf. Beide *Kosten-Kurven* verlaufen unterhalb der Kosten-Kurve des C 4.5 Entscheidungsbaumes.

Wird nun von der Annahme ausgegangen, dass Tendenzaussagen über Anwendungsparameter der Scorecard vorliegen (siehe Beispiel *schiefe Verteilungsstrukturen*), so kann ein konkreter Kostenvorteil gegenüber dem C 4.5 Entscheidungsbaum angegeben werden. Bei einem Kostenverhältnis von 5:1 und einer *a priori*-Wahrscheinlichkeit „schlechter" Risiken von 30% ergibt sich, wie bereits erwähnt, ein *PC*(+)-Wert von 0,6818. Ein Vergleich mit dem C 4.5 Entscheidungsbaum ergibt einen relativen Kostenvorteil für die Random-Forest-Scorecard in Höhe von 0,1352 und für die Logistische Regression in Höhe von 0,1105. Damit werden die Kosten aufgrund fehlerhafter Kreditvergabeentscheidungen bei Wahl der besser geeigneten Random-Forest-Scorecard (Logistische Regression) gegenüber des C 4.5 Entscheidungsbaumes um circa 13,52% (11,05%) reduziert. Dies entspricht einer erheblichen Einsparung und unterstreicht, welche Bedeutung der Scorecard-Auswahlprozess für das Credit-Scoring besitzen kann.

Wie dargelegt werden konnte, liegen Einschätzungen über Klassen- und Kostenverteilungen in der Regel nur in eingeschränkter Form vor. Eine Abbildung der Klassifikationsgüte über alle *PC*(+)-Bereiche hinweg, wie sie die x-Achse des Kosten-Raumes zur Verfügung stellt, gewährleistet eine adäquate Berücksichtigung dieser Situation. Durch den aufgespannten Kosten-Raum können gänzlich fehlende Verteilungsannahmen (*PC(+)*-Achse) bis hin zu Tendenzaussagen (*PC(+)*-Wert) der Anwendungsparameter für die Klassifikationsgüte der einzelnen Scorecards unter einer kostenminimierenden Bewertungsperspektive betrachtet werden. Dies stellt einen entscheidenden Vorteil für *Kosten-Kurven* dar.

Instationarität: Auch bei sich ändernden Anwendungsparametern kann die Darstellung der *Kosten-Kurven* im Kosten-Raum aufrecht gehalten werden, da die Klassifikationsleistung über sämtliche *PC(+)*-Werte abgebildet wird und sich bei Verteilungsveränderungen lediglich der betrachtete Bereich auf der x-Achse verschiebt. Würde sich beispielsweise das Kostenverhältnis von 5:1 auf 4:1 ändern, so würde sich ein neuer *PC(+)*-Wert in Höhe von

0,6316 ergeben. Damit verschiebt sich der Fokus der Klassifikationsbewertung nun auf neu ermittelte Schnittpunkte mit den *Kosten-Kurven* der Scorecards (vgl. *Tabelle 4*):

Tabelle 4: Normalisierte erwartete Kosten der Logistischen Regression, Random-Forest-Scorecard und dem C 4.5 Entscheidungsbaum für veränderte Anwendungsparameter

Verteilungsstruktur	***Klassen 30:70***		
	Logistische Regression	*Random-Forest*	*Entscheidungs-Baum*
Kosten 4:1	*0,2365*	*0,2128*	*0,3419*
Kosten 5:1	*0,2204*	*0,1957*	*0,3308*

Wie in *Tabelle 4* zu erkennen ist, liefert die Random-Forest-Scorecard für die aufgezeigten Verteilungskonstellationen geringere (normalisierte erwartete) Kosten als die Scorecard der Logistischen Regression und die des C 4.5 Entscheidungsbaums. Demnach kann, wenn von veränderten Parameterkonstellationen ausgegangen wird, auch für die neue Situation von einer Überlegenheit der Random-Forest-Scorecard ausgegangen werden.

Informationsgrad: Um nun eine finale Auswahl für die überlegene Scorecard treffen zu können, sollten mögliche Unsicherheiten in der Parameterschätzung berücksichtigt werden. Darauf basierend könnte der Entscheidungsträger eine Spanne „plausibler" Anwendungsparameter in Form eines wahrscheinlichen Schwankungsintervalls definieren und den Vergleich der Scorecards auf diesen Bereich eingrenzen. Wird zum Beispiel unterstellt, dass das Kostenverhältnis in der Anwendung zwischen 3:1 und 6:1 und die *a priori*-Wahrscheinlichkeit schlechter Risiken zwischen 20% und 30% liegen wird, dann folgt aus diesen Informationen ein relevanter *PC*(+)-Bereich (vgl. *Tabelle 5*):

Tabelle 5: Mögliches Schwankungsintervall von Anwendungsparameter und daraus folgende *PC(+)*-Werte

Verteilungsstruktur	***Klassen 30:70***	***Klassen 20:80***
Kosten 3:1	*0,5625*	***0,4286***
Kosten 4:1	*0,6316*	*0,5000*
Kosten 5:1	*0,6818*	*0,5556*
Kosten 6:1	***0,7200***	*0,6000*

Für die angenommenen Anwendungsparameter liegt der relevante *PC*(+)-Bereich gemäß *Tabelle 5* zwischen 0,4286 und 0,7200. *Tabelle 6* zeigt die resultierenden (normalisierten erwarteten) Kosten der Random-Forest-Scorecard und der Logistischen Regression auf. Von einer weiteren Untersuchung des C 4.5 Entscheidungsbaumes wird abgesehen, da dieser klar von den beiden anderen Scorecards dominiert wird.

Tabelle 6: Normalisierte erwartete Kosten zweier Scorecards auf Basis der Random-Forest-Scorecard und der Logistischen Regression für eine Menge erwarteter Anwendungsparameter

Verteilungsstruktur	***Klassen 30:70***		***Klassen 20:80***	
	Logistische Regression	*Random-Forest*	*Logistische Regression*	*Random-Forest*
Kosten 3:1	*0,2401*	*0,2219*	*0,2464*	*0,2316*
Kosten 4:1	*0,2365*	*0,2128*	*0,2433*	*0,2288*
Kosten 5:1	*0,2204*	*0,1957*	*0,2404*	*0,2227*
Kosten 6:1	*0,2017*	*0,1813*	*0,2381*	*0,2171*

Es kann nun der relevante *PC(+)*-Abschnitt der *Kosten-Kurven* beider Scorecards analysiert werden, um die relative Dominanz beziehungsweise Vorteilhaftigkeit des Scorecard-Einsatzes genauer zu quantifizieren. Im relevanten *PC*(+)-Bereich ergibt sich ein (maximaler) Kostenvorteil für die Random-Forest-Scorecard in Höhe von 0,0247 gegenüber der Logistischen Regression. Demnach können, unter den vom Entscheidungsträger als plausibel betrachteten Anwendungsparametern, die Kosten falscher Entscheidungen durch die Scorecard des Random Forest im Vergleich zur Logistischen Regression um maximal 2,47% gesenkt werden.

Über die verschiedenen Eignungskriterien kann somit dargelegt werden, dass die Random-Forest-Scorecard eine konsequent bessere Klassifikationsleistung erbringt als der C 4.5 Entscheidungsbaum und die Logistische Regression. Dies verdeutlicht, dass *Kosten-Kurven* einen Informationsgewinn ermöglichen, indem die Scorecard-Güte unmittelbar für verschiedene Anwendungsszenarien betrachtet werden kann und zudem Informationen darüber liefert, unter welchen *PC(+)*-Werten eine Scorecard eine andere dominiert.

3.3 Betriebswirtschaftliche Würdigung von *Kosten-Kurven*

Der sinnvolle Einsatz von *Kosten-Kurven* hängt im Wesentlichen von der Fragestellung ab, inwiefern ein (ökonomischer) Mehrwert durch *Kosten-Kurven* generiert werden kann. In diesem Zusammenhang sollte sich der Anwender vor Augen führen, dass sich jedweder Vorteil aus der Auswahl der bestgeeigneten Scorecard und nicht unmittelbar aus der Anwendung von *Kosten-Kurven* ergibt. Der Einsatz von *Kosten-Kurven* weist jedoch dahingehend einen unmittelbaren Vorteil auf, indem *Kosten-Kurven* Schwächen (wie in Kapitel 2 erläutert) von gebräuchlichen Bewertungsinstrumenten ausgleichen und eine Auswahlentscheidung unter ökonomischen Gesichtspunkten erleichtern beziehungsweise erst ermöglichen.

In der Simulationsstudie ergab sich ein Einsparungspotenzial von circa 13,52% durch die Wahl der besser geeigneten Scorecard Random-Forest im Vergleich zur schlechtesten

Alternative (beziehungsweise von circa 80,43% im Vergleich zum Worst-Case Szenario). Dieser Wert hängt allerdings von den unterstellten Ausprägungen der Anwendungsparameter (und anderen Faktoren) ab und kann somit nicht verallgemeinert werden. Vielmehr muss das Einsparungspotenzial unter Berücksichtigung der eingesetzten Methoden (Random Forest im Vergleich zur Logistischen Regression) als auch der Ausprägungen der Anwendungsparameter (Kostenverhältnis 5:1 und Klassenverteilung 30:70) betrachtet werden. Die Simulationsstudie lässt den Schluss zu, dass durch *Kosten-Kurven* die Entscheidungsqualität bei Scorecard-Wahlentscheidungen tendenziell verbessert werden kann. Dieser Schluss begründet sich insbesondere aus der Möglichkeit, beliebige Informationsstände des Entscheidungsträgers flexibel in den Bewertungsprozess zu integrieren (siehe *Konkretisierungsgrad*). Es kann also angenommen werden, dass bei Nutzung von *Kosten-Kurven* die für eine gegebene Anwendung bestgeeignete Scorecard öfter gewählt werden dürfte.

Ein weiterer Vorteil der *Kosten-Kurven*-Methodik ist darin zu sehen, dass dem Entscheidungsträger zentrale Hinweise gegeben werden, ob eine Scorecard unter bestimmten Ausprägungen an Anwendungsparametern überhaupt eingesetzt werden darf (siehe *Informationsgrad*). Auch hier bieten *Kosten-Kurven* einen wichtigen Erkenntnisgewinn. Vor diesem Hintergrund erscheint es zulässig davon auszugehen, dass der Einsatz von *Kosten-Kurven* insgesamt zu einer höheren Qualität der gewählten und eingesetzten Scorecards führt und sich hieraus ein substanzieller ökonomischer Mehrwert ergeben kann. Wie groß dieser Vorteil ist, hängt letztendlich von der Anzahl der Scorecard-Wahlentscheidungen und damit der Größe des Finanzdienstleisters ab. Beispielsweise dürften speziell Großunternehmen von *Kosten-Kurven* profitieren, da insbesondere hier häufig neue Scoring-Anwendungen beziehungsweise regelmäßige Aktualisierungen der eingesetzten Modelle erforderlich sind. Folglich müssen viele Auswahlentscheidungen getroffen werden, aus denen ein systematischer Vorteil durch den Einsatz von *Kosten-Kurven* resultieren könnte. Ferner bieten *Kosten-Kurven* die Möglichkeit, zeitliche Veränderungen der Anwendungsparameter dahingehend zu berücksichtigen, dass zwar der Fokus der (Klassifikations-) Betrachtung verschoben wird, die eigentliche Darstellung der Prognosegüte aber beibehalten werden kann (siehe *Instationarität*). So führen Änderungen der Anwendungsparameter einer Scorecard lediglich zu einer Verschiebung eines relevanten *PC*(+)-Bereichs.

Obigen Vorteilen steht der notwendige Implementierungsaufwand zur Anwendung von *Kosten-Kurven* gegenüber. Die *Kosten-Kurven*-Methodik ist im Rahmen einer freien, Java-basierten Software öffentlich zugänglich. Implementierungsaufwendungen beschränken sich somit auf die Integration der Methodik in die Unternehmens-IT sowie die Schulung von

Mitarbeitern. Auch hier gilt, dass letztendlich die Unternehmensgröße beziehungsweise Anzahl an Auswahlentscheidungen determiniert, wie schnell sich ein positiver ROI ergeben würde. Speziell im Kreditwesen ist es darüber hinaus bedeutsam, dass jegliche Planungsinstrumente regulatorisch akzeptabel sind und die Anforderungen bestehender Verordnungen erfüllen. *Kosten-Kurven* sind auch unter diesem Gesichtspunkt sehr gut für einen Einsatz in der Finanzdienstleistungsindustrie geeignet. In der als *Basel II* bekannten Rahmenverordnung für das Risikomanagement von Finanzprodukten wird explizit auf die ROC-Analyse als sinnvolles Bewertungsinstrument hingewiesen[51]. *Kosten-Kurven* besitzen gegenüber der ROC-Analyse den Vorteil, dass der Kosten-Raum ökonomisch interpretierbar ist und verschiedene Informationsstände bezüglich vorliegender Anwendungsparameter sehr viel intuitiver berücksichtigt werden können. Mathematisch sind beide Techniken äquivalent[52], sodass die regulatorische Konformität der ROC-Analyse auch den Einsatz von *Kosten-Kurven* abdecken sollte.

4 Zusammenfassung und Ausblick

Scorecards werden im Finanzdienstleistungsbereich vielfach eingesetzt, um verschiedene Entscheidungsprozesse in Marketing und Risikomanagement zu unterstützen. Im Mittelpunkt des Beitrags steht die Frage, wie Entscheidungsträger vorgehen sollten, um zwischen alternativen Scorecards zu wählen. Die mittels einer Scorecard generierten Prognosen besitzen einen erheblichen ökonomischen Wert, so dass es von großer Wichtigkeit ist, für ein gegebenes Entscheidungsproblem, unter Berücksichtigung aller relevanten Rahmen- und Anwendungsbedingungen, eine bestgeeignete Scorecard zu identifizieren. Der Beitrag entwickelt zunächst einen Kriterienkatalog, der diese Anforderungen zusammenfasst. Eine Analyse gängiger Bewertungspraktiken verdeutlicht, dass die vornehmlich eingesetzten statistischen Güteindikatoren nur eingeschränkt für Anwendungen im Kreditwesen geeignet sind. *Kosten-Kurven* repräsentieren an dieser Stelle ein überlegenes Instrument. Eine analytische und empirische Betrachtung dieser neuen Bewertungsmethode dokumentiert, dass die für das Kreditgeschäft typischen Anforderungen deutlich besser erfüllt werden.

Für die unternehmerische Praxis liefert der vorliegende Beitrag einen Erkenntnisgewinn, indem anhand eines konkreten Anwendungsbeispiels verdeutlicht wird, wie *Kosten-Kurven* im Kreditwesen genutzt werden können und welche Vorteile sie gegenüber gebräuchlichen Alternativen besitzen. Zum einen erlaubt die ökonomische Bewertung anhand erwarteter Fehlerkosten eine anwendungsorientierte Betrachtungsperspektive, mit der ein erheblicher

[51] Vgl. *Basel Committee on Banking Supervision* (2005).
[52] Vgl. *Drummond/Holte* (2006).

Informationsgewinn gegenüber statistischen Gütemaßen einhergeht. Zum anderen erweisen sich *Kosten-Kurven* als äußerst flexibel bei der Berücksichtigung von Kenntnissen über die konkrete Entscheidungssituation. Liegen keinerlei Informationen über die in der Anwendungsdomäne erwarteten Kosten- und Klassenverteilungen vor, können alternative Modelle über alle denkbaren Rahmenbedingungen hinweg kontrastiert werden. Gleichermaßen lassen sich spezifische und semistrukturierte Erwartungen in die Bewertung integrieren, indem ein relevanter *PC(+)*-Bereich ermittelt und die Betrachtungsperspektive auf diesen begrenzt wird. Mit diesen Eigenschaften ermöglichen *Kosten-Kurven* eine transparente, effektive und robuste Bewertung alternativer Scorecards, wie sie für die Abschätzung von Kreditrisiken unabdingbar ist.

Ferner ergibt sich ein wissenschaftlicher Erkenntnisgewinn, indem konzeptionelle Schwächen gebräuchlicher Gütemaße zur Erhebung der Prognosegüte von Scorecards im Credit-Scoring offenbart werden. Hierdurch wird die gängige Praxis, eben diese Gütemaße durch immer neue Klassifikationsverfahren zu optimieren, in Frage gestellt. Ferner wird mit den *Kosten-Kurven* auf ein überlegenes Instrumentarium zur Scorecard-Bewertung hingewiesen. Insofern liefert der Beitrag wertvolle Impulse für das Design neuer Klassifikationsverfahren. Beispielsweise legt die anwendungsbezogene Visualisierung der Prognosegüte im Kostenraum nahe, dass die Fläche unter einer Kosten-Kurve eine sinnvolle Zielgröße darstellt, die im Rahmen der Modellerstellung minimiert werden könnte. Zukünftige Arbeiten könnten sich folglich mit dem Entwurf und der Evaluation von Verfahren befassen, die diese Zielsetzung verfolgen.

Literatur

Abdou, Hussein (2009), Genetic programming for credit scoring: The case of Egyptian public sector banks, in: Expert Systems with Applications, 36, 9, S. 11402–11417.

Abdou, Hussein/Pointon, John/El-Masry, Ahmed (2008), Neural nets versus conventional techniques in credit scoring in Egyptian banking, in: Expert Systems with Applications 35, 3, S. 1275–1292.

Adams, Niall M./Hand, David J. (1999), Comparing classifiers when the misallocation costs are uncertain, in: Pattern Recognition, 32, 7, S. 1139-1147.

Baesens, Bart/Gestel, Tony van (2009), Credit Risk Management, Oxford University Press, Oxford.

Baesens, Bart/Setiono, Rudy/Mues, Christophe/Vanthienen, Jan (2003), Using neural network rule extraction and decision tables for credit-risk evaluation, in: Management Science, 49, 3, S. 312-329.

Banasik, John /Crook, Jonathan/Thomas, Lyn (2003), Sample selection bias in credit scoring models, in: Journal of the Operational Research Society 54, 8, S. 822–832.

Basel Committee on Banking Supervision (2005), Studies on the validation of internal rating systems, Bank for International Settlements, Working Paper 14, http://www.bis.org/publ/bcbs_wp14.htm.

Burez, Jonathan/Van den Poel, Dirk (2007), CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services, in: Expert Systems with Applications, 32, 2, S. 277-288.

Drummond, Chris/Holte, Robert C. (2000), Explicitly representing expected cost: An alternative to ROC representation, 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, U.S.A.

Drummond, Chris/Holte, Robert C. (2006), Cost curves: An improved method for visualizing classifier performance, in: Machine Learning, 65, 1, S. 95-130.

Duda, Richard O./Hart, Peter E./Stork, David G. (2001), Pattern Classification, Wiley, New York.

Fawcett, Tom (2006), An introduction to ROC analysis: ROC analysis in pattern recognition, in: Pattern Recognition Letters, 27, 8, S. 861-874.

Hand, David J. (2001), Modelling consumer credit risk, in: IMA Journal of Management Mathematics, 12, 2, S. 139-155.

Hand, David J. (2005), Good practice in retail credit scorecard assessment, in: Journal of the Operational Research Society, 56, 9, S. 1109-1117.

Hartmann-Wendels, Thomas/Pfingsten, Andreas/Weber, Martin (2010), Bankbetriebslehre, Springer, Berlin.

Hastie, Trevor/Tibshirani, Robert/Friedman, Jerome H. (2009), The Elements of Statistical Learning, Springer, New York.

Hofmann, Gerhard (2007), Basel II und MaRisk, Bankakademie-Verlag, Frankfurt a.M.

Khashman, Adnan (2010), Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes, in: Expert Systems with Applications, 37, 9, S. 6233-6239.

Larivière, Bart/Van den Poel, Dirk (2005), Predicting customer retention and profitability by using random forests and regression forests techniques, in: Expert Systems with Applications, 29, 2, S. 472-484.

Lee, Tian-Shyug/Chen, I-Fel (2005), A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines, in: Expert Systems with Applications 28, 4, S. 743–752.

Loterman, Gert/Brown, Iain/Martens, David/Mues, Christophe/Baesens, Bart (2011), Benchmarking regression algorithms for loss given default modeling, in: International Journal of Forecasting, 28, 1, S. 161-170.

Martens, David/Baesens, Bart/van Gestel, Tony/Vanthienen, Jan (2007), Comprehensible credit scoring models using rule extraction from support vector machines, in: European Journal of Operational Research, 183, 3, S. 1466-1476.

Matuszyk, Anna/Mues, Christophe/Thomas, Lyn C. (2009), Modelling LGD for unsecured personal loans: Decision tree approach, in: Journal of the Operational Research Society, 61, 3, S. 393-398.

Provost, Foster/Fawcett, Tom (2001), Robust classification for imprecise environments, in: Machine Learning, 42, 3, S. 203-231.

Provost, Foster, J./Fawcett, Tom/Kohavi, Ron (1998), The Case against Accuracy Estimation for Comparing Induction Algorithms, Proceedings of the Fifteenth International Conference on Machine Learning, San Francisco, CA, U.S.A.

Reichheld, Frederick F./Sasser, W. Earl (1990), Zero defections: Quality comes to service, in: Havard Business Review, 68, 5, S. 105-111.

Sinha, Atish P./May, Jerrold H. (2004), Evaluating and tuning predictive data mining models using receiver operating characteristic curves, in: Journal of Management Information Systems, 21, 3, S. 249-280.

Swets, John A. (1988), Measuring the accuracy of diagnostic systems, in: Science, 240, 4857, S. 1285-1293.

Thomas, Lyn C. (2000), A survey of credit and behavioural scoring: Forecasting financial risk of lending to consumers, in: International Journal of Forecasting, 16, 2, S. 149-172.

Thomas, Lyn C./Edelman, David B./Crook, Jonathan N. (2002), Credit Scoring and its Applications, Siam, Philadelphia.

Thomas, Lyn C./Oliver, Richard W./Hand, David J. (2005), A survey of the issues in consumer credit modelling research, in: Journal of the Operational Research Society, 56, 9, S. 1006- 1015.

Van den Poel, Dirk/Lariviere, Bart (2004), Customer attrition analysis for financial services using proportional hazard models, in: European Journal of Operational Research, 157, 1, S. 196-217.

Viaene, Stijn/Dedene, Guido (2004), Cost-sensitive learning and decision making revisited, in: European Journal of Operational Research, 166, 1, S. 212-220.

West, David (2000), Neural network credit scoring models, in: Computers & Operations Research, 27, 11-12, S. 1131-1152.

Zhang, Defu/Zhou, Xiyue/Leung, Stephen C. H./Zheng, Jiemin (2010), Vertical bagging decision trees model for credit scoring, in: Expert Systems with Applications, 37, 12, S. 7838-7843.

Summary:

Credit scorecards are routinely used in the financial service industry to guide decision making in marketing and risk management. The paper is concerned with the problem of identifying an appropriate scorecard among a set of alternatives. To that end, a requirement specification for scorecard assessment in the credit industry is developed. Examining the compliance of current assessment practices with these requirements, the authors find that standard performance measures suffer important limitations. The cost-curve methodology is introduced as a more powerful tool for scorecard selection. Its unique advantages are illustrated by means of an empirical study. A key implication of the paper is that Cost Curves facilitate a business oriented scorecard selection and, thereby, contribute toward increasing decision quality in scorecard-supported business processes.

III. Crowdsourcing: Systematisierung praktischer Ausprägungen und verwandter Konzepte

Authors: **Nicole Martin**

Stefan Lessmann

Stefan Voß

Year: *2008*

Bibliographic Status: *Published,*

Proceedings of the Multikonferenz Wirtschaftsinformatik, Munich, Germany

Crowdsourcing: Systematisierung praktischer Ausprägungen und verwandter Konzepte

Nicole Martin, Stefan Lessmann, Stefan Voß

Institut für Wirtschaftsinformatik
Universität Hamburg
Von-Melle-Park 5, 20146 Hamburg
nicole_martin1@yahoo.de,
lessmann@econ.uni-hamburg.de
stefan.voss@uni-hamburg.de

Abstract: Die Arbeit betrachtet das Crowdsourcing als ein aktuell diskutiertes Konzept für die Organisation eines überbetrieblichen, interaktiven Leistungsaustauschs auf der Basis von Web 2.0. In der wissenschaftlichen Literatur wurde dieser Ansatz bisher wenig beachtet, wohingegen sich in der betrieblichen Praxis bereits einige, z. T. aber stark unterschiedliche „Crowdsourcing Plattformen" finden. In Ermangelung eines allgemeinen Begriffsverständnisses ist es das Ziel der vorliegenden Arbeit, das Crowdsourcing Konzept zu systematisieren. Dazu werden ein Definitionsansatz sowie ein Klassifikationsschema vorgeschlagen, welche aus der Analyse bestehender Crowdsourcing Formen und angrenzender theoretischer Konzepte abgeleitet werden.

1 Einleitung

Globalisierung und Deregulierung haben zu einer spürbaren Wettbewerbsverschärfung geführt und zwingen Unternehmen, neue Gestaltungen der wirtschaftlichen Realität zu entdecken [Fe07]. Ein Aspekt ist dabei die effiziente Ausgestaltung der Geschäftsprozesse entlang der Wertschöpfungskette, z. B. zum Erreichen einer Kostenführerschaft [Po80]. Dies beinhaltet in zunehmendem Maße auch eine Öffnung der Unternehmensgrenzen, das heißt eine aktive Einbeziehung externer Partner in den Leistungserstellungsprozess, um strategische Wettbewerbsvorteile zu erschließen (siehe auch [HS06]). Dabei schaffen moderne Informations- und Kommunikationssysteme (IuK-Systeme) die technischen Voraussetzungen für eine unternehmensübergreifende Daten-, System- und ggf. Prozessintegration, wie sie beispielsweise im Supply Chain Management oder dem Collaborative Planning, Forecasting and Replenishment vorgesehen ist. In ähnlicher Weise erlauben Kollaborationssysteme eine – zumeist asynchrone – Zusammenarbeit räumlich/geografisch verteilter Personen und leisten ebenfalls einen wichtigen Beitrag zur Ermöglichung interaktiver Wertschöpfungsformen.
Im Mittelpunkt der Arbeit steht das sog. Crowdsourcing, welches den Grundgedanken einer kollektiven Interaktion aufgreift und eine Nutzbarmachung von Wissen und/oder Arbeitskraft einer großen Anzahl externer Partner für die Leistungserstellung anstrebt. Plakativ wird in diesem Zusammenhang auch von einer Erschließung der „Weisheit der Vielen“ [Su04] gesprochen. Crowdsourcing steht dabei in engem Zusammenhang mit dem Web 2.0 [O'R05], welches in Anlehnung an [RKK07] als eine Kombination aus neuen Techniken und Anwendungstypen sowie einer sozialen Bewegung und neuen Geschäftsmodellen verstanden werden kann. Das heisst Crowdsourcing repräsentiert eine auf dem Web 2.0 basierende Wertschöpfungsform.
In der wissenschaftlichen Literatur wurde das Crowdsourcing Konzept bisher wenig beachtet, wohingegen sich in der betrieblichen Praxis bereits zahlreiche Umsetzungen finden lassen. Obgleich sich insg. ein sehr heterogenes Bild ergibt, ist sämtlichen Crowdsourcing Plattformen gemein, dass sie ein asynchrones Zusammenwirken dislozierter Individuen durch Einsatz web-basierter Technologien unterstützen bzw. ermöglichen. Folglich besteht eine konzeptionelle Ähnlichkeit zwischen Crowdsourcing Plattformen und klassischen Groupware-Systemen. Wie zu zeigen sein wird, wird das bisherige Verständnis computergestützter Gruppenarbeit beim Crowdsourcing allerdings u. a. dahingehend erweitert, dass die Partizipieren nicht notwendigerweise kollaborativ agieren, sondern oftmals in einer Wettbewerbsbeziehung stehen.
In Ermangelung einer systematischen Aufarbeitung des Crowdsourcing Konzeptes in der wissenschaftlichen Literatur ist es das Ziel der vorliegenden Arbeit, einen Definitionsansatz zu entwickeln und einen Beitrag zur Konzeptionalisierung dieses neuen Forschungsgebietes zu leisten. Im Mittelpunkt steht dabei eine adäquate Berücksichtigung von Begriffsinhalten, die sich aus bestehenden Crowdsourcing Umsetzungen ergeben. Gleichermaßen bedeutsam ist eine Analyse der Beziehungen zwischen Crowdsourcing und verwandten Ansätzen, welche sich ebenfalls mit dem interaktiven Zusammenwirken mehrerer Individuen zu Wertschöpfungszwecken befassen. Die Arbeit soll damit auch aufzeigen, welche neuen Formen computergestützter Zusammenarbeit existieren, in wie weit diese eine Erweiterung des bisherigen Verständnis von Kooperationssystemen notwendig machen und welche Anforderungen sich daraus für entsprechende Anwendungssysteme ergeben.
Der Aufbau der Arbeit gestaltet sich wie folgt: In Kapitel 2 werden zunächst bestehende Crowdsourcing Plattformen katalogisiert. Darauf aufbauend wird in Kapitel 3 ein Definitionsansatz entwickelt. Diese empirische Perspektive wird dann durch eine Gegenüberstellung mit angrenzenden Ansätzen, erweitert, um das induzierte Begriffsverständnis zu validieren und die Berechtigung einer eigenständigen Begriffsbildung zu verifizieren. Darauf aufbauend wird in Kapitel 4 ein Klassifikationsschema für die

derzeitigen Formen von Crowdsourcing entwickelt. Kapitel 5 fasst die Ergebnisse der Arbeit zusammen und zeigt weiterführende Forschungsbedarfe auf.

2. Crowdsourcing in der betrieblichen Praxis

Dieses Kapitel dokumentiert bestehende Plattformen, um die Vielfältigkeit von auf Crowdsourcing basierenden Geschäftsmodellen aufzuzeigen. Die Plattformen wurden anhand ihrer Größe (gemessen in der Anzahl der Mitglieder) und ihrer Medienpräsenz in mit Crowdsourcing assoziierten Zeitungsartikeln sowie Internetforen bzw. -blogs ausgewählt:
InnoCentive[1] ist ein Portal, welches für Unternehmen, die spezielle F&E Lösungen suchen, und Wissenschaftler bzw. Experten, welche ihr Wissen und ihre Erfahrung zur Verfügung stellen möchten, als Vermittlungsplattform dient. Demnach nimmt InnoCentive eine Mediatorenrolle ein und ermöglicht es dem lösungssuchenden Unternehmen, seine Fragestellung zielgerichtet, ähnlich einer Ausschreibung, zu publizieren. Die Bearbeitung der Aufgabenstellung erfolgt i.d.R. parallel durch mehrere Interessenten, wobei nur die aus Unternehmenssicht beste Lösung mit einer Geldprämie entlohnt wird (wettbewerbsorientierte Bearbeitung). Die Aufgabenstellungen zielen zumeist auf den Entwurf und/oder die Entwicklung einer – zumindest aus Unternehmensperspektive – innovativen Lösung ab.[2]
NineSigma[3] ist eine Plattform, welche, ähnlich wie InnoCentive, Problemstellungen mit Innovationscharakter bearbeiten lässt und Spezialisten einbezieht, die in einem direkten Wettbewerb miteinander konkurrieren. Allerdings verfolgt NineSigma neben der Ausschreibung von Unternehmensaufgabenstellungen auch das Ziel, die richtigen Adressaten ausfindig zu machen; beispielsweise Personen, die eine ähnliche Problemstellung bereits in anderem Kontext bearbeitet haben. Dementsprechend wird hier neben der Mittlerrolle auch eine Koordinationsfunktion ausgeübt. Wesentliche Motivation für die Teilnahme sind erneut monetäre Anreize.
Das Konzept von **Cambrian House**[4] repräsentiert eine Richtung des Crowdsourcing, in dem die Gruppe der Partizipierenden gleichzeitig die Rolle des Auftraggebers und des Problemlösers übernimmt. Die Grundidee besteht darin, einerseits Bedarfe proaktiv zu identifizieren und andererseits Individuen zusammenzubringen, welche diese in Form entsprechender Produktinnovationen gemeinschaftlich kommerzialisieren können. Demnach basiert Cambrian House auf der Idee einer Nischenstrategie [Po80] und bietet Unterstützungs- bzw. Vermittlungsleistungen an, eine Solche zu verwirklichen. Die Leistungserstellung ist grundsätzlich kollaborativ organisiert, wobei auf das Know-how von Spezialisten zurückgegriffen wird, die Fachkenntnis der beteiligten Individuen aber insg. heterogener ist als bei InnoCentive/Ninesigma. Trotz dieser betriebswirtschaftlich orientierten Grundausrichtung ist zu beobachten, dass bei den Mitgliedern auch intrinsische Motivationsanreize relevant sind.
Marketocracy[5] ist eine Plattform, welche die Evaluation von Investmentstrategien zum Ziel hat. Dazu werden im Rahmen einer virtuellen Börse die Investitionstätigkeiten von Akteuren analysiert und ausgewertet. Den Teilnehmern werden virtuelle 1 Mio. USD für ihre Transaktionen zur Verfügung gestellt. Die erfolgreichsten Investmentstrategien fließen dem Marketocracy Capital Management, einem Investmentberater für (reale) Fonds, zu. Für diejenigen Strategien, die in realen Fonds übernommen werden, wird der virtuelle Investor finanziell entlohnt. Da diese Perspektive de facto nur für Teilnehmer besteht, die über gewisse Vorkenntnisse im Bereich Anlagestrategien und Spekulationsgeschäfte verfügen, werden auch

[1] http://www.innocentive.com/
[2] Vgl. zu den verschiedenen Dimensionen des Innovationsbegriffs auch [VB00].
[3] http://www.ninesigma.com/
[4] http://www.cambrianhouse.com/
[5] http://www.marketocracy.com/

hier eher Spezialisten angesprochen. Diese erarbeiten Anlagestrategien parallel und wettbewerbsorientiert auf der Basis monetärer Anreize.

Rent a Coder[6] stellt einen Marktplatz dar, auf dem Unternehmen oder Einzelpersonen Programmierer suchen und einstellen können. Anfragen können direkt an die Plattform gestellt werden, welche dann deren Distribution an einen Pool von Entwicklern (über 150.000 weltweit) übernimmt. Diese unterbreiten dem Auftraggeber jeweils ein Angebot, aus denen das Unternehmen auswählen kann. Rent a Coder übernimmt hierbei eine Vermittlungsposition, wobei die Teilnehmer rein extrinsisch motiviert sind.

Trendwatching[7] hat über 8.000 so genannter „Trend Spotter", die über innovative Trends und Geschäftsideen in ihrem Land berichten. Diese Berichte werden kostenlos oder auch kostenpflichtig angeboten. Die „Spotter" erhalten für ihre Recherchen Leistungspunkte, die gegen Sachprämien eingetauscht werden können. Eine finanzielle Entlohnung ist ebenfalls möglich. Da die Aufgaben einer Journalistentätigkeit gleichkommen, sind die Partizipierenden auch hier eher spezialisiert und primär extrinsisch motiviert.

Die Plattform **iStockphoto**[8] ist eine Art Marktplatz und stellt eine große Sammlung von professionellen Fotografien für einen geringen Preis (< 1 USD) zur Verfügung. Dabei können sowohl professionelle Fotografen als auch Hobbyisten ihre Arbeiten zur Verfügung stellen. Ein Hauptanreiz für die Teilnehmer besteht in der internen „Community-Dynamik", die die Mitglieder zum regen Austausch und zur eigenen Fortbildung nutzen. Allerdings ist der extrinsische Anreiz nicht vollständig zu vernachlässigen, da Plattform-Teilnehmer auch hier (geringfügig) entlohnt werden.

Threadless[9] ist eine Plattform für Künstler oder kreative Laien, die T-Shirt Designs entwerfen und die Möglichkeit besitzen, diese auf der Threadless-Seite zu veröffentlichen und anschließend von der Community bewerten zu lassen. Die best bewerteten Entwürfe werden anhand einer monetären Prämie entlohnt, gehen dann in Produktion und stehen anschließend jedem zum Kauf zur Verfügung. Darüber hinaus übernehmen die Teilnehmer die Werbung der T-Shirts und sind für Katalogfotos verantwortlich. Somit ist die Community vom Entwurf bis hin zur Distribution der T-Shirts involviert.

Bei **John Fluevog**[10] handelt es sich um einen Schuhhändler und Produzenten, der eine Plattform zur Verfügung stellt, um sich die Kreativität eines jeden Teilnehmers, vom Laien bis zum Künstler, für neue Designs zu Nutze zu machen. Analog zu Threadless werden Designvorschläge gesucht und anschließend von anderen Teilnehmern bewertet. Die Namen der Gewinner werden auf die gefertigten Schuhe gedruckt, wobei keine weitere monetäre Entlohnung für die geleisteten Designvorschläge erfolgt (intrinsische Motivation).

Mechanical Turk[11] ist ein von Amazon initiiertes Netzwerk, das bei der Lösung von routinemäßigen Aufgaben helfen soll, welche bisher nicht oder nur begrenzt maschinell gelöst werden können, beispielsweise die Identifizierung von Objekten auf Fotografien. Die gestellten Aufgaben werden hierbei als HIT (Human Intelligence Task) bezeichnet und an die Teilnehmer ausgeschrieben, welche für die Bearbeitung bezahlt werden. Diese Plattform könnte auch als Marktplatz für „Mikro-Dienstleistungen" verstanden werden, bezieht Amateure ein und vergibt ausschließlich Aufgabenstellungen, die keinen Innovationscharakter besitzen.

[6] http://www.rentacoder.com/
[7] http://www.trendwatching.com/
[8] http://www.istockphoto.com/
[9] http://www.threadless.com/
[10] http://www.fluevog.com/
[11] http://www.mturk.com/mturk/welcome

3. Konzeptualisierung von Crowdsourcing

Im Folgenden sollen die Unzulänglichkeiten des derzeitigen Crowdsourcing Begriffs aufgezeigt und ein eigener Definitionsansatz vorgestellt werden. Dieser wird nachfolgend verwandten Konzepten zur interaktiven Leistungserstellung gegenübergestellt, um die wesentlichen Unterschiede hervorzuheben.

3.1 Derzeitiges Begriffsverständnis und Definition

Der Begriff Crowdsourcing ist in [Ho06] erstmals aufgeführt worden und wird dort als „(…) the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call" definiert.
Vor dem Hintergrund von Kapitel 2 wird unmittelbar deutlich, dass dieser Definitionsansatz die praktische Realität nur unzureichend abbildet. Zum einen ist die Fokussierung auf die Unternehmensperspektive im Sinne einer Weiterentwicklung oder Ergänzung von Outsourcing nicht mit Ansätzen wie Cambrian House oder iStockphoto kompatibel. Diese dokumentieren vielmehr, dass keinesfalls ein unternehmens- bzw. auftraggeberinitiierter Tätigkeitsanstoß vorliegen muss. Anstatt dessen entwickelt die (virtuelle) Community selbsttätig und von sich heraus Dienstleistungen und Produkte, die zur wirtschaftlichen Nutzung angeboten werden. Ferner wird der dem Crowdsourcing inhärente Bezug zur IuK-Technologie in der Definition von [Ho06] komplett vernachlässigt, obwohl der Informations- und/oder Leistungsaustausch bei allen Plattformen durchgängig webbasiert bzw. auf Basis von Web 2.0 erfolgt. Da ein interaktives Zusammenwirken vieler, geografisch verteilter Gruppen/Personen durch webbasierte Informationssysteme überhaupt erst ermöglich wird, repräsentiert der Technologieaspekt ein konstituierendes Merkmal von Crowdsourcing, welches in einer definitorischen Abgrenzung berücksichtigt werden muss. Weiterhin mag kritisiert werden, dass es sich bei dem „open call" streng genommen nicht um einen offenen Aufruf handelt, sondern sich dieser stets nur an Mitglieder der jeweiligen Plattform richtet.
Dieser Dissens zwischen theoretischem Verständnis und unternehmerischer Wirklichkeit legt es nahe, den Crowdsourcing Begriff deutlich weiter zu fassen. Dabei erscheint es sinnvoll, dass zwischen den bestehenden Crowdsourcing Formen divergierende Merkmale, z. B. das zugrundeliegende Anreizsystem oder der Wissensstand der Teilnehmer, in den Definitionsansatz mit aufgenommen werden, um das Begriffsverständnis zu präziseren und eine Abgrenzung gegenüber verwandten Konzepten zu erleichtern. Vor diesem Hintergrund soll folgender Definitionsansatz vorgeschlagen werden:

Crowdsourcing ist eine interaktive Form der Leistungserbringung, die kollaborativ oder wettbewerbsorientiert organisiert ist und eine große Anzahl extrinsisch oder intrinsisch motivierter Akteure unterschiedlichen Wissensstands unter Verwendung moderner IuK-Systeme auf Basis von Web 2.0 einbezieht. Leistungsobjekt sind Produkte oder Dienstleistungen unterschiedlichen Innovationsgrades, welche durch das Netzwerk der Partizipierenden reaktiv aufgrund externer Anstöße oder proaktiv durch selbsttätiges Identifizieren von Bedarfslücken bzw. Opportunitäten entwickelt werden.

3.2 Angrenzende Konzepte zur interaktiven Leistungserstellung

Im Zuge der fortschreitenden Entwicklung von IuK-Systemen sind in den letzten Jahren zahlreiche Konzepte entstanden, die mit Crowdsourcing in Beziehung stehen. Hier sind insb. die Interaktive Wertschöpfung, Open Innovation oder Open Source zu nennen, wobei erstere vornehmlich im Marketingbereich angesiedelt sind, während Open Source ausschließlich die Entwicklung von Software zum Gegenstand hat.

Die Interaktive Wertschöpfung nach [RP06] umfasst den Prozess einer kooperativen Zusammenarbeit von Herstellern und Kunden. Diese kann sich zwischen den Extrempunkten einer vollkommenen hersteller- und/oder einer kundendominierten Wertschöpfung bewegen. Abhängig davon, in welchem Stadium des Wertschöpfungsprozesses sich das Unternehmen befindet, kann entweder von Open Innovation oder von der Produktindividualisierung (Mass Customization) gesprochen werden [RP06]. Reichwald und Piller gehen sogar von einer Identität der Interaktiven Wertschöpfung und Crowdsourcing aus.[12] Interaktive Wertschöpfung bzw. Crowdsourcing lägen dann vor, „wenn ein Unternehmen (oder eine andere Institution) eine Aufgabe, die bislang intern durch die Mitarbeiter erstellt wurde, an ein undefiniertes, großes Netzwerk von Kunden und Nutzern in Form eines offenen Aufrufs zur Mitwirkung vergibt."[13] Dagegen betrachten [RKK07] Crowdsourcing als ein Teilgebiet des Social Commerce, welcher eine logische Weiterentwicklung von Electronic Commerce darstellt und mit der Interaktiven Wertschöpfung gleichzusetzen sei.

In wie weit es sich bei Social Commerce und der Interaktiven Wertschöpfung um äquivalente Ansätze handelt, soll an dieser Stelle nicht thematisiert werden. Einer Gleichheit von Interaktiver Wertschöpfung und Crowdsourcing ist jedoch zu widersprechen, da erstere ausschließlich die Unternehmensperspektive repräsentiert und selbstorganisierte, proaktiv agierende Communities vernachlässigt, obwohl diese eine wichtige Komponente der Crowdsourcing Praxis repräsentieren. Widersprüche ergeben sich ebenfalls auch hinsichtlich der Motivationsanreize der beteiligten Individuen. Die Interaktive Wertschöpfung beruft sich hier insb. auf den Lead-User-Ansatz [Hi86], in dem sich Nutzer durch den „Zustand der Unzufriedenheit" [Ol80] dazu veranlasst fühlen, eine innovative Lösung selbst zu realisieren bzw. innerhalb einer Plattform Innovationen zu entwickeln [RP05]. Andere Motivationsanreize bleiben ebenso unberücksichtigt wie eine Interaktion mit sonstigen Individuen, die nicht unmittelbar zum Nutzerkreis eines Produktes bzw. einer Dienstleistung zählen. Kapitel 2 dokumentiert aber, dass beides für Crowdsourcing charakteristisch ist.

Überschneidungen zwischen Crowdsourcing und der Interaktiven Wertschöpfung ergeben sich aber hinsichtlich der Innovationsperspektive, welche innerhalb der Interaktiven Wertschöpfung vornehmlich durch Open Innovation repräsentiert wird. Allgemein formuliert steht Open Innovation für einen offenen Innovationsprozess, d.h. es handelt sich um eine aktive Einbindung Externer in Wertschöpfungsprozesse. Die ursprüngliche Definition in [Ch03] lautet: „Open Innovation is a paradigm that assumes that firms can and should use external ideas as well as internal ideas, and internal and external paths to market, as the firms look to advance their technology". Es handelt sich also wie beim Crowdsourcing um eine Erweiterung der Unternehmensgrenzen. Hierbei basiert der Ansatz auf dem Einbezug Externer, deren Motivationsanreize extrinsisch wie intrinsisch ausgeprägt sein können. Open Innovation weist damit große Ähnlichkeiten mit Crowdsourcing auf, wobei die Leistungserstellung bei letzterem aber nicht notwendigerweise auf Innovationen ausgerichtet sein muss. Ferner wird die organisatorische und insb. technische Umsetzung bei Open Innovation nicht thematisiert, wohingegen Crowdsourcing de facto den Einsatz webbasierter Plattformen nach dem Prinzip des Web 2.0 vorsieht.

Im Bereich der Softwareentwicklung ist eine interaktive Leistungserstellung seit vielen Jahren als Open Source bekannt. Dieses Prinzip beinhaltet folgende Aspekte: Eine Software liegt in einer für den Menschen lesbaren Form vor und darf ohne Beschränkung genutzt sowie weiter verbreitet werden [Di99]. Ferner sind Modifikationen, mit geringen Auflagen, grundsätzlich gestattet und sogar erwünscht. Dementsprechend wird bei Open Source eine große Anzahl, fachlich versierter Personen, aber nicht ausschließlich Nutzer/Kunden, in den Prozess der Softwareentwicklung aktiv integriert. [Ra01] bezeichnet das mitwirkende Individuum als einen „in its true and original sense of an enthusiast, an artist, a thinkerer, a problem solver, an expert". Daher kann postuliert werden, dass die Übertragung des Gedankens von Open Source

[12] http://www.open-innovation.com/iws/
[13] http://www.open-innovation.com/iws/faq.html#1

auf andere Produktbereiche, unter Erweiterung des Spezialisierungsgrades der beteiligten Individuen und einer abgeänderten Regelung der Benutzerrechte, eine Form von Crowdsourcing darstellt. Erfolgt eine Beschränkung auf den Teilbereich von Open Source, welcher sich auf die Entwicklung grundsätzlich neuartiger Software bezieht, ergibt sich eine analoge Beziehung zu Open Innovation; vgl. auch [Pi03].
Es lässt sich also feststellen, dass Crowdsourcing konzeptionelle Unterschiede zu Open Innovation und Open Source aufweist bzw. diese erweitert. Überschneidungen mit der Interaktiven Wertschöpfung ergeben sich lediglich hinsichtlich der Innovationsperspektive, so dass insg. davon ausgegangen werden kann, dass die Bildung eines eigenständigen Crowdsourcing Begriffs gerechtfertigt ist.

4. Crowdsourcing Klassifikationsschema

Der Crowdsourcing Ansatz stellt aus Unternehmensperspektive ein interessantes und aktuelles Themengebiet dar, dessen zukünftiger Erfolg aber maßgeblich von einem klaren Verständnis der Inhalte, Potentiale und Grenzen abhängen dürfte. Daher, soll im Folgenden ein Klassifikationsschema in Form einer Portfoliomatrix entworfen werden, welches eine Systematisierung praktischer Crowdsourcing Formen einerseits und angrenzender theoretischer Ansätze andererseits, erlaubt. Hierzu ist zunächst zu klären, welche der in Kapitel 3.1 identifizierten Merkmale für eine Abgrenzung von Crowdsourcing gegenüber anderen Formen der interaktiven Leistungserstellung besonders geeignet sind.

4.1 Dimensionen des Crowdsourcing Portfolios

Open Innovation beschränkt sich nicht auf eine bestimmte Form der Incentivierung und deckt das ganze Spektrum intrinsisch motivierter und ökonomisch agierender Individuen ab. Das klassische Open Source Verständnis (vgl. auch [Ra01]) geht eher von einer intrinsischen Motivation der Teilnehmer aus. Allerdings legt die erhebliche kommerzielle Bedeutung von Open Source nahe, dass in der heutigen Zeit sehr wohl auch monetäre Anreize bestehen. Gleichermaßen ist hinsichtlich der Tätigkeitsinitiierung festzustellen, dass sowohl Open Innovation als auch Open Source externe Anstöße und selbsttätiges Handeln beinhalten. Beispielsweise betont der Lead-User Ansatz das Handeln aus eigener Unzufriedenheit heraus, wohingegen das in [Ch03] zum Ausdruck kommende Verständnis das nach Lösungen suchende Unternehmen in den Fordergrund stellt.
Eine kompetitiv organisierte Leistungserstellung, wie sie bei Crowdsourcing teilweise stattfindet, kennt Open Source nicht. Ähnlich stehen bei Open Innovation andere Aspekte im Vordergrund und es wird allgemein von einer Kollaboration zwischen Unternehmen und Kunden/Partnern ausgegangen [Pi03]. Unterschiede ergeben sich ferner bezüglich des (fachlichen) Erfahrungsstandes der Partizipierenden. Da Open Source die Entwicklung von Software zum Gegenstand hat, verfügen die Akteure über – i.d.R. weitreichende – Programmierkenntnisse. Open Innovation dient allgemein der Einbeziehung externer Partner in den Innovationsprozess, ohne deren Wissensstand bzw. Expertise näher zu spezifizieren. Dies verdeutlicht auch, dass Open Innovation eine Ausgliederung routinemäßiger, geringwertiger Tätigkeiten, wie z. B. bei Mechanical Turk, nicht beinhaltet und die Innovationsdimension ein wichtiges Unterscheidungsmerkmal zu Crowdsourcing ist. Eine Beschränkung auf innovative Produkte liegt bei Open Source nicht vor, vielmehr wird der Innovationsgrad des Leistungsobjekts von der Neuartigkeit der entwickelten Software determiniert.
Aus den obigen Ausführungen wird deutlich, dass vor allem die Dimensionen Innovationsgrad der erbrachten Leistung und Kenntnisstand der partizipierenden Individuen als Kriterien herangezogen werden können, anhand derer eine schematische Einordnung der

vorgestellten Konzepte sowie eine Klassifikation der vorgestellten Praxisbeispiele erfolgen kann. Dabei ist zu berücksichtigen, dass lediglich Crowdsourcing auch wettbewerblich organisierte Leistungserstellungsformen beinhaltet, so dass auch dieses Merkmal große Relevanz für eine Abgrenzung besitzt. Da sich aber in dieser Dimension keine Unterschiede zwischen Open Source und Open Innovation ergeben, wurde aus Gründen einer vereinfachten Darstellung davon abgesehen, diese dritte Dimension im Klassifikationsschema zu berücksichtigen.

Innovationen werden hinsichtlich der Wirkungsaspekte häufig in „inkrementell" und „radikal" unterschieden [Ha04; VB00]. Hierbei umfasst der Begriff Innovation die Neuartigkeit einer Problemslösungsfunktion, d.h. es liegt eine neuartige Zweck-Mittel-Kombination vor [Ha04]. Bei einer radikalen Innovation geht es um die Entwicklung einer absoluten Neuerung, die keine Ähnlichkeiten zu bestehenden Prozessen oder Produkten aufweist und deren Zweck- und Mitteleinsatz in hohem Maße neuartig ist. Als inkrementell werden Innovationen charakterisiert, welche aus einer neuen Kombination von Zweck und Mittel oder einem verbesserten Zweck/Mittel-Verhältnis bestehen und z. B. zu Produkten mit leicht verbesserten Funktionen führen. Es ist fraglich, in wie fern radikale Innovationen Gegenstand von Crowdsourcing/Open Innovation sein können. Konzeptionell zielt die Integration Externer in den Innovationsprozess eher auf inkrementelle Verbesserungen. Fundamentale Neuerungen im Sinne einer radikalen Innovation sind unwahrscheinlich, aber nicht grundsätzlich auszuschließen. Daher werden im Folgenden lediglich die Ausprägungen „innovativ" und „nicht-innovativ" unterschieden.

Neben dem Aspekt der Innovation, wird der Wissensstand der Partizipierenden als Dimension herangezogen. Hierbei geht es um die Kenntnisse bzw. Erfahrungen, welche Akteure benötigen, um die vorgegebenen Aufgabenstellungen zu bewältigen. Crowdsourcing, Open Innovation sowie Open Source beinhalten hierbei Tätigkeiten, deren Komplexitätsgrad ein ausgeprägtes Fachwissen erfordert. Darüber hinaus beziehen einige Crowdsourcing Formen und Open Innovation aber auch Amateure in die Leistungserstellung mit ein. Diese Generalisierung bestehender Ansätze durch Crowdsourcing soll in Abbildung 1 verdeutlicht werden.

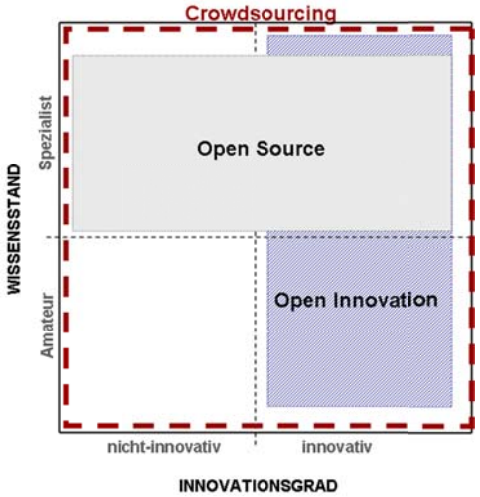


Abbildung 4: Crowdsourcing Klassifikationsschema

4.2 Klassifikation von Crowdsourcing Plattformen

Das oben abgeleitete Klassifikationsschema soll nachfolgend für eine Gruppierung der beschriebenen Crowdsourcing Umsetzungen genutzt werden, um zu analysieren, in wie weit die Praxis bereits das volle Spektrum an Crowdsourcing Formen abdeckt. Auf der Basis von Kapitel 2 ergibt sich hierbei die in Abbildung 2 dargestellte Einteilung:

Im Quadranten I beziehen InnoCentive und NineSigma ausschließlich Hochqualifizierte in den Entwicklungsprozess innovativer Produkte/Dienstleistungen ein. In diesem Zusammenhang werden die Plattformen zur Öffnung der Unternehmensgrenzen genutzt, insb. im Hinblick auf eine Erweiterung der internen Forschungs- und Entwicklungsabteilungen. Diese Plattformen verkörpern damit auch den Grundgedanken von Open Innovation, wobei die kompetitiv organisierte Leistungserstellung zu bedenken ist. Cambrian House zielt ebenfalls auf Innovationen ab und bedient sich der Kompetenz von Spezialisten. Allerdings ist der Wissensstand der beteiligten Individuen heterogener. Ferner kommt dieser Form von Crowdsourcing eine gewisse Eigenständigkeit zu, da der Tätigkeitsanstoß aus der Community selbst kommt.

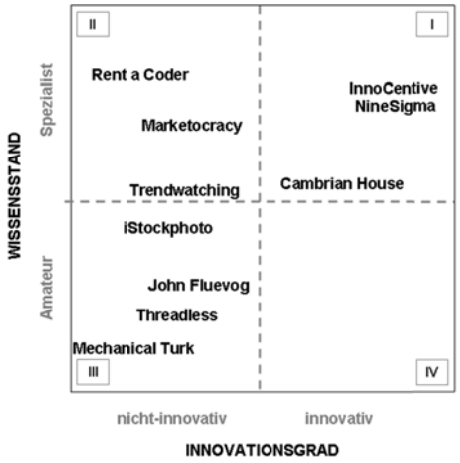


Abbildung 5: Klassifikation derzeitiger Crowdsourcing Plattformen[14]

Quadrant II steht stellvertretend für Plattformen, die als eine Alternative zum traditionellen Branchenanbieter verstanden werden können. Bei Trendwatching berichten Journalisten über vermeintlich innovative Trends und bei Marketocracy werden fähige Investoren und deren Strategien gesucht. Rent a Coder repräsentiert insb. das in der Definition von [Ho06] zum Ausdruck kommende Crowdsourcing Verständnis und kann als Weiterentwicklung von Outsourcing verstanden werden. Anstatt mit einem Outsourcing Partner zusammenzuarbeiten, werden Programmierer für begrenzte Zeit direkt über die Crowdsourcing Plattform angeworben.
Quadrant III beinhaltet Plattformen, die ebenfalls keine besondere Innovativität aufweisen und durch die Einbeziehung von Amateuren und/oder Hobbyisten gekennzeichnet sind. Beispiele wie John Fluevog und Threadless verdeutlichen, dass der Kreativsektor bei dieser Form von Crowdsourcing eine wichtige Rolle spielt. Eine Unterscheidung anhand der Wissensdimension, z. T. auch gegenüber Trendwatching, ist bei diesen Plattformen schwierig und orientiert sich eher an der derzeitigen Nutzerschaft[15] als an grundsätzlichen Unterschieden der jeweils im Fokus stehenden Aufgaben. Eine Sonderrolle nimmt jedoch die Plattform Mechanical Turk ein, welche ebenfalls eine große Nähe zum klassischen Outsourcing aufweist. Interessant ist, dass gerade der Quadrant III stark besetzt ist, weil dieser konzeptionell ausschließlich durch Crowdsourcing besetzt ist.[16] Hieran wird die Notwendigkeit, Crowdsourcing als eigenständiges Wissenschaftsgebiet wahrzunehmen, noch einmal besonders deutlich.
Schlussendlich repräsentiert Quadrant IV einen Bereich, für den derzeit noch keine Plattformen existieren. Im Rahmen von Open Innovation bezieht sich die Entwicklung (inkrementell) neuartiger Produkte unter Einbeziehung von Teilnehmern ohne spezielles

[14] Für die bestehenden Crowdsourcing Plattformen gilt umso mehr, dass sich die Innovationsdimension auf inkrementelle Neuerungen im Sinne von [Ha04] bezieht.
[15] Dieser Schluss basiert auf einer Analyse der Webseiten der Plattformen (z. B. Beiträge in Foren).
[16] Vgl. hierzu auch Abbildung 1.

Fachwissens zumeist auf die Adaption eines Produktes zur verbesserten Bedienung von Kundenwünschen. Crowdsourcing könnte hier einen wertvollen Beitrag liefern, indem moderne IuK-Systeme genutzt werden, um die Präferenzen und Anregungen einer großen Masse an Kunden zu erheben und zu verdichten.

5. Schlussbetrachtung

In der vorliegenden Arbeit wurde das Konzept des Crowdsourcing betrachtet und ein Klassifikationsschema für dessen verschiedene Ausprägungen entwickelt. Dabei wurde aufgezeigt, worin die Unterschiede zu ähnlichen Ansätzen bestehen und in welchen Bereichen Überschneidungen existieren. Während das Open Source Prinzip aufgrund des Betrachtungsobjekts Software in der Wirtschaftsinformatik wohl bekannt ist, werden Konzepte wie die Interaktive Wertschöpfung vornehmlich im Marketing behandelt. Gemein ist diesen Ansätzen der hohe Grad an Interaktivität, d.h. die Einbeziehung eines großen Kreises externer Personen. Crowdsourcing generalisiert diese Ansätze hinsichtlich der Motivation der beteiligten Individuen (intrinsisch/extrinsisch), der Organisation der Leistungserstellung (kollaborativ/kompetitiv), des Bezugsobjekts (jegliche Form von Produkten oder Dienstleistungen) sowie der Projektinitiation (reaktiv, z. B. bei Aufruf oder proaktiv, z. B. bei Entdecken einer Bedarfslücke). Darüber hinaus ist der Ansatz aus wirtschaftsinformatischer Sicht insbesondere auch wegen der intensiven Nutzung moderner Informationssysteme im Zusammenhang mit dem Web 2.0 interessant. Die vorgestellten Plattformen dokumentieren die vielfältigen Möglichkeiten, bestehende Geschäftsprozesse (z. B. im Rahmen des Innovationsmanagement) oder neue Geschäftsmodelle auf Basis von Crowdsourcing zu verbessern bzw. zu realisieren.
Allerdings besteht noch erheblicher Forschungsbedarf, um das Potential und die Grenzen von Crowdsourcing systematisch zu analysieren. So ist aus Unternehmensperspektive die Frage, welche Art von Aufgabenstellungen auch durch eine Form von Crowdsourcing wirtschaftlich erledigt werden können, von primärem Interesse. In diesem Zusammenhang ist beispielsweise zu klären, welche Anforderungen an eine Problemspezifikation zu stellen und wie Vertragsvereinbarungen zu gestalten sind, damit extern durch anonyme Akteure erbrachte Leistungen nahtlos im Unternehmen genutzt werden können.
Ferner ist zu untersuchen, ob und ggf. unter welchen Umständen die Bereitstellung eines dedizierten IuK-System für Crowdsourcingzwecke für ein Unternehmen sinnvoll ist. Mit Ausnahme von John Fluevog werden derzeitige Lösungen von einem Drittanbieter betrieben. Letztendlich kann aber auch Crowdsourcing als eine Form computergestützter Zusammenarbeit aufgefasst werden, wobei sich neue Dimensionen hinsichtlich der Anzahl der Systemnutzer sowie deren Nähe zum Unternehmen ergeben. Weiterhin beinhaltet das derzeitige Verständnis von Kollaborationssystemen keine kompetitiv organisierten Formen der Leistungserbringung. Andererseits ergeben sich hinsichtlich der Ermöglichung eines asynchronen Zusammenwirkens dislozierter Individuen erhebliche Überschneidungen zwischen Crowdsourcing Plattformen und klassischen Kollaborationssystemen. Folglich wäre zu diskutieren, ob eine entsprechende Ausweitung des Kooperationssystembegriffs zweckmäßig ist und welche neuen funktionalen Anforderungen an solche Systeme zu stellen wären.

Literaturverzeichnis

[Ch03] Chresbrough, H.W.: Open Innovation: The New Imperative for Creating and Profiting from Technology. Harvard Business School Press, Boston, 2003.

[Da89] Davis, S.: From Future Perfect: Mass Customization. Planning Review 17, 1989; S. 9-28.

[Di99] DiBona, C.; Ockman, S.; Stone, M.: Open Source: Voices from the Open Source Revolution. OReilly Media, Beijing, 1999.

[Fe07] Fettke, P.: Supply Chain Management: Stand der empirischen Forschung. Zeitschrift für Betriebswirtschaft 77(4), 2007; S. 417-461.

[Ha04] Hauschildt, J.: Innovationsmanagement. 3. Auflage, Vahlen, München, 2004.

[Hi86] von Hippel, E.: Lead Users. A Source of Novel Product Concepts. Management Science 32, 1986; S. 781-805.

[Ho06] Howe, J.: The Rise of Crowdsourcing. Wired Magazine, http://www.wired.com/wired/archive/14.06/crowds.html, 2006.

[HW02] Hippner, H.; Wilde, K.D.: CRM – Ein Überblick. In (Helmke, S.; Uebel, M.; Dangelmaier, W., Hrsg.): Effektives Customer Relationship Management. Gabler, Wiesbaden, 2002; S. 3-28.

[HS06] Huston, L.; Sakkab, N.: Connect and Develop: Inside Procter & Gamble's New Model for Innovation. Harvard Business Review 84, 2006; S. 58-66.

[Ol80] Oliver, R.: A Cognitive Model of the Antecedents and Consequences of Satisfaction Divisions. Journal of Marketing Research 417, 1980; S. 140-146.

[O`R05] O'Reilly, T.: What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. http://www.oreillynet.com/pub/a/oreilly/tim/news/-2005/09/30/ what-is-web-20.html, 2005.

[Pi03] Piller, F.: Von Open Source zu Open Innovation. Harvard Business Manager 25, 2003; S.114-115.

[Pi93] Pine, B.J.: Mass Customization: The New Frontier in Business Competition. Harvard Business School Press, Boston, 1993.

[Po80] Porter, M. E.: Competitive Strategy: Techniques for Analyzing Industries and Competitors. The Free Press, New York, 1980.

[Ra01] Raymond E. S.: The Cathedral and the Bazar: Musings on Linux and Open Source by an Accidental Revolutionary. O'Reilly Media, Bejing, 2001.

[RKK07] Richter, A.; Koch, M.; Krisch, J.: Social Commerce – Eine Analyse des Wandels im E-Commerce. Technischer Bericht Nr. 2007-03, Fakultät für Informatik, Universität der Bundeswehr München, http://www.kooperationssysteme.de, 2007.

[RP05] Reichwald, R.; Piller, F.: Open Innovation: Kunden als Partner im Innovationsprozess. http://www.impulse.de/downloads/open_innovation.pdf, 2005.

[RP06] Reichwald, R.; Piller, F.: Interaktive Wertschöpfung. Gabler Verlag, Wiesbaden, 2006.

[Sc02] Schwabe, G.: Kooperationssysteme und Wissensmanagement, Bellmann, M.; Somerlatte, T.; Kramer, H. (Hrsg.); Wissensmanagement, Symposium, 2002.

[Su04] Surowieki, J.: Die Weisheit der Vielen. 2. Auflage, Bertelsmann, München, 2004.

[VB00] Voß, S.; Böse, J.: Innovationsentscheidungen bei logistischen Dienstleistern - Praktische Erfahrungen in der Seeverkehrswirtschaft. In (Dangelmaier, W.; Felser, W., Hrsg.): Das reagible Unternehmen. HNI, Paderborn, 2000; S.253-282.

Sämtliche im Text angegebenen Internetquellen wurden, wenn nicht anders angegeben, am 19.09.2007 zuletzt abgerufen.

Appendix: Summaries / Zusammenfassungen

Paul, P., Martin, N., Sattler, H., & Hennig-Thurau, T. (2012). Customer-related assets and their contribution to firm value: A theoretical framework and empirical application. To be submitted to *Journal of Marketing*.

What is the value of marketing? Many CEOs want an answer to this question and pressure marketers to identify the assets generated by marketing and to monetize the contributions of these assets to firm value. The extant research on marketing-related assets has been limited in two respects: the lack of a comprehensive, *non-overlapping* and *measurable* framework for these assets and the unknown applicability of marketing-related assets to financial accounting. Building on prior research, this research offers a customer-related assets (CRAs) framework that uses a customer-centric perspective to identify a comprehensive and mutually exclusive set of customer-related assets and that integrates these assets with financial accounting standards. In addition to offering a method of monetizing the contribution of marketing to firm value, the CRAs framework clarifies the controversial relationship between brand equity and customer equity. The authors demonstrate the practical use of this framework for a major European corporation, combining a large-scale empirical study involving survey data with information from the firm's internal databases.

Die angestrebte Maximierung des Unternehmenswertes stellt ein zentrales und übergeordnetes Ziel einer Unternehmung dar. Hierbei stehen Unternehmenseinheiten zunehmend im Wettbewerbskampf um knappe Ressourcen zueinander. Insbesondere für das Marketing lässt sich dabei die Schwierigkeit beobachten, eine ökonomische Wertschöpfung von intangiblen, kundenbezogene Vermögensgegenständen nachzuweisen. Als Konsequenz einer erschwerten und nicht-trivialen Bewertbarkeit lässt sich feststellen, dass das Marketing droht an relativer Bedeutung im Unternehmenskontext einzubüßen. Bisherige Forschungsbemühungen stellen keinen Systematisierungsansatz zur Verfügung, um kundenbezogene Vermögensgegenstände ganzheitlich und überschneidungsfrei zu definieren und messbar zu machen. Deshalb ist es das Ziel dieses Beitrags einen definitorischen Rahmen und einen Messansatz für kundenbezogene Vermögensgegenstände einzuführen, der mit Finanz- bzw. Rechnungslegungsstandards im Einklang steht. Hierdurch können intangible, kundenbezogene Vermögensgegenstände messbar gemacht und von anderen Wertkomponenten trennscharf dargestellt werden (z. B. Markenwert, Kundenwert). Die Autoren verdeutlichen die Anwendbarkeit des eingeführten Messansatzes mithilfe eines empirischen Fallbeispiels, das aus einer Unternehmens-kooperation resultiert.

Zenker, S., & Martin, N. (2011). Measuring success in place marketing and branding. ***Place Branding & Public Diplomacy, 7(1),*** **32-41.**

As the competition between cities increases, cities focus more and more on establishing themselves as brands. Consequently, cities invest an extensive amount of taxpayers' money into their marketing activities. Unfortunately, cities still lack a proper success measurement, which has raised questions regarding the efficient and effective use of the taxpayers' money. With this contribution the authors want to highlight some existing, but primarily new possibilities for a complex success measurement in place marketing, referring to the extant literature on place marketing and the general field of marketing. Therewith, the authors strive to translate different concepts like customer equity or customer satisfaction into the lexicon of place marketing, thus identifying empirical gaps for further research, as well as existing fruitful approaches.

Städte werben um diverse Zielgruppen, wie Einwohner, Investoren oder Besucher und stehen zunehmend unter dem Druck, sich als Marken zu etablieren. Hieraus entsteht die Notwendigkeit, einen stärkeren Fokus auf Investitionen in Marketingmaßnahmen, finanziert aus Steuergeldern, zu legen und deren Effektivität und Effizienz nachzuhalten. Allerdings lässt sich feststellen, dass Städte bisher nur unzureichend den Erfolg von Marketinginvestitionen erfassen. Es bleibt demnach ungeklärt, wie effektiv und effizient Steuergelder für Marketingzwecke eingesetzt werden. Vor diesem Hintergrund adressiert dieser Beitrag die Erfolgsmessung von Marketinginvestitionen, indem neue Möglichkeiten einer effektiven und effizienten Wertmessung von Marketingmaßnahmen im Kontext von Stadtmarketing aufgezeigt werden. Es wird sowohl ein Überblick über bisherige Forschungsbeiträge aus dem Forschungsfeld des Stadtmarketings als auch eine Übertragung von ausgewählten Wertmetriken aus dem klassischen Marketing in den Kontext des Stadtmarketings präsentiert. Damit erfolgt eine Einführung neuer Konzepte, wie die des „Citizen Equity" und eine Diskussion bereits bestehender Werterfassungsansätze.

Martin, N. (2012). Assessing scorecard performances: A literature review and classification. To be submitted soon to *Expert Systems with Applications*.

The assessment of scorecard performance in the field of credit scoring is of major relevance to firms. This study presents the first systematic academic literature review of how empirical benchmark studies assess scorecard performance in the field of credit scoring. By analysing 62 comparative studies, this study provides two main contributions. First, this study provides a systematic overview of the assessment-related decisions of all the reviewed studies based on a classification framework. Second, the assessment criteria of consistency, application fit, and transparency are introduced and used to discuss the observed assessment-related decisions. As the findings show, researchers often pay insufficient attention to ensuring the consistent assessment of scorecard performance. Moreover, the majority of the reviewed studies choose performance indicators that failed to fit the application context and provided non-transparent assessment documentation. In conclusion, these researchers pay a great deal of attention to the development of scorecards, but they often fail to implement a straightforward assessment procedure.

Scorecards werden im Kreditwesen routinemäßig eingesetzt, um Entscheidungsprozesse zu unterstützen. Wie eine Vielzahl an Publikationen zeigen, nimmt die Entwicklung und Evaluierung solcher Scorecards eine zentrale Rolle ein. Neben unzähligen Vergleichsstudien von Scorecards erscheinen jedoch nur vereinzelnd Diskussionen, die die Gütebewertung thematisieren, obgleich die Bewertung von Scorecards eine zentrale Aufgabe darstellt. Der vorliegende Beitrag behandelt die Frage, wie Scorecards im Anwendungskontext des Kreditwesens bewertet und ausgewählt werden sollten. Auf Basis einer Analyse von 62 ausgewählten Vergleichsstudien wird zum einen ein systematischer Überblick über den bisherigen Umgang mit der Gütebewertung gegeben. Zum anderen wird ein Kriterienkatalog eingeführt mithilfe dessen die beobachteten Entscheidungen kritisch diskutiert, Fehler aufgedeckt und zukünftige Forschungsfelder aufgezeigt werden. Es zeigt sich, dass der Großteil bisheriger Forschungsarbeiten die Weiterentwicklung immer komplexer werdender Scorecards vorantreibt, jedoch eine stringente, passende und transparente Gütebewertung vernachlässigt und hierdurch die Aussagekraft von Forschungsbemühungen eingeschränkt wird.

Martin, N., & Lessmann, S. (2012). Bewertung und Auswahl von Scorecards im Kreditwesen: Eine Analyse zur Eignung von *Kosten-Kurven*. Submitted to *Zeitschrift für betriebswirtschaftliche Forschung.*

Credit scorecards are routinely used in the financial service industry to guide decision making in marketing and risk management. The paper is concerned with the problem of identifying an appropriate scorecard among a set of alternatives. To that end, a requirement specification for scorecard assessment in the credit industry is developed. Examining the compliance of current assessment practices with these requirements, the authors find that standard performance measures suffer important limitations. The Cost Curve methodology is introduced as a more powerful tool for scorecard selection in credit scoring applications. Its unique advantages are illustrated by means of an empirical study. A key implication of the paper is that Cost Curves facilitate a business oriented scorecard selection and, thereby, contribute toward increasing decision quality in scorecard-supported business processes.

Scorecards werden im Kreditwesen routinemäßig eingesetzt, um Entscheidungsprozesse im Marketing und im Risikomanagement zu unterstützen. Dem Einsatz einer Scorecard geht ein Auswahlprozess voraus, in dessen Rahmen alternative Modelle entwickelt und verglichen werden. Der vorliegende Beitrag behandelt die Frage, wie diese Scorecardbewertung bzw. Scorecardauswahl erfolgen sollte. Hierzu wird ein Kriterienkatalog entwickelt, der die spezifischen Anforderungen des Kreditwesens zusammenfasst. Auf dieser Basis werden gebräuchliche Instrumente zur Scorecardauswahl und Scorecardbewertung analysiert und deren Schwächen offenbart. Mit den *Kosten-Kurven* wird ein neues Bewertungsinstrument für das Kreditwesen vorgestellt und empirisch verdeutlicht, welche Vorteile sich aus seinem Einsatz ergeben. Eine wesentliche Implikation des Beitrags ist, dass *Kosten-Kurven* eine ökonomisch motivierte Scorecardbewertung ermöglichen und damit zu einer höheren Entscheidungsqualität in Scorecard-gestützten Geschäftsprozessen beitragen.

Martin, N., Lessmann, S., & Voß, S. (2008). Crowdsourcing: Systematisierung praktischer Ausprägungen und verwandter Konzepte. In: Martin Bichler, Thomas Hess, Helmut Krcmar, Ulrike Lechner, Florian Matthes, Arnold Picot, … (Eds.), ***Multikonferenz Wirtschaftsinformatik, MKWI 2008, München, 26.2.2008 - 28.2.2008, Proceedings.*** **GITO-Verlag, Berlin.**

This paper focusses on the concept of Crowdsourcing which refers to an organizational concept of an interactive product and/or service creation process based on web 2.0 technology. By analyzing different Crowdsourcing communities, the authors introduce the first academic definition of the Crowdsourcing concept and a classification framework to distinguish between different types of Crowdsourcing communities. Accordingly, systematical differences between Crowdsourcing and related concepts, namely Open Source and Open Innovation, are detected. It is argued that Crowdsourcing generalizes Open Source and Open Innovation regarding: the motivation of the included persons, the organization of the product and/or service creation process, the aimed objective and the project initiation.

Die Arbeit betrachtet das Crowdsourcing als ein aktuell diskutiertes Konzept für die Organisation eines überbetrieblichen, interaktiven Leistungsaustauschs auf der Basis von Web 2.0. In der wissenschaftlichen Literatur wurde dieser Ansatz bisher wenig beachtet, wohingegen sich in der betrieblichen Praxis bereits einige, z. T. aber stark unterschiedliche „Crowdsourcing Plattformen“ finden. In Ermangelung eines allgemeinen Begriffsverständnisses ist es das Ziel der vorliegenden Arbeit, das Crowdsourcing Konzept zu systematisieren. Dazu werden ein Definitionsansatz sowie ein Klassifikationsschema vorgeschlagen, welche aus der Analyse bestehender Crowdsourcing Formen und angrenzender theoretischer Konzepte abgeleitet werden.

Appendix: Liste der Veröffentlichungen

Appendix: Liste der Veröffentlichungen

Publizierte Aufsätze:

Martin, N., Lessmann, S., & Voß, S. (2008). Crowdsourcing: Systematisierung praktischer Ausprägungen und verwandter Konzepte. In: Martin Bichler, Thomas Hess, Helmut Krcmar, Ulrike Lechner, Florian Matthes, Arnold Picot, … (Eds.), *Multikonferenz Wirtschaftsinformatik, MKWI 2008, München, 26.2.2008 - 28.2.2008, Proceedings.* GITO-Verlag, Berlin.

Zenker, S., & Martin, N. (2011). Measuring success in place marketing and branding. *Place Branding & Public Diplomacy, 7(1),* 32-41.

Bisher unplublizierte Aufsätze:

Martin, N. (2012). Assessing scorecard performances: A literature review and classification. Working Paper, to be submitted soon to *Expert Systems with Applications*.

Martin, N, & Lessmann, S. (2012). Bewertung und Auswahl von Scorecards im Kreditwesen: Eine Analyse zur Eignung von *Kosten-Kurven.* Working Paper, Submittet to *Zeitschrift für betriebswirtschaftliche Forschung*.

Paul, P., Martin, N., Sattler, H., & Hennig-Thurau, T. (2012). Customer-related assets and their contribution to firm value: A theoretical framework and empirical application. Working Paper, targeted to submission to *Journal of Marketing*.

Appendix: Eidesstattliche Versicherung

Appendix: Eidesstattliche Versicherung

Hiermit erkläre ich, Dipl.-Kffr. Nicole Martin, an Eides statt, dass ich die Dissertation mit dem Titel:

„*Brands and customers as drivers of firm value*“

selbstständig und ohne fremde Hilfe verfasst habe.

Andere als die von mir angegebenen Quellen und Hilfsmittel habe ich nicht benutzt. Die den herangezogenen Werken wörtlich oder sinngemäß entnommenen Stellen sind als solche gekennzeichnet.

Hamburg, den	Unterschrift