The Management of Direct Material Cost During New Product Development: A Case Study on the Application of Big Data, Machine Learning, and Target Costing

Zur Erlangung des akademischen Grades eines

Doktors der Wirtschaftswissenschaften

Dr. rer. pol.

von der KIT-Fakultät für Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation

von

M.Sc. Dominik Hammann

Tag der mündlichen Prüfung:	23.03.2022
Erster Gutachter:	Prof. Dr. Marc Wouters
Zweiter Gutachter:	Prof. Dr. Gerhard Satzger

Karlsruhe, 2022

Abstract

This dissertation thesis investigates the application of big data, machine learning, and the target costing approach for managing costs during new product development in the context of high product complexity and uncertainty. A longitudinal case study at a German car manufacturer is conducted to examine the topic. First, we conduct a ***systematic literature review***, which analyzes use cases, issues, and benefits of big data and machine learning technology for the application in management accounting. Our review contributes to the literature by providing an overview about the specific aspects of both technologies that can be applied in managerial accounting. Further, we identify the specific issues and benefits of both technologies in the context management accounting. Second, we present a ***case study*** on the applicability of machine learning and big data technology for product cost estimation, focusing on the material costs of passenger cars. Our case study contributes to the literature by providing a novel approach to increase the predictive accuracy of cost estimates of subsequent product generations, we show that the predictive accuracy is significantly larger when using *big* data sets, and we find that machine learning can outperform cost estimates from cost experts, or produce at least comparable results, even when dealing with highly complex products. Third, we conduct an ***experimental study*** to investigate the trade-off between accuracy (predictive performance) and explainability (transparency and interpretability) of machine learning models in the context of product cost estimation. We empirically confirm the often-implied inverse relationship between both attributes from the perspective of cost experts. Further, we show that the relative importance of explainability to accuracy perceived by cost experts is important when selecting between alternative machine learning models. Then, we present four factors that significantly determine the perceived relative importance of explainability to accuracy. Fourth, we present a ***proprietary archival study*** to investigate the target costing approach in a complex product development context, which is characterized by product design interdependence and uncertainty about target cost difficulty. We find that target cost difficulty is related to more cost reduction performance during product development based on archival company data, and thereby complement results from earlier studies, which are based on experimental studies. Further, we demonstrate that in a complex product development context, product design interdependence and uncertainty about target cost difficulty may both limit the effectiveness of target costing.

Kurzfassung

Die vorliegende Arbeit untersucht die Anwendung von Big Data, Machine Learning und der Zielkostenrechnung für das Kostenmanagement für Produkte mit hoher Komplexität und Unsicherheit während der Produktentwicklung. Zur Untersuchung des Forschungsthemas wurde eine dreijährige Fallstudie bei einem deutschen Automobilhersteller durchgeführt. Im ersten Abschnitt werden Anwendungsfälle, Chancen und Risiken von Big Data und Machine Learning im Controlling anhand einer ***systematischen Literaturrecherche*** untersucht. Zunächst stellt die Literaturrecherche einen Überblick zu den spezifischen Aspekten beider Technologien im Bereich des Controllings dar. Darüber hinaus werden die spezifischen Chancen und Risiken beider Technologien im Bezug des Controllings identifiziert. Im zweiten Abschnitt wird anhand einer ***Fallstudie*** die Anwendbarkeit von Machine Learning and Big Data für die Produktkostenschätzung von Materialeinzelkosten hochkomplexer Produkte untersucht. Zunächst wird ein neuer Ansatz zur Erhöhung der Genauigkeit von Produktkostenschätzungen nachfolgender Produktgenerationen vorgestellt. Zudem wird die signifikante Verbesserung der Genauigkeit durch die Anwendung von Big Data aufgezeigt. Ein Vergleich mit manuellen Verfahren zeigt, dass auch bei hochkomplexen Produkten Machine Learning in der Lage ist, bessere oder zumindest vergleichbare Kostenschätzungen zu generieren. Im dritten Abschnitt wird anhand eines ***Experiments*** der Zielkonflikt zwischen der Genauigkeit (Vorhersage) und Erklärbarkeit (Transparenz und Interpretierbarkeit) von Machine Learning Modellen im Kontext der Produktkostenschätzung untersucht. Dabei kann die oft unterstellte inverse Beziehung zwischen Genauigkeit und Erklärbarkeit empirisch bestätigt werden. Zudem stellt sich heraus, dass die wahrgenommene relative Wichtigkeit von Erklärbarkeit zu Genauigkeit ein wichtiger Faktor für die Auswahl von Machine Learning Modellen ist. Zuletzt werden vier Faktoren vorgestellt, welche die relative Wichtigkeit von Erklärbarkeit und Genauigkeit während der Produktentwicklung bestimmen. Die ***Archivstudie*** im vierten Abschnitt untersucht die Zielkostenrechnung im Kontext komplexer Produktentwicklung auf der Basis von Wechselwirkungen innerhalb der Produkte und Unsicherheiten bezüglich der Zielkostenlücke. Zunächst wird gezeigt, dass die Höhe der Zielkostenlücke im Zusammenhang mit der Kostenreduktion während der Produktentwicklung steht, was bisherige Forschungsergebnisse auf der Basis von Experimenten komplementiert. Zudem wird gezeigt, dass die Wechselwirkungen innerhalb der Produkte und Unsicherheiten bezüglich der Zielkostenlücke die Effektivität der Zielkostenrechnung limitieren.

Contents

List of Figures

List of Tables

Abbreviations

ABDC	Australian Business Deans Council
ANN	Artificial Neural Network
BD	Big Data
CBR	Case-Based Reasoning
CC	Current Cost
CRPD	Cost Reduction Performance during Development
CRPDoP	Cost Reduction Performance during Development over Production
CRPP	Cost Reduction Performance during Production
CRPT	Total Cost Reduction Performance
CTC	Change of Target Cost
DMC	Direct Material Cost
DTR	Decision Tree Regression
EE	Elementary Effect
ELR	Elastic Net Regression
EPD	Early Product Development
EtA	Explainability to Accuracy
EVS	Explained Variance Score
FFE	Fuzzy Front-End
GAP	Target cost difficulty
GBR	Gradient Boosted Regression
IF	Influencing product Features
LAR	Lasso Regression
LSVR	Linear Support Vector Regression
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MLR	Multiple Linear Regression
NE	Net Earnings
NMAE	Normalized Mean Absolute Error
NRMSE	Normalized Root Mean Square Error

PFI	Permutation Feature Importance
SD	Standard Deviation
SE	Simultaneous Engineering
SOP	Start Of Production
TAM	Technology Acceptance Model
TC	Target Cost
TCA	Target Cost Attainment
TTF	Task-Technology Fit
XAI	eXplainable Artificial Intelligence

Acknowledgment

First and foremost, I would like to thank my supervisor, Prof. Dr. Marc Wouters for the invaluable advice and guidance throughout this dissertation. His extraordinary engagement, recommendations, and thoughtful comments contributed greatly to this project.

I would also like to thank Mr. Jörg Felten for the many helpful practical suggestions and valuable insights in order to connect research and practice. I also want to thank Mr. Kai Siedlatzek for supervising the research project at the case company. I would also like to thank Dr. Frank Rösler for the initial support and setting off the research project. Further, I want to thank all participants from the case company who agreed to be interviewed and took part in the experiment.

I would also like to thank Mr. Michael Disch for the helpful comments and valued time while jointly undergoing the research project at the case company. Additionally, I want to thank my colleagues form the Institute of Management at KIT, Dr. Ana Mićković, Dr. Frank Stadtherr, Mr. Dominik Ebinger, Dr. Frederik Riar, and Ms. Nadine Fleischer, for their support and helpful advice to this dissertation.

I am deeply grateful to my parents for their unconditional support and encouragement throughout the years. I also appreciate all the support I received from the rest of my family.

Thanks also to my girlfriend Julia, for the unwavering support and the positive attitude that helped me overcome the little struggles throughout my studies.

1 Introduction

Cost management during new product development is one of the most important tasks of manufacturing companies and a principal task of management accounting. However, the management of costs becomes increasingly difficult due to higher complexity, modularity, and uncertainty during new product development (Gopalakrishnan et al., 2015; Stadtherr & Wouters, 2021). Thereby, direct material costs play a major role for new product development, as material cost usually have the largest leverage for optimizing the profitability of products (Fischer, 2017). Two important success factors of new product development performance are the tools and methods available. New tools that incorporate ***big data*** and ***machine learning***, yield promising benefits toward the management of costs (Chou et al., 2010; Fosso Wamba et al., 2015; Loyer et al., 2016). However, empirical research reporting on the actual applicability and realized benefits of these technologies is scarce. On the other hand, cost management methods, such as ***target costing***, can improve customer satisfaction and the cost reduction performance during product development (Dekker & Smidt, 2003). However, little is known about the impact of high product complexity and uncertainty on the effectiveness of these methods (Ax et al., 2008). Therefore, this dissertation involves two research foci, namely 1) the *applicability of big data and machine learning technology for cost management during new product development* and 2) the *impact of complexity and uncertainty on the effectiveness of the target costing approach*.

Global data is expected to grow exponentially by about 40% a year (McKinsey Global Institute, 2011). This avenue of large and complex data sets are referred to as big data, which can be characterized by its high volume, variety, velocity, and veracity (IBM, 2013). The usage of big data technology enables companies to change from mostly intuitive-based to data-based decision making (O'Leary, 2013). Further, the progress of machine learning opens new possibilities for processing information and creating additional value from data. On the other side, machine learning establishes algorithms that enable computers to learn by finding statistical regularities and patterns in data (Oladipupo, 2010). Machine learning and data analytics are considered the top two game-changing technologies for businesses (Carlton Sapp et al., 2019). Top performing companies use data analytics five times more than low performing companies (LaValle et al., 2011). New data sources and advanced analytical techniques are expected to yield great impact on the finance profession by increasing the influence in their businesses (Economist Intelligence Unit, 2013; Richins et al., 2017). In practice, however, companies often gather massive amounts of data without knowing how to make use of it (Earley, 2015; LaValle et al., 2011). In addition, complex machine learning techniques are often referred to as *black boxes* as they lack transparency and interpretability. Explainability is especially important when managing costs during new product development with its cross-functional nature (Cavalieri et al., 2004; Verlinden et al., 2008).

Target costing is one of the most important techniques to manage life-cycle costs during the product design and development stage (Kato, 1993). Target costs are calculated as the maximum product cost, which ensures meeting the profitability goals and customer requirements (Everaert et al., 2006). Therewith, the target costing method is a strategically important cost management method, which secures the competitiveness of manufacturing companies (Cooper & Slagmulder, 1999). In particular, the method improves the coordination of efforts within companies toward a joint cost goal (Ewert & Ernst, 1999). However, this approach may not enhance motivation and performance in a more complex and uncertain organizational product development context (A. Davila & Wouters, 2004; Gopalakrishnan et al., 2015; Henri & Wouters, 2020; Mihm, 2010). On the one hand, complex product development environment can correspond to large interdependencies between parts, which demand more coordination of shared resources from development teams. On the other hand, longer lead times of product development create uncertainty about future sales prices, product attributes, and costs. So far, little is known about the effectiveness of the target costing approach in such a complex product development context.

To investigate both research foci, we conduct a longitudinal case study at a German car manufacturer. The author was part of the controlling department full-time for three years. The case study provided access to large and sensitive data sources and cost experts with whom findings were evaluated. Case study research is appropriate in situations of asking *how* and *why* research questions and involves multiple sources of data such as interview data, archival data, and participant observation (Yin, 2018). The case study approach is often used to examine how techniques and processes from theory are applied in practice (Scapens, 1990). This dissertation thesis consists of 6 chapters as showcased in **Figure 1**. Chapter 2, 3, and 4 address the first research focus; chapter 5 covers the second one. The 4 main chapters are briefly introduced in the following sections.

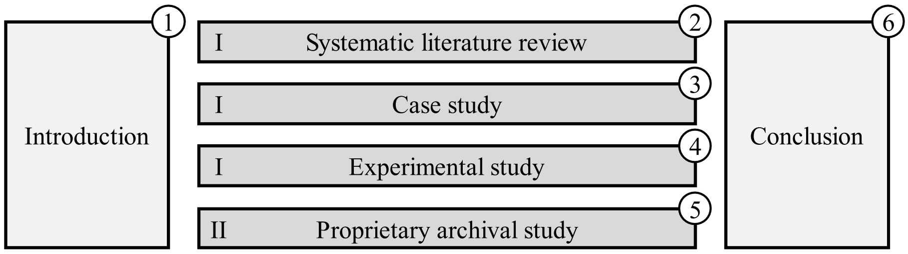


Figure 1: Overview of the six chapters that constitute the dissertation. The dissertation investigates two research foci (I and II).

In the second chapter, we conduct a systematic literature review on the use cases, issues, and benefits of big data and machine learning for management accounting. We aim to provide a better understanding of which managerial accounting tasks big data and machine learning technology can be applied on. We systematically review 36 articles, which deal with management accounting-related tasks. Content analysis is applied to identify categories about the issues and benefits of big data and machine learning for management accounting.

In the third chapter, we conduct a case study on the usage of big data and machine learning technology for the management of direct material cost of passenger cars. Specifically, we investigate the impact of big data on the cost estimation accuracy, compare the predictive accuracy of machine learning methods with manual cost calculations from cost experts, and examine the validity of cost-engineering insights such as cost driver selection and the cost behavior of product features. The results are discussed with cost experts from the case company to validate the results.

In the fourth chapter, we analyze the applicability of complex machine learning models during new product development. We evaluate the interpretability problem of machine learning techniques, which commonly describes the trade-off between accuracy and explainability. In an experiment with 40 participants from the controlling department of the case company, we investigate several factors that influence the importance of explainability relative to accuracy in the context of product cost estimation during product development.

In the fifth chapter, we analyze the impact of interdependencies between parts and uncertainty during new product development on the effectiveness of the target costing approach. The analysis is based on proprietary archival data from the case company. For several hundreds of parts and components, we examine the cost reduction performance and target cost attainment during product development and during the production stage.

This dissertation thesis contributes to the literature in several ways. In the second chapter, we confirm the revolutionary potential of big data and machine learning for management accounting and extend the body of literature by identifying new benefits and issues of both technologies in this context. In the third chapter, we empirically prove the benefit of big data for machine learning-based product cost estimation, introduce a novel method to improve multi-generational cost estimates, and empirically assess the validity of cost engineering insights of machine learning models. In the fourth chapter, we empirically show that the interpretability problem is indeed perceived by people, that the trade-off between explainability and accuracy has a major impact on the machine learning selection process and present four factors that determine the relative importance of explainability to accuracy. In the fifth chapter, we complement prior research by providing empirical evidence on the positive association between target cost difficulty and cost reduction performance on the basis of proprietary archival company data. We also add to the literature by showing that product design interdependence and uncertainty about target cost difficulty may limit the effectiveness of target costing.

2 Big Data and Machine Learning in Management Accounting: A Systematic Literature Review

Abstract

This systematic literature review analyzes use cases, issues, and benefits of machine learning and big data technology for the application in management accounting. Currently, the literature lacks in an overview about the use cases of big data and machine learning for managerial accounting and the specific issues and benefits that can be expected from both technologies in this context. First, this review contributes to the literature by providing an overview about the specific aspects of big data (volume, velocity, variety, and veracity) and machine learning methods that can be applied in managerial accounting. We find that both technologies can actually be applied on almost any of the main management accounting tasks. Thereby, several machine learning methods at different levels of complexity and any of the four big data aspects are used. Further, we find that both technologies are used for different tasks. Machine learning is primarily used for predictive tasks, while big data is mostly used for descriptive analyses. Simultaneous usage of both technologies is rather limited, which is surprising since both technologies benefit from each other. In the second stage of this review, we identify issues and benefits of both technologies in management accounting. Our systematic review finds that poor data quality and the lack of the management accountants' skills are the most critical issues of big data in managerial accounting. For machine learning the interpretability problem and the complex training process are considered problematic. Big data offers opportunities by providing new insights, better decision making and increasing the influence of accountants. In the case of machine learning, we find that higher accuracy, time-specific advantages, and greater independence from expert knowledge are the main benefits. The issues and benefits of both technologies highly depend on the specific management accounting task.

Keywords: Big data; Machine learning; Management accounting; Literature review

2.1 Introduction

Machine learning and big data technology are expected to have a significant positive impact on the performance of corporate management and are considered game-changers for businesses (Carlton Sapp et al., 2019; LaValle et al., 2011). The collection of massive amounts of data at high volume, velocity, variety, and veracity—short big data—can support management accountants in many aspects, such as decision making, visual analytics, and new insights (McAfee et al., 2012; Saggi & Jain, 2018). With internal and external data sources, management accountants can better conduct descriptive analytics (explaining the past), predictive analytics (forecasting the future) and prescriptive analytics (identifying optimal decisions) (Appelbaum et al., 2017). At the same time, more sophisticated machine learning models can be applied to improve corporate forecasting, planning tasks, and cost control (Coussement et al., 2015; Kuzey et al., 2019). Moreover, machine learning can automate repetitive tasks and thereby free up administrative capacities in companies (Gotthardt, M., Koivulaakso, D., Paksoy, O. et al., 2020; Losbichler & Lehner, 2021). Big data is usually an important enabler and an input for successful machine learning applications. The combination of machine learning and big data has the potential to transform the function of finance professionals and change their role within businesses (Economist Intelligence Unit, 2013; Richins et al., 2017). Thereby, accounting professionals are not required to act as technical specialists but they should understand the potential of big data and its implications on corporate decision making (Bhimani & Willcocks, 2014). In practice, however, many companies often collect massive amounts of data without knowing how to apply the data to operate their businesses (Earley, 2015; LaValle et al., 2011). Big data is a growing topic in business research that is however still in its infancy (Y. Zhang et al., 2021). Further, the expected benefits are set against the inherent problems of big data and machine learning such as data quality of big data sets and acceptance problems of complex data analytics techniques. Therefore, management accountants are often reluctant to apply machine learning and big data technology to operate their business (Arnaboldi et al., 2017; Losbichler & Lehner, 2021).

Currently, the literature lacks in an overview about the use cases for the management accounting profession and the specific issues and benefits that can be expected from these technologies on the various management accounting tasks. So far most literature reviews on this topic investigate related technologies such as business intelligence or data mining. Rikhardsson and Yigitbasioglu (2018) for example investigate applications of business intelligence and analytics however they mostly focus on the identification of research gaps. Nielsen (2018) discusses the application of business analytics for managerial accounting. Amani and Fadlalla (2017) investigate the utilization of data mining in accounting. In addition, current literature on big data often presents an optimistic view focusing mainly on the opportunities of the technology (Rikhardsson & Yigitbasioglu, 2018). Gärtner and Hiebl (2017) provide a literature review on both—benefits and issues—in management accounting, however, they primarily investigate the impact along technological aspects of big data without linking opportunities and challenges to specific management accounting tasks. Mardini and Alkurdi (2021) provide a brief overview of the impact of *artificial intelligence* (AI), which is closely related to machine learning, on managerial accounting, however the literature review lacks in a systematic categorization of use cases, benefits, or issues.

To address this gap, we provide a holistic overview of the use cases of big data and machine learning in managerial accounting. We address the specific issues and benefits that both technologies yield for various management accounting tasks. This chapter presents a systematic literature review to address two research questions: *What are use cases of big data and machine learning in management accounting? What are the issues and benefits of these technologies for managerial accounting?*

To answer the research questions a systematic literature review is conducted. We therefore search for relevant studies within the journals covered in the *Australian Business Deans Council* (ABDC) journal list. Content analysis is used to identify categories about the issues and benefits. Systematic literature reviews are commonly used to organize the complexity and variety of knowledge and aim to describe and evaluate the existing body of literature and identify knowledge gaps for further research (Tranfield et al., 2003). This literature review makes two contributions to the literature. First, we provide a comprehensive overview of the specific aspects of big data (volume, velocity, variety, and veracity) and machine learning (various methods) that can be used to solve the main management accounting tasks. Second, we identify and categorize issues and benefits in the context of managerial accounting. We facilitate deciding on the adoption of big data and machine learning by presenting various use cases and showcasing potential issues and benefits.

In the remainder of this chapter, we first provide comprehensive descriptions of management accounting, big data, machine learning, and related technologies to avoid any ambiguity on this topic. Section 2.3 explains the literature search method and the framework for content analysis. In Section 2.4, we review the literature and analyze the identified categories. Section 2.5 concludes the chapter.

2.2 Conceptualizing Management Accounting, Big Data, and Machine Learning

2.2.1 Management Accounting

In 1981, the Institute of Management Accountants characterized management accounting as a profession, responsible for the identification, accumulation, and analysis of financial data (Institute of Management Accountants, 2008). By now, management accounting emerged into a more strategic role acting as close business partners of management. This new role is described in the current definition by the Institute of Management Accountants (2008, p. 1):

> *"Management accounting is a profession that involves partnering in management decision making, devising planning and performance management systems, and providing expertise in financial reporting and control to assist management in the formulation and implementation of an organization's strategy."*

By providing information, supporting decisions, and acting as a business partner, managerial accounting is a central specialist to guide management (Langmann, 2019). The business orientation of management accounting covers the willingness and ability to provide more added value to the management by supporting decision-making and control (Järvenpää, 2007). Strategic management accounting involves an external perspective, the use of non-financial measures, and the usage of different time periods (Rom & Rohde, 2006). Driven by the ongoing digitization process, the transformation from a purely transaction-based perspective to a broader understanding of relevant information is taking place (Bhimani & Willcocks, 2014). Thereby, the application of extensive data mining and analytics enables accountants to take a larger strategic role within companies (Pickard & Cokins, 2015).

2.2.2 Big Data, Machine Learning, and Related Disciplines

In the following, we provide an overview of big data, machine learning, and related disciplines. ***Big data*** refers to large and complex data sets and was first characterized by the 3Vs: volume, variety, and velocity (Laney, 2001; McAfee et al., 2012). This characterization has been expanded by the dimension veracity, describing the uncertainty and credibility of data (IBM, 2013). More recently, also value, variability and visualization are considered.[1] According to the *HACE* theorem, big data "starts with large-volume, *heterogeneous*, *autonomous* sources with distributed and decentralized control, and seeks to explore *complex* and *evolving* relationships among data" (X. Wu et al., 2014, p. 98). However, the definition and classification of big data differs across various domains. Whether a data set can be regarded as big data depends on the capabilities of the information system (Vasarhelyi et al., 2015). Hence, there is no minimum size for data to be considered as *big*. ***Big data analytics*** describes the statistical analysis techniques deployed when dealing with large and complex data sets. Kwon et al. (2014, p. 387) defined big data analytics as "technologies (e.g., database and data mining tools) and techniques (e.g., analytical methods) that a company can employ to analyze large scale, complex data for various applications intended to augment firm performance in various dimensions".

Machine learning is about designing algorithms that allow a computer to learn by finding statistical regularities or other patterns in data sets (Oladipupo, 2010). Machine learning thereby enable computers to automatically improve through training examples (Horvitz & Mulligan, 2015). Machine learning is used to acquire knowledge on their own by discovering patterns from raw data (Goodfellow et al., 2016). Machine learning can be categorized into supervised and unsupervised learning (VanderPlas, 2016). Supervised learning involves the modeling of relationships between known features and a corresponding target variable. Common use cases for supervised machine learning methods are classification (discrete categories) and regression (continuous quantities). Unsupervised learning involves the identification of patterns in data sets without any target variable. The goal is to detect hidden structures and patterns in data sets, including clustering (identifying distinct groups within data) and dimensionality reduction (generate a more concise representation of data). Machine learning originates from artificial intelligence, statistics, and computer science, but it has established as a scientific discipline on its own (Luxburg & Schölkopf, 2011). A popular machine learning technique for many applications in practice and science are ***artificial neural networks*** (ANNs) (Hwang & Ding, 1997). Neural networks build complex representations of data

[1] Note that the definition of big data varies across researchers. As an example, Fosso Wamba et al. (2015) describe the 4Vs by volume + velocity + variety + value and the 5Vs by volume + velocity + variety + value + veracity.

inspired by the architecture of human brains. ANNs have been applied in many areas, including pattern matching, forecasting, and classification. The machine learning discipline is closely related to ***statistical learning***, which refers to techniques for understanding and evaluating complex data sets. It has emerged as a new subfield in statistics focusing on supervised and unsupervised modeling (James et al., 2013). In essence, statistical learning methods solve prediction and inference problems.

Artificial intelligence is characterized by the ability to learn from data and solve specific tasks (Kaplan & Haenlein, 2019). The ultimate achievement would be to develop a machine that can mimic or outperform human mental capabilities, including: understanding, reasoning, and recognition (Hopgood, 2005). Since there is no standard definition of intelligence, likewise there is much debate about the term artificial intelligence (Legg & Hutter, 2007). Commonly artificial intelligence refers to human-like thinking and acting (Russell & Norvig, 2016). In practice, artificial intelligence-based systems often lack in explainability and are being regarded as *black boxes*. This key impediment triggered the development of *explainable artificial intelligence* (XAI). This research field aims to make outcomes from artificial intelligence systems more understandable to humans and improves trust into artificial intelligence-based systems (Adadi & Berrada, 2018).

In the following, we introduce three closely related disciplines on this topic: data mining, data science, and data analytics. We incorporate similar technologies to enable diffusion on this topic. ***Data mining*** uses historical data to discover patterns and improve decision making (T. M. Mitchell, 1999). Data mining can be depicted as computer automated exploratory data analysis of data sets with high complexity and volume (Friedman, 1997). ***Data science*** is an interdisciplinary field, which turns data into value for businesses (van der Aalst, 2016). Value can be provided in many ways, including: visualizations that provide new insights, forecasting models, and automated decision-making systems. The data science discipline covers data storage, extraction, preparation, exploration, computing infrastructures, visualizations, analytics, and the utilization of findings in businesses. ***Data analytics*** is a highly interdisciplinary field incorporating aspects from many other scientific areas such as statistics, machine learning, pattern recognition, operations research, and artificial intelligence, which describes the analysis of large data sets with computer systems in order to support decision making (Runkler, 2016).

2.3 Method

2.3.1.1 Literature Search

To analyze the intersection between both technologies and management accounting, we conduct a systematic literature review. The literature selection process is based on the query-based approach of Saggi and Jain (2018). The literature search incorporates two sets of journals. First, we incorporate studies exclusively published in accounting journals. Second, we include studies published in business journals dealing with management accounting-relevant problems. The selection of the accounting and business journals is based on the ABDC journal quality list from 2019 (ABDC, 2019). The accounting journals are selected according to the ABDC code 1501 with an ABDC rating of A*, A, and B. The sample of business journals is based on all journals in the ABDC list with a rating of A*. The accounting journals already covered in the first sample are omitted from the second set of journals. In so doing, we obtain 58 accounting journals and 182 business journals.

Next, the sample of accounting (A) and business (B) journals are searched for relevant articles. Therefore, we formulate three search term queries. The queries are numbered as (1) “big data”, (2) “machine learning”, “artificial intelligence”, “neural net”, and “statistical learning”, and (3) “data science”, “data analytics”, and “data mining”. The third query is applied to also incorporate related disciplines and cast the net as widely as possible. We additionally include the search term “case study” in the query of business journals, as we are specifically interested in practical applications of big data and machine learning. **Table 1** provides an overview of the literature search queries.

Table 1: Literature search queries

Accounting literature (A)		
A1	**A2**	**A3**
"big data"	"machine learning" *AND* "artificial intelligence" *AND* "neural net" *AND* "statistical learning"	"data science" *AND* "data analytics" *AND* "data mining"
Business literature (B)		
B1	**B2**	**B3**
"big data" *AND* "case study"	("machine learning" *AND* "artificial intelligence" *AND* "neural net" *AND* "statistical learning") *AND* "case study"	("data science" *AND* "data analytics *AND* "data mining") *AND* case study"

The literature search has been carried out on the Scopus citation and abstract database (Elsevier B.V., 2019). With 22,800 titles from more than 5,000 publishers, Scopus is "the largest abstract and citation database of peer-reviewed literature" in the fields of science, technology, and social science (Elsevier B.V., 2017, p. 3). We search for matching search terms within the article's title, abstract, and keyword list. By combining the two journal samples and the three search term queries, we obtain six literature queries (A1, A2, A3, B1, B2, and B3). The execution of the queries A1, A2, and A3 on Scopus returned 47, 57, and 44 papers, respectively. The execution of the queries B1, B2, and B3 returned 30, 204, and 83 papers.

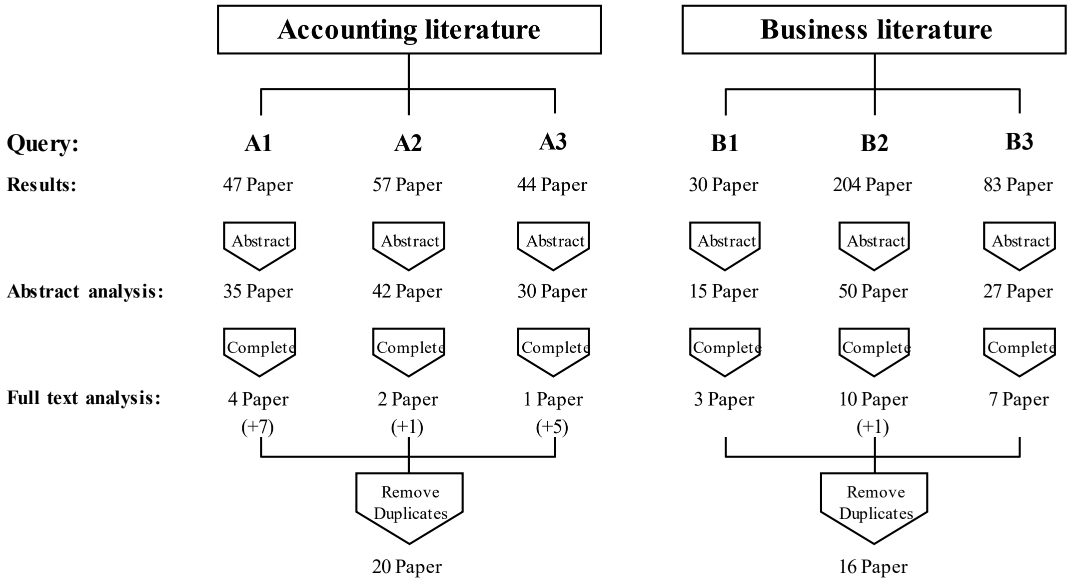


Figure 2: Paper selection process. Parentheses specify secondary studies. Representation adopted from Saggi and Jain (2018).

We select relevant articles according to the inclusion criteria adopted from Amani and Fadlalla (2017). An article is included when it 1) describes a specific use case of big data or machine learning in a managerial accounting-related task, 2) specifies what big data or machine learning techniques have been utilized, and 3) discusses issues or benefits of these technologies. Since we are also interested in potential use cases, we additionally include conceptional papers such as interpretive articles, commentaries, and literature reviews. We exclude studies in civil engineering or public management as we focus on an industry-specific context. The selection process is depicted in **Figure 2**. First, the results are filtered by abstract reading. The remainder is selected by full text reading, resulting in 20 studies from the accounting journal set and 16 articles from the set of business journals. The sample includes 14 secondary studies obtained by forward and backward search of the primary studies. **Table 2** provides an overview of the journals that research the application of big data and machine learning in managerial accounting. A large proportion of articles (28%) is published in the *International Journal of Production Economics* (IJPE).

Table 2: List of journals/sources included in this literature review. The asterisk indicates journals/sources obtained by forward and backward search.

No.	Journal/Sources	Abbreviation	Count
1.	*Accounting and Business Research*	*ABR*	1
2.	*Accounting Horizons*	*AH*	1
3.	*Association of Chartered Certified Accountants**	*ACCA*	1
4.	*Chartered Global Management Accountant**	*CGMA*	2
5.	*Computers & Industrial Engineering*	*CIE*	1
6.	*Contemporary Accounting Research*	*CAR*	1
7.	*Decision Support Systems*	*DSS*	4
8.	*European Journal of Operational Research*	*EJOR*	2
9.	*Harvard business review*	*HBR*	1
10.	*IMA Journal of Management Mathematics**	*IMA JMM*	1
11.	*Information & Management*	*IM*	1
12.	*International Journal of Accounting & Information Management**	*IJAIM*	3
13.	*International Journal of Accounting Information Systems*	*IJAIS*	1
14.	*International Journal of Agile Manufacturing**	*IJAM*	1
15.	*International Journal of Hospitality Management*	*IJHM*	1
16.	*International Journal of Production Economics*	*IJPE*	10
17.	*Journal of Emerging Technologies in Accounting*	*JETA*	1
18.	*London: Economist Intelligence Unit**	*EIU*	1
19.	*MIT sloan management review*	*MIT SMR*	1
20.	*Neurocomputing**	*Neuroc*	1

We observe a sharp increase in literature in the past 3 decades (**Figure 3**). While in the nineties three articles were published, two decades later the body of literature has increased by a factor of seven.

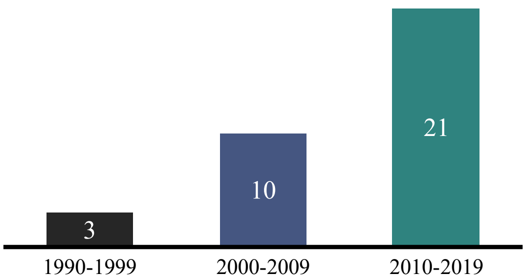


Figure 3: Increase in literature of big data and machine learning applications in managerial accounting in the past 30 years. Two papers from 2020 are discarded in this illustration.

An overview of the research methods on this topic is provided in **Figure 4**. The characterization of research methods is based on the classification of Wouters and

Morales (2014). The assignment of research methods per article is summarized in **Table 6** in Appendix A. The analysis shows that most research relies on archival or market data, while there is hardly any field research on this topic.

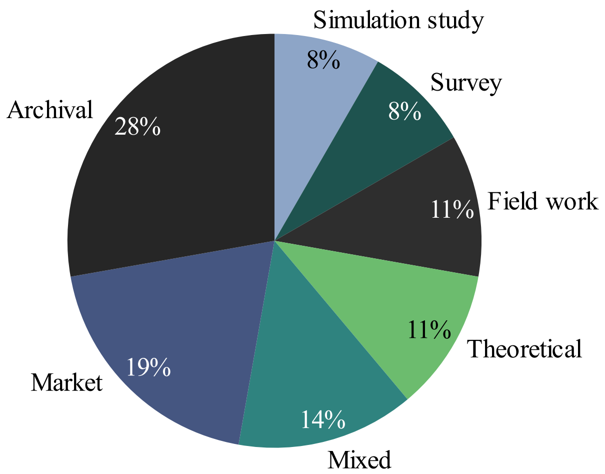


Figure 4: Overview of research methods for the 36 articles

2.3.1.2 Content Analysis and Organizing Framework

Content analysis is conducted to investigate use cases, issues, and benefits of big data and machine learning in management accounting. To analyze the articles, we apply ***deductive category development*** for the big data and machine learning techniques and the management accounting tasks. The big data categorization is based on the 4Vs (volume, velocity, variety, and veracity) (IBM, 2013). If an article utilizes social media data as the main data source, we consider all four Vs to be satisfied. Usually, social media data involves huge amounts of data that are produced in real-time at large variety and veracity. The categorization of machine learning techniques is based on the classification scheme of Brownlee (2019). The categorization of management accounting tasks is based on the main processes of the *Controlling Process Model* (International Group of Control, 2012). As we expect cost estimation to be a major use case of machine learning and big data, we introduce a separate category for this task. Cost estimation can be depicted as a sub-category of cost accounting. The categorization of issues and effects of big data and machine learning is based on ***inductive category development***. The categories of the content analysis are summarized in **Table 3**. The classification of each article according to the coding scheme is presented in **Table 4**. For each article a brief summary regarding the research focus is provided.

Table 3: Categories for content analysis

Use Case (*deductive approach*)	
Big Data	Volume (VOL), velocity (VEL), variety (VAR), and veracity (VER).
Mach Learn.	Regression analysis (REGR), instance-based algorithms (INST), decision tree algorithms (TREE), Bayesian methods (BAYES), artificial neural networks (ANN), and other techniques (OTHER).
Mgmt. Acc.	Strategic planning (STR.PLAN), operational planning and budgeting (OPR.PLAN), forecasting (FORECAST), cost accounting (COST.ACC), cost estimation (COST.EST), management reporting (REPORT), project and investment controlling (INVEST), risk management (RISK.MGT), and management support (MGT.SUPP).
Issues (*inductive approach*)	
Big Data	Poor data quality (QUALITY), lacking skills and know how (SKILL), privacy issues (PRIVACY), information overload (OVERLOAD), and problems with the preprocessing of raw data (PREPROC).
Mach Learn.	Interpretability problem (INTERPRET), issues related to the training process (TRAINING), and (hyper-)parameter setting (PARAMS).
Benefits (*inductive approach*)	
Big Data	New insights (INSIGHT), better corporate decision making (DEC.MAK), and increasing influence of management accountants (INFLUENCE).
Mach Learn.	Accuracy of outcomes (ACCURACY), time-related advantages (TIME), independence from expert knowledge (INDEP), and new insights (INSIGHT).

Table 4: Use cases, issues, and benefits of big data (BD) and machine learning (ML) in management accounting

Author (Year)	Summary (regarding research focus)	BD Technique	ML Technique	Use Case	BD Issues	ML Issues	BD Benefits	ML Benefits
Appelbaum et al. (2017)	This paper introduces a framework for integrating data analytics and business intelligence into the management accounting task, focusing on decision making and performance measurement.	VOL, VEL, VAR, VER		COST.ACC, REPORT, MGT.SUPP	QUALITY, PRIVACY			
Bhimani and Willcocks (2014)	This article discusses potentials and challenges of big data for the finance function and, specifically, for the preparation of management accounting information.	VEL, VAR		MGT.SUPP	QUALITY		INFLUENCE	
CGMA (2013)	The report examines challenges, opportunities, and implications of big data technology for businesses, by conducting interviews and a survey with CFOs and finance experts.	VOL, VEL, VAR		COST.ACC	QUALITY, SKILL		INFLUENCE	
CGMA (2016)	A survey is conducted to analyze the potentials of big data for strategic decision making in today's uncertain, vague, volatile, and complex business environment.	VOL, VEL, VAR		STR.PLAN	QUALITY, OVERLOAD			
Chua (2014)	This report provides an overview of the challenges, potentials, and effects of big data on the finance and accounting function in the coming 5 to 10 years.	VOL, VEL, VAR		MGT.SUPP, RISK.MGT, REPORT	PRIVACY		INSIGHT, DEC.MAK, INFLUENCE	
Economist Intelligence Unit (2013)	The article examines the effects of the increasing amount of data within companies on the finance function and, in particular, CFOs, building on a survey and interviews with business professionals.	VOL		FORECAST, OPR.PLAN, RISK.MGT, REPORT	QUALITY, SKILL, OVERLOAD		INFLUENCE	

Table 4: (Continued)

Author (Year)	**Summary** (regarding research focus)	BD Technique	ML Technique	Use Case	BD Issues	ML Issues	BD Benefits	ML Benefits
Green et al. (2018)	Big data will increase the demand for more atomized (not aggregated) and transparent forms of accounting information to gain more flexibility when conducting analyses and making managerial decisions.	VEL, VAR		STR.PLAN, REPORT, MGT.SUPP	SKILL		DEC.MAK	
Holton (2009)	This study uses text mining techniques to detect disgruntled employee communications. The model can support fraud risk prediction to realize and avert fraudulent behavior.	VOL,VEL, VAR,VER	BAYES	RISK.MGT	PRIVACY			
LaValle et al. (2011)	The article explores the issues and potentials of business analytics for companies by conducting a survey with 3,000 business professionals and analysts.	VOL, VAR		OPR.PLAN	SKILL		DEC.MAK	
McAfee et al. (2012)	This research investigates the effect of big data on corporate performance by analyzing annual reports and interviews of over 300 companies.	VOL, VAR, VEL		MGT.SUPP			DEC.MAK	
Warren et al. (2015)	This paper describes the usage of new types of data, such as video and audio files, for the accounting profession. In managerial accounting, the application of big data will improve budgeting and management control.	VOL, VAR, VEL		REPORT, OPR.PLAN	QUALITY, SKILL			
Desai (1997)	This paper applies genetic algorithms and artificial neural networks to predict loan credit-scores (good, poor, bad) of customers.		ANN, OTHER	FORECAST		PARAMS		
Kuzey et al. (2019)	This research uses several machine learning models, including logistic regression, support vector machines, and decision trees, to analyze influencing factors of cost system functionality (i.e., accuracy, detail of cost data).		TREE, INST, REGR	COST.ACC				ACCURACY, INSIGHT

Table 4: (Continued)

Author (Year)	**Summary** (regarding research focus)	**BD Technique**	**ML Technique**	**Use Case**	**BD Issues**	**ML Issues**	**BD Benefits**	**ML Benefits**
Liang et al. (1992)	This paper analyzes the application of machine learning models for the selection between LIFO (last in, first out) and FIFO (first in, first out) accounting methods.		REGR, TREE, ANN	OPR.PLAN		INTERPRET		
Bansal, K., Vadhavkar, S. and Gupta (1998)	This article explores the application of artificial neural networks for the optimization of inventory management of distribution companies.	VOL	ANN	OPR.PLAN		TRAINING, PARAMS		
Chou et al. (2010)	This study compares the accuracy of various machine learning models for the cost prediction of TFT-LCD (thin-film transistor liquid-crystal display) manufacturing equipment.		REGR, ANN, OTHER	COST.EST		INTERPRET, TRAINING, PARAMS		ACCURACY, TIME
Kostakis et al. (2008)	This research paper combines simulation modeling and association rule mining to define cost drivers for activity-based costing systems.		OTHER	COST.ACC				INSIGHT, ACCURACY
Partovi and Anandarajan (2002)	This article describes the application of artificial neural networks for the classification of stock-keeping units according to the ABC ranking scheme.		ANN	OPR.PLAN		TRAINING		ACCURACY
Shawver (2005)	This study analyzes the usage of artificial neural networks for the prediction of bank merger premiums.		ANN	INVEST		INTERPRET		ACCURACY, TIME
Tang (2009)	The case study examines the application of artificial neural networks and fuzzy logic theory to improve corporate budget allocation.		ANN	OPR.PLAN		TRAINING		

Table 4: (Continued)

Author (Year)	**Summary** (regarding research focus)	BD Technique	ML Technique	Use Case	BD Issues	ML Issues	BD Benefits	ML Benefits
Qi et al. (2016)	This research uses online reviews of products to retrieve customer requirements to improve new product development.	VOL, VEL, VAR, VER		MGT.SUPP	QUALITY		INSIGHT	
Tan et al. (2015)	This paper introduces an analytical infrastructure based on deduction graph techniques to support managerial decision making.	VOL, VEL, VAR, VER		MGT.SUPP	QUALITY, OVERLOAD, PREPROC		INSIGHT	
Zhan and Tan (2020)	This study proposes an infrastructure for big data to support operational decision making and product development.	VOL, VEL, VAR, VER		MGT.SUPP	PREPROC		INSIGHT	
Caputo and Pelagagge (2008)	This paper compares parametric and neural network methods for the estimation of manufacturing cost of pressure vessels.		ANN	COST.EST		INTERPRET		ACCURACY, TIME, INDEP
Cavalieri et al. (2004)	This paper compares parametric and neural network methods for estimating manufacturing cost of automotive brake disks.		ANN	COST.EST		INTERPRET		ACCURACY, TIME, INDEP
Loyer et al. (2016)	This paper compares several machine learning techniques for the estimation of manufacturing costs of jet engine components.		REGR, ANN, INST, OTHER	COST.EST		TRAINING, INTERPRET		ACCURACY, INSIGHT
K. Coussement et al. (2015)	The paper proposes a Bayesian framework for integrating expert knowledge into decision support systems in the context of customer satisfaction prediction.	VOL, VEL, VAR, VER	BAYES	FORECAST				ACCURACY, INDEP
Deng and Yeh (2011)	This paper compares neural networks, regression models, and support vector machines for estimating manufacturing cost of airframe structural parts.		INST, ANN, REGR	COST.EST		PARAMS		
Golmohammadi (2011)	This article proposes a decision-making model integrating fuzzy logic and neural networks to support the managerial decision-making process.		ANN	MGT.SUPP				INDEP

Table 4: (Continued)

Author (Year)	**Summary** (regarding research focus)	**BD Technique**	**ML Technique**	**Use Case**	**BD Issues**	**ML Issues**	**BD Benefits**	**ML Benefits**
Gruss et al. (2018)	The paper introduces a novel text mining approach based on social media data to support the reporting on product quality.	VOL, VEL, VAR, VER	BAYES	REPORT	QUALITY	INTERPRET		ACCURACY
Hua Tan et al. (2006)	The article combines case-based reasoning with neural networks to support investment decision making on new manufacturing technologies.		ANN, OTHER	INVEST				INDEP, INSIGHT
Verlinden et al. (2008)	This paper compares regression analysis and artificial neural networks for the estimation of manufacturing cost of sheet metal parts.		REGR, ANN	COST.EST		INTERPRET		ACCURACY, TIME
Q. Wang (2007)	This study proposes a cost model approach based on artificial neural networks for collaborating manufacturing companies.		ANN	COST.EST				ACCURACY
S. Wu and Akbarov (2011)	The article explores the application of support vector regression models for predicting warranty claims, supporting the financial planning of manufacturers and warranty providers.		ANN, INST	FORECAST, OPR.PLAN				ACCURACY
Kristof Coussement et al. (2017)	The paper analyzes alternative data preparation techniques to improve the performance of regression-based churn prediction models.	VOL	REGR, INST, TREE, BAYES, ANN, OTHER	FORECAST		INTERPRET, TRAINING		
H. Xia et al. (2020)	This study uses online reviews to evaluate the competitiveness and unique features of hotel brands or other branded products in the tourism sector.	VOL, VEL, VER, VAR		REPORT			INSIGHT	

2.4 Application of Big Data and Machine Learning in Management Accounting

2.4.1 Use Cases

In the following, we provide an overview of the use cases of big data and machine learning on the main management accounting tasks. The first category covers the ***strategic planning process***. The usage of big data in combination with profound analytics tools provide companies with new insights, creating competitive advantages in their strategic planning (CGMA, 2016). Specifically, the provision of more atomized and reconfigurable accounting information can be used to improve strategic decision making (Green et al., 2018). The second category covers the ***operational planning and budgeting task***. Machine learning methods, such as neural networks and ID3 algorithms, can support the choice between LIFO (last in, first out) and FIFO (first in, first out) accounting methods (Liang et al., 1992). Further, artificial neural networks are able to optimize the inventory management of distribution companies by learning from past operations and decisions (Bansal, K., Vadhavkar, S. & Gupta, 1998). ANNs can also support the classification of stock-keeping units according to the ABC ranking scheme to improve the allocation of material resources (Partovi & Anandarajan, 2002). Tang (2009) combined artificial neural networks and fuzzy logic theory to improve corporate budget allocation. Finally, machine learning can be applied to estimate the number of warranty claims to support the financial planning of manufacturers and warranty providers (S. Wu & Akbarov, 2011). The third category addresses the ***forecasting*** task. The largest positive impact of large amounts of data from the perspective of chief financial officers lies in scenario planning and forecasting (Economist Intelligence Unit, 2013). Coussement et al. (2015) introduces a Bayesian framework for integrating expert knowledge into decision support systems in the context of customer satisfaction forecasting of products or services. Coussement et al. (2017) explores several data preparation techniques to improve the performance of logistic regression models compared to state-of-the-art machine learning techniques for the prediction of customer churn. In addition, genetic algorithms and artificial neural networks were applied for predicting loan credit-scores (good, poor, bad) of customers (Desai, 1997). The fourth category addresses ***cost accounting***. New types of data, where much stems from external sources, enable companies to identify cost reduction potentials (CGMA, 2013). Kuzey et al. (2019) uses several machine learning models, such as logistic regression, support vector machines, and decision trees, to analyze influencing factors of cost system functionality (i.e., accuracy, detail of cost data) and optimize the design of cost systems. Finally, the combination of simulation modeling and association rule mining can be applied to define cost drivers for activity-based costing systems (Kostakis et al., 2008). The fifth category covers ***cost estimation***. Chou et al. (2010) applies regression models and neural networks on cost forecasting of liquid-crystal display equipment. Deng and Yeh (2011) use support vector machines, regression models, and neural networks to estimate manufacturing costs for airframe structural parts. Further, machine learning techniques have been applied on the estimation of manufacturing cost of jet engine components, providing relevant insights for cost engineers such as cost driver ranking and partial dependence plots for cost drivers (Loyer et al., 2016). Cavalieri et al. (2004) compares parametric and neural network methods for estimating manufacturing cost of automotive brake disks to optimize product concepts during the early stages of product development. Regression analysis and artificial neural networks were used for the estimation of manufacturing cost of sheet metal parts to obtain first cost forecasts during the early pricing stages (Verlinden et al., 2008). The sixth category deals with the ***management reporting*** task of managerial accounting. Gruss et al. (2018) introduces a novel text mining approach based on social media postings to support the management reporting on product quality. Big data in the form of online reviews can be applied in the reporting of competitiveness of hotel brands and other branded products (H. Xia et al., 2020). New types of data, such as video and audio files can be used in managerial accounting to improve management control systems by incorporating better performance measures (Warren et al., 2015). The seventh category covers ***project and investment controlling***. Artificial neural networks can be applied for the estimation of bank merger premiums to improve corporate investment decisions (Shawver, 2005). Hua Tan et al. (2006) combines case-based reasoning with neural networks to support investment decision making on new manufacturing technologies. The eighth category encompasses ***risk management***. Big data can support accountants in the area of risk management by real-time risk identification and seeing the bigger picture (Chua, 2013). Holton (2009) applies text mining techniques to detect disgruntled employee communications, supporting fraud risk prediction and prevention. Finally, the ninth category addresses ***management support***. Big data can be used for the preparation of management information incorporating highly different data sources (Bhimani & Willcocks, 2014). Thereby, the usage of big data enables managers to measure and learn more about their companies (McAfee et al., 2012). Tan et al. (2015) introduces an analytical infrastructure based on deduction graph techniques to support managerial decision making by providing important product requirements. Golmohammadi (2011) proposes a decision-making model, integrating fuzzy logic and neural networks to support the managerial decision-making process.

2.4.2 Issues

In the following, we describe the identified issues of big data applications in management accounting. The first issue of big data is related to the poor ***data quality***. Reliable data quality and integrity is an important requirement for the usefulness of big data applications (Appelbaum et al., 2017; Warren et al., 2015). However, large data sets are often disorganized and difficult to manage (Bhimani & Willcocks, 2014). For example, social media posts are difficult to work with due to abbreviations and different formatting (Gruss et al., 2018). In addition, the reliability of such information is often problematic (Tan et al., 2015). The second category addresses the required ***skills*** and know-how in order to operate both technologies. The adoption of big data technology is often limited by the lacking knowledge of retrieving valuable information from massive amounts of data (Warren et al., 2015). Most business professionals have difficulties getting insights from big data and work with non-financial data sources (CGMA, 2013). The adoption of big data and the development toward a data-driven company is often inhibited by the lack of internal skills and the understanding to use analytics for business operations (LaValle et al., 2011). However, teaching accountants how to work with unfamiliar data sources and integrating IT in the accounting curriculum is challenging (Green et al., 2018). The third category covers ***privacy issues***. Integrating and sharing many data sources throughout a company can be the cause of privacy and security problems (Appelbaum et al., 2017). The risks of privacy and ethical problems are difficult to measure in big data applications (Chua, 2013). Hence, big data systems must be implemented with care to prevent privacy violations (Holton, 2009). The fourth category is ***information overload***. Providing too much information at higher degrees of variety reduces the usability for effective decision making (CGMA, 2016). When applying big data, often much effort and time is required to identify relevant information (Tan et al., 2015). The fifth category deals with complex ***preprocessing*** steps. It is difficult to integrate information from different sources (i.e., customer feedback, website data) and utilizing the separate pieces of information for decision making (Tan et al., 2015). Further, there are certain data formats, such as videos and images, that yield potential information, but they are generally difficult to process and interpret (Zhan & Tan, 2020).

In the following, we provide an overview of the identified issues when applying machine learning in management accounting. The first category deals with the

interpretability problem. A major limitation of machine learning is the lack of interpretability and transparency, which is necessary to improve the model and utilize the output in practice (Cavalieri et al., 2004). The perception of machine learning models as a black box leads to a reduction in trust into the model and its results (Liang et al., 1992). One example of a hard to interpret and opaque machine learning models are artificial neural networks (Chou et al., 2010). In cost estimation, neural networks do not provide any direct information about the association between input and output variables that can be used to understand cost relationships and identify cost drivers (Caputo & Pelagagge, 2008). Therefore, cost engineers still prefer more transparent and interpretable cost estimation models for their explanatory power (Verlinden et al., 2008). The second category deals with the complex ***training*** process of machine learning. Complex machine learning models require many hours to train (Chou et al., 2010). Further, machine learning models are often limited to numerical representations only. Hence, the inclusion of important qualitative features into a model is challenging (Partovi & Anandarajan, 2002). Massive training examples are required when training machine learning models; if the size of a training set is not large enough, the machine learning algorithm might not solve the task sufficiently (Tang, 2009). The final category covers the (hyper-)***parameter*** setting problem. Selecting the right machine learning model for a task and adjusting the parameters is a manual and iterative process, where developers sometimes go through thousands of versions until identifying an adequate model (Bansal, K., Vadhavkar, S. & Gupta, 1998). Deng and Yeh (2011) states that the adjustment of parameters is the most critical problem in neural networks, where the predictive performance is directly dependent on the parameter selection. Often heuristic approaches are necessary, where the identification of the most appropriate model with an optimized set of parameters however cannot be ensured (Desai, 1997).

2.4.3 Benefits

In this section, we provide an overview of the identified positive effects of big data and machine learning applications in management accounting. The first category of benefits covers the gathering of new ***insights***. Big data can help manufacturers to better understand their customers and identify important customer requirements, which can be used to improve product designs (Qi et al., 2016). Further, the application of big data enables manufactures to get new ideas for new product development (Tan et al., 2015). Additional insights from big data can also be used to get a holistic perspective about the operational capabilities of a company (Zhan & Tan, 2020). The second category covers benefits related to better ***decision making***. Big data enables decision making based on facts instead of relying only on personal judgment (LaValle et al., 2011). McAfee et al. (2012, p. 6) state that "data-driven decisions are better decisions" as managers can judge based on proof instead of hunches. Therefore, recipients of accounting data will increasingly demand for raw data, which provide more flexibility and autonomy when conducting analyses according to the specific decision-making needs (Green et al., 2018). Further, the velocity aspect of big data can provide real time insights, which can support time critical decision making processes (Chua, 2013). The third category deals with an increasing ***influence*** of accountants. Big data and more sophisticated data analytics can transform finance professionals into information specialists (Bhimani & Willcocks, 2014). Therefore, management accountants are required to partner more closely with the IT department (data collection), data scientists (data analytics) and managers (transforming insights into actions) (CGMA, 2013). Finance and accounting managers see a strategic opportunity in the application of big data leading to an increase of their influential power within companies (Economist Intelligence Unit, 2013).

The first category of benefits toward the application of machine learning in management accounting is the increasing ***accuracy*** of outcomes. Results from machine learning models proof to have higher predictive accuracy than conventional techniques in the classification of stock-keeping units according to the ABC ranking system (Amani & Fadlalla, 2017). Similarly, the application of artificial neural networks significantly increases the predictive accuracy when estimating merger premiums compared to simple linear regression models (Shawver, 2005). Further, the utilization of machine learning models lead to an increase in predictive accuracy in the context of product cost estimation (Caputo & Pelagagge, 2008; Cavalieri et al., 2004; Loyer et al., 2016). The second category covers ***time***-specific advantages. Machine learning enables product managers to produce sufficiently accurate product cost estimates in the early phase of new product development, which support the first negotiations with clients and reduces cost changes at later stages (Chou et al., 2010). Further, machine learning models estimate product costs much leaner and faster than more traditional methods (Cavalieri et al., 2004). Verlinden et al. (2008) state that in the sheet metal industry the application of machine learning can be beneficial, since fast product cost estimates are necessary as customers often expect price quotations right away. The third category covers benefits regarding the ***independence*** of machine learning systems from expert knowledge. Caputo and Pelagagge (2008) point out that the utilization of ANNs saves time and expenses of cost experts as supervised learning algorithms are independent from prior set cost relationships. In cost management the results of machine learning models are independent from the abilities of cost experts, selecting the most important cost drivers and defining cost estimation relationships (Cavalieri et al., 2004). Golmohammadi (2011) claims that artificial neural networks can be used for many multi-attribute decision making problems without personal judgment from management. Coussement et al. (2015) point out that the integration of machine learning methods into decision support systems is beneficial as such systems can operate mostly independently from expert opinions. On many occasions the manual evaluation of problems would be too expensive and time consuming. The final category covers new ***insights*** gained from machine learning models. Machine learning can be used to gather robust and reliable insights of influencing factors for cost system functionality, including the importance and sensitivity of variables (Kuzey et al., 2019). Kostakis et al. (2008) used association rule mining to gain insights into the associations between cost drivers for activity-based costing systems. The findings can be used to develop latent cost drivers for more difficult-to-measure cost drivers. Loyer et al. (2016) used machine learning to gain relevant insights for cost engineers; for example, machine learning can be used to rank cost drivers according to their importance and approximate the functional relationship between cost drivers and manufacturing costs. Machine learning systems can support the prioritization of investment decisions by systematically providing insights about past projects and experiences (Hua Tan et al., 2006).

2.4.4 Category Frequencies and Analysis

In the following, we analyze the category frequencies of both technologies, the management accounting tasks, and the identified issues and benefits. **Figure 5** depicts the number of articles for each category in the sample of literature. The percentage frequencies refer to the relative number of articles within a category group.

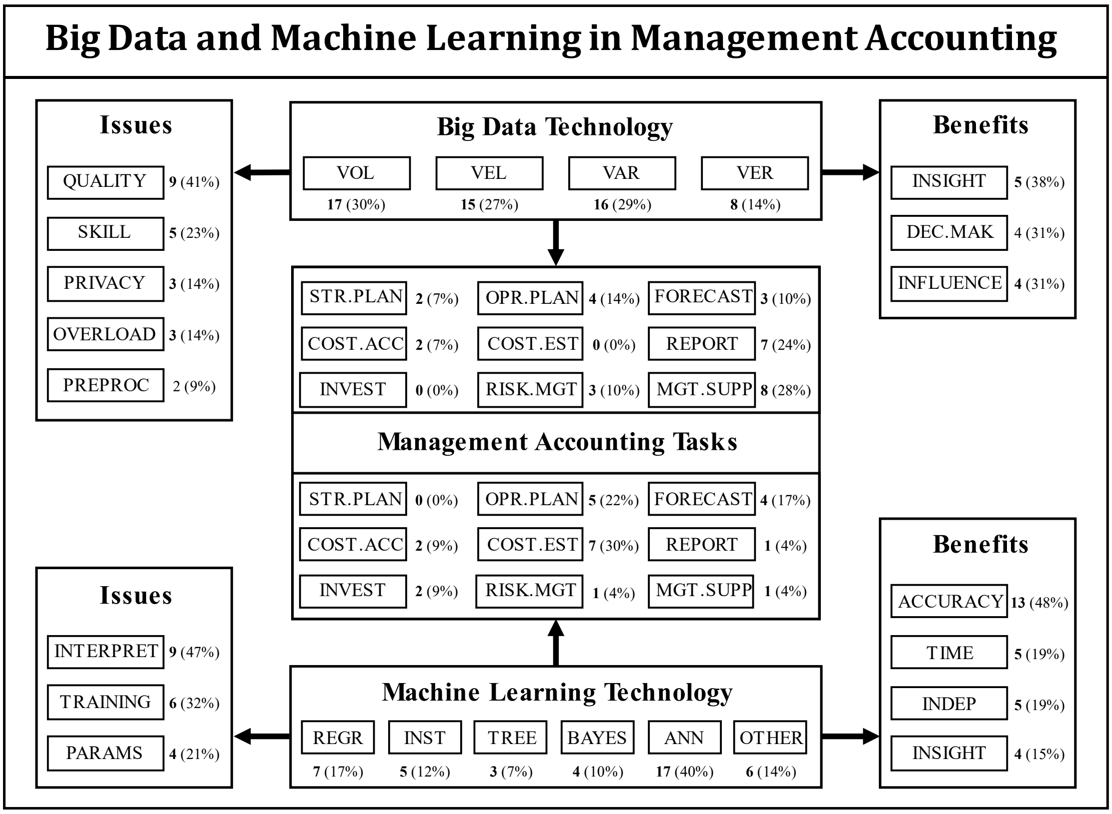


Figure 5: Overview of big data and machine learning application in managerial accounting

2.4.4.1 Use Cases of Big Data and Machine Learning in Management Accounting (Research Question 1)

The content analysis shows that big data and machine learning technology can be used to support all major management accounting tasks. All four aspects of big data (4Vs) can be deployed in managerial accounting. Thereby, big data is mostly specified by high volume (30%), variety (29%), and velocity (27%). Several machine learning methods, either classification or regression, at different levels of complexity can be deployed. The most common machine learning techniques are the rather complex artificial neural network (40%) followed by the rather simple regression model (17%). **Figure 6** depicts the number of articles describing use cases of big data, machine learning and the simultaneous usage of both technologies on the nine management accounting tasks. The deployment of big data and machine learning technology strongly depends on the corresponding management accounting task. Big data is primarily applied for *management support*, *management reporting*, and *operational planning and budgeting*. For tasks that involve predictive analytics such as *cost estimation*, *forecasting*, and *operational planning and budgeting* primarily machine learning is used. The simultaneous deployment of big data and machine learning is scarce. From 36 articles only 5 (14%) combined machine learning and big data technology. **Figure 6** also shows that most applications of big data and machine learning in management accounting have been investigated with empirical research.

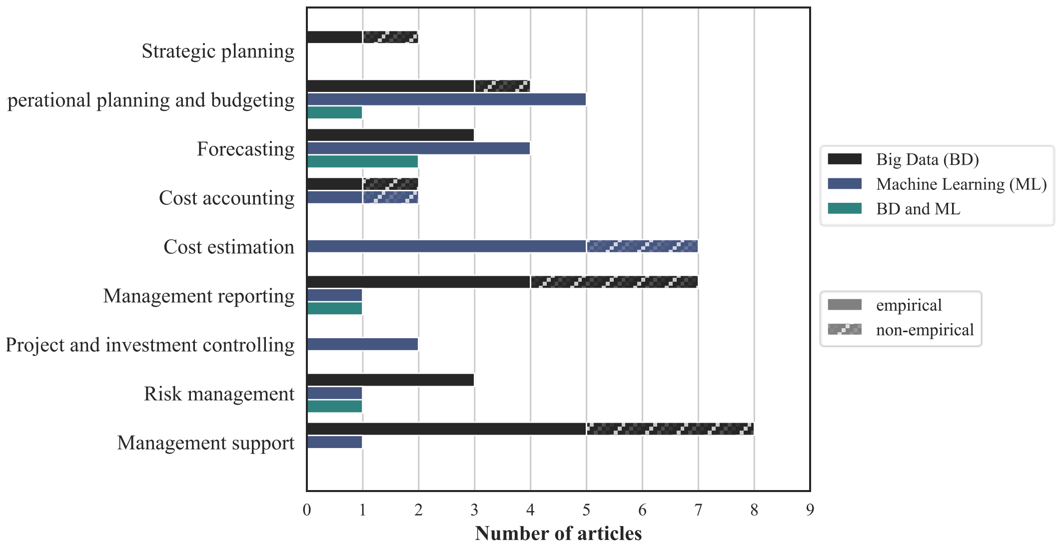


Figure 6: Number of articles that describe applications of big data and machine learning in management accounting. Non-empirical articles are illustrated separately.

2.4.4.2 Issues and Benefits of Big Data and Machine Learning Application in Management Accounting (Research Question 2)

Next, we discuss the issues and benefits of both technologies in the context of managerial accounting. The most critical issues of big data applications in management accounting are poor *data quality* (41%) and the lack of required *skills* and know-how (23%). In most cases of machine learning applications, the *interpretability problem* (47%) and the complex *training* process (32%) are considered problematic. Overall, most issues are related toward the lack of know-how and experience of management accountants to work with these technologies. Therefore, most issues could be addressed with additional training and education. However, some issues, such as the interpretability problem of machine learning and data quality of big data, remain to be critical. In these cases, compromising between different factors (i.e., accuracy versus interpretability) is necessary.

The most important benefits of big data application are new *insights* (38%), better *decision making* (31%) and an increasing *influence* of accountants in their businesses (31%). In the case of machine learning, mostly an increase in *accuracy* of outcomes is reported (48%), followed by *time*-specific advantages (19%) and a larger *independence* from expert knowledge (19%). Therefore, most benefits point toward an increase of effectivity to solve common management accounting tasks. Some benefits, such as the increase of influential power and independence from expert knowledge, indicate a rather disruptive impact on the management accounting profession. Big data and machine learning can change the role of management accountants toward information specialists. **Table 5** depicts the categories of issues and benefits of both technologies over the nine managerial accounting tasks. The categories of issues and benefits are highly dependent on the corresponding tasks. For example, the *interpretability problem* and the *accuracy* benefit of machine learning is almost entirely relevant in cost estimation. Whereas the *data quality* issue and benefits from new *insights* of big data are mostly effective in the tasks of management reporting and management support.

Table 5: Issues and benefits of big data and machine learning over managerial accounting tasks. The table depicts for each category of issues and benefits the number of articles that describe a specific management accounting task. The most common combination of management accounting task and category of issues and benefits is indicated by underline.

	Big Data								**Machine Learning**						
	Issues					Benefits			Issues			Benefits			
	QUALITY	SKILL	PRIVACY	OVERLOAD	PREPROC	INSIGHT	DEC.MAK	INFLUENCE	INTERPRET	TRAINING	PARAMS	ACCURACY	TIME	INDEP	INSIGHT
Total	**15**	**11**	**7**	**6**	**2**	**7**	**8**	**9**	**9**	**6**	**4**	**14**	**5**	**5**	**4**
Strategic planning	1	1		1			1								
Operat. plan. and budgeting	2	3		1			1	1	1	3	1	2			
Forecasting	1	1		1				1	1	1	1	2		1	
Cost accounting	2	1	1					1				2			2
Cost estimation									5	2	2	6	4	2	1
Management reporting	4	3	2	1		2	2	2	1			1			
Project and investment contl.									1			1	1	1	1
Risk management	1	1	2	1		1	1	2							
Management support	4	1	2	1	2	4	3	2						1	

2.5 Discussion and Conclusion

In this chapter, we investigate the body of literature on the application of big data and machine learning technology in management accounting. The study answers two research questions: 1) *What are use cases of big data and machine learning in management accounting?* 2) *What are the issues and benefits of these technologies for managerial accounting?* The systematic literature review has identified and analyzed 36 articles.

The literature so far has considered big data and machine learning as game-changers for businesses (Carlton Sapp et al., 2019; LaValle et al., 2011). We find several empirical studies, which suggest that big data technology can actually be applied in the managerial accounting tasks of *strategic planning*, *operational planning and budgeting*, *forecasting*, *cost accounting*, *management reporting*, *risk management*, and *management support*. Further we find that machine learning technology can actually be applied in *operational planning and budgeting*, *forecasting*, *cost accounting*, *cost estimation*, *management reporting*, *project and investment controlling*, *risk management*, and *management support*. Since both technologies can be applied on almost any management accounting task, we can confirm the game-changing potential of big data and machine learning technology.

However, we also find that big data and machine learning are mostly used for different tasks. Big data is primarily used for *management reporting* and *management support*, while machine learning is primarily used for *cost estimation*, *operational planning and budgeting*, and *forecasting*. This suggests that the revolutionary potential of big data and machine learning is different between the tasks. Big data yields a larger impact on tasks where descriptive analyses are involved, while machine learning yields a larger potential for tasks with predictive analytics. As both technologies are often primarily used for different tasks, there are not many studies that evaluate or apply simultaneous usage. Literature suggests that the simultaneous usage is beneficial as big data and machine learning are two technologies reinforcing each other (Y. Zhang et al., 2021). We assume that the limited research on the simultaneous usage can be explained by the different knowledge bases needed in order to work with big data and machine learning. Since researchers might either be familiar with big data or machine learning, the simultaneous usage is rather limited.

Our findings suggest two interesting avenues for further research. First, we identified three managerial accounting tasks where either the application of big data or machine learning have not been covered. Therefore, further research could investigate the applicability of big data in *cost estimation* and *project and investment controlling* and machine learning in *strategic planning*. Second, further research is needed to examine the potential benefits and challenges of the simultaneous usage of big data and machine learning in management accounting.

The literature so far has identified several management accounting specific challenges and opportunities of big data and machine learning (Gärtner & Hiebl, 2017; Mardini & Alkurdi, 2021). Our systematic review finds that poor *data quality*, lack in *skills*, *privacy issues*, *information overload*, and the *preprocessing* of raw data are critical issues of big data in management accounting. On the other hand, big data offers new *insights*, better *decision making* and increasing *influence* of accountants. In the case of machine learning, we find that the *interpretability problem*, complex *training* process, and *parameter* setting are considered problematic, while higher *accuracy* of outcomes, *time*-specific advantages, larger *independence* from expert knowledge, and new *insights* are important opportunities.

Our systematic literature review identified several new benefits and issues of big data and machine learning in managerial accounting not considered in literature reviews so far (Gärtner & Hiebl, 2017; Mardini & Alkurdi, 2021). As an example, for big data we identified the increasing *influence* of accountants as a new important benefit and the *privacy issues* a new critical problem. For machine learning, we identified the *independence* from expert knowledge as an additional benefit and the *interpretability problem* as a highly critical issue. Further, our findings suggest a slightly different ranking of issues of big data in management accounting relative to the literature. While in Gärtner and Hiebl (2017) the large data volume and the loss of data sovereignty was considered challenging in managerial accounting, we find that *data quality* and the lack in *skills* as particularly problematic.

In addition, we find that the issues and benefits are highly dependent on the management accounting tasks. As an example, the *data quality* is particularly problematic when dealing with management reporting, while the *interpretability problem* is particularly critical in the case of cost estimation. This differentiation suggests that there are no overall issues and benefits of big data and machine learning, but they rather depend on the managerial accounting task. We can explain this result with the unique characteristics of the nine management accounting tasks. Managerial accounting tasks differ in various aspects such as time, quality, and accuracy requirements. As big data and machine learning have benefits and issues related to these aspects, the significance of opportunities and challenges can be different as well.

Two more research avenues can be suggested. First, most issues of big data and machine learning point toward the general lack of knowledge and experience of accountants, which can be counteracted with education and training or partnering more closely with IT. However, some issues require compromising between different factors (i.e., interpretability versus accuracy). The interpretability problem, which describes the trade-off between interpretability and accuracy of machine learning models, was the most concerning issue of machine learning in management accounting. Future research could analyze solutions to the interpretability problem specifically in the field of managerial accounting. Second, only little field work is applied on this topic. Field research on big data and machine learning from the perspective of management accounting could investigate how such technology is used, how to overcome challenges, and report on actual benefits.

The literature review has some limitations. First, the literature search is limited by the focus on high-ranking journals of the ABDC list in combination with backward and forward search. Of course, we cannot claim completeness since we risk missing relevant studies in lower ranked journals or other journals not included in the ABDC journal list, which are not covered in the backward and forward search. The second limitation refers to the selection of search terms. Some authors might apply big data or machine learning techniques without using our search terms. Still, we assume that a significant part of the relevant studies published in reputable academic journals is included in this review.

2.6 Appendix A

Table 6: Overview of articles in this review

Author (year)	Query	Journal	Research method
Appelbaum et al. (2017)	A1	IJAIS	Theoretical
Bhimani and Willcocks (2014)	A1	ABR	Theoretical
CGMA (2013)	A1	CGMA	Qualitative, Survey
CGMA (2016)	A1	CGMA	Survey
Chua (2014)	A1	ACCA	Theoretical, Qualitative
Economist Intelligence Unit (2013)	A1	EIU	Survey, Qualitative
Green et al. (2018)	A1	IJAIM	Theoretical
Holton (2009)	A1	DSS	Market
LaValle et al. (2011)	A1	MIT SMR	Survey
McAfee et al. (2012)	A1	HBR	Market, Qualitative
Warren et al. (2015)	A1	AH	Theoretical
Desai (1997)	A2	IMA JMM	Archival
Kuzey et al. (2019)	A2	IJAIM	Survey
Liang et al. (1992)	A2	CAR	Market
Bansal, K., Vadhavkar, S. and Gupta (1998)	A3	IJAM	Archival
Chou et al. (2010)	A3	IJPE	Archival
Kostakis et al. (2008)	A3	IJAIM	Simulation study
Partovi and Anandarajan (2002)	A3	CIE	Archival
Shawver (2005)	A3	JETA	Market
Tang (2009)	A3	Neuroc	Field work
Qi et al. (2016)	B1, B3	IM	Market
Tan et al. (2015)	B1, B3	IJPE	Field work
Zhan and Tan (2020)	B1, B3	EJOR	Field work
Caputo and Pelagagge (2008)	B2	IJPE	Archival
Cavalieri et al. (2004)	B2	IJPE	Archival
Loyer et al. (2016)	B2	IJPE	Archival
Coussement et al. (2015)	B2, B3	DSS	Market
Deng and Yeh (2011)	B2	IJPE	Simulation study
Golmohammadi (2011)	B2	IJPE	Archival
Gruss et al. (2018)	B2, B3	DSS	Market
Hua Tan et al. (2006)	B2	IJPE	Field work
Verlinden et al. (2008)	B2	IJPE	Archival
Q. Wang (2007)	B2	IJPE	Simulation study
S. Wu and Akbarov (2011)	B2	EJOR	Market, Archival
Coussement et al. (2017)	B3	DSS	Archival
H. Xia et al. (2020)	B3	IJHM	Market

3 Big Data and Machine Learning in Cost Management: A Case Study

Abstract

This study presents a case study on the applicability of machine learning and big data technology for product cost estimation, focusing on the material costs of passenger cars, using actual company data. The study provides contributions on six research aspects. First, we show what machine learning algorithms are appropriate when dealing with product cost estimation of highly complex products with more than 2,000 parts and hundreds of cost drivers. Second, our case study contributes to the literature by providing a novel approach to increase the predictive accuracy of cost estimates of subsequent product generations. Third, we show that the accuracy is up to 3.5 times higher when using big data compared to an intermediate size of data. Fourth, machine learning can outperform cost estimates from cost experts, or produce at least comparable results, even when dealing with highly complex products. Then, we add to the current literature by evaluating use cases, issues, and benefits of machine learning and big data from the perspective of cost experts. Specifically, the case study shows that machine learning can reliably select the most important cost drivers (fifth aspect) and calculate the average cost of cost drivers over thousands of product configurations (sixth aspect). However, cost experts must be knowledgeable about the product and remain careful when interpreting machine learning outcomes as they can yield misleading outcomes for some exceptional cases. In conclusion, machine learning and big data empirically proved to be able to generate additional values in many aspects for managing cost during the early phase of new product development.

Keywords: Machine learning; Big Data; Product cost estimation; Cost management; Case study

3.1 Introduction

Machine learning and big data technology offer significant potential benefits for managing costs and economic decision-making in manufacturing companies (Chou et al., 2010; Fosso Wamba et al., 2015; Loyer et al., 2016). While the literature offers many theoretical approaches, frameworks, or conceptual applications of big data and machine learning (Bhimani & Willcocks, 2014; Rikhardsson & Yigitbasioglu, 2018; Saggi & Jain, 2018), far less empirical studies report experienced implementation and realized benefits, and these typically do not provide specific insights regarding cost management and economic decision-making (Fosso Wamba et al., 2015; Hua Tan et al., 2006; Tan et al., 2015). The lack of actual implementations and empirical evaluation also encompasses the field of cost estimation. We need to better understand how the potential of these new technologies can be realized in this area. This chapter presents a case study to address the following questions: *How can big data and machine learning technology be applied to complex product cost estimation? What are the actual benefits and insights for such technology for the product development process?*

To answer this research question, we present a case study describing the application of these technologies in the automotive industry using real industrial data. We cooperated with a German car manufacturer to investigate the applicability of machine learning for cost estimation of passenger cars for subsequent product generations. This study describes the implementation for such technology, how to overcome issues, and evaluates potential benefits. Two state-of-the-art machine learning models, namely *Artificial Neural Networks* and *Gradient Boosted Regression Trees*, were applied on high volume data to analyze several aspects of cost estimation and cost management. The research aspects were analyzed based on real industrial data. In addition, the results are discussed and evaluated with cost experts from the case company.

The analysis contributes to the literature in several ways. First, the study shows that machine learning and big data can produce higher predictive accuracy than manual calculations from cost experts or produce at least comparable results even when dealing with highly complex products. We thereby add to the cost estimation literature that mostly considered products with intermediate complexity (Bendul & Apostu, 2017; Caputo & Pelagagge, 2008). Second, the case study introduces a novel approach to increase predictive accuracy for cost estimates for subsequent product generations in the context of multi-generational product development (Cai & Tyagi, 2014; Tyagi et al., 2015). Third, this case study shows that the accuracy is considerably higher when using *big* data sets compared to intermediate-sized data sets. In the context of cost estimation, we consider an amount of data in the few hundreds as intermediate (Caputo & Pelagagge, 2008; Chou et al., 2010; Loyer et al., 2016). Too our knowledge, considerably larger amounts of data in the tens of thousands or hundreds of thousands have not explicitly examined in the context of product cost estimation and are further referred to as big data. Further, we add to the current literature by evaluating benefits and issues of the application of machine learning at a manufacturing company from the perspective of cost experts. Research in the field of product cost estimation already used machine learning models to examine cost engineering insights such as cost driver identification and cost behavior analysis, however, did not empirically validate the results (Chan et al., 2018; Loyer et al., 2016). The evaluation shows that machine learning and big data can reliably select the most important cost drivers and calculate the average cost of product features over thousands of product configurations. Yet, we also highlight the importance of cost estimators and cost managers to be knowledgeable about the product and remain careful when interpreting machine learning outcomes as they yield in some cases misleading outcomes.

The remainder of this chapter is structured as follows: In the next section, we provide a short overview about product cost estimation methods and the application of machine learning in the product cost estimation task. In Section 3.3 we motivate six research aspects in the context of cost estimation and cost management during new product development. In Section 3.4, we conduct a case study approach to analyze the research aspects. Section 3.5 concludes the work.

3.2 Literature Review

3.2.1 Product Cost Estimation

Product cost estimation is a topic of great importance for all manufacturing and producing companies. Especially during the product development process, it is a main interface between engineering and cost management divisions. When it comes to evaluating the cost of a product, several problems arise. An overestimation of cost could mistakenly discard profitable products, while an underestimation of cost causes the risk of producing at loss. Therefore, generating accurate cost estimates at any stage of product development is necessary to remain competitive.

For an overview of product cost estimation methods, the framework of Niazi et al. (2006) is helpful, who divides methods into qualitative and quantitative techniques. *Qualitative techniques* compare new products to previous ones to find similarities. Based on that, past designs and manufacturing data can be used to generate cost estimates for new products. Qualitative techniques can be divided into intuitive and analogical techniques. *Intuitive techniques* are based on expert experience. The knowledge is usually stored in rule sets, decision trees, or case-databases. *Analogical techniques* use historical cost data to train statistical models. Estimations are based on correlations between the features and cost, where causality does not necessarily exist. *Quantitative techniques* make use of the knowledge about the design, materials, and processes of a product. Costs are then calculated using a predefined equation. Quantitative techniques can be divided into parametric and analytical methods. *Parametric techniques* derive an equation for costs, using relevant features as input variables. *Analytical techniques* consider the product as the sum of all necessary units and operations (i.e., break-down approach, activity-based costing).

Machine learning has been applied to product cost estimation in several industries for a great variety of products. In the manufacturing sector, machine learning was used to estimate the cost of machined rotational parts (Li Qian & Ben-Arieh, 2008) and sheet metal parts (Verlinden et al., 2008). In the aerospace industry, machine learning algorithms were applied to predict the manufacturing cost of jet engine components (Loyer et al., 2016) and airframe structural projects (Deng & Yeh, 2011). In the electronics industry, several machine learning techniques were compared for the prediction of thin-film transistor liquid-crystal display manufacturing equipment (Chou et al., 2010; Chou & Tsai, 2012). Caputo and Pelagagge (2008) used neural networks and parametric methods in the heavy carpentry sector to forecast manufacturing cost of large pressure vessels. Y. F. Zhang et al. (1996) applied neural networks to estimate the cost of packaging products. In the automotive industry, Cavalieri et al. (2004) compared parametric methods and neural networks to predict the unitary manufacturing costs of brake disks. Further, discrete event simulation has been applied to the cost analysis of composite cross car beams (Kendall et al., 1998). Farineau et al. (2001) used regression analysis to predict manufacturing cost of gear box casings (i.e., clutch housing, differential carrier, gear box housing). In the study of Stockton et al. (2013), data mining techniques were used to extract cost estimation relationships for automated spray painting and turning processes.

The application of machine learning for product cost estimation proved to be beneficial in many ways. The application of such technology can generate accurate cost estimates (Caputo & Pelagagge, 2008; Loyer et al., 2016), is fast (Cavalieri et al., 2004), can be conducted in the very early phase of new product development (Chou et al., 2010), can lead to more autonomy (Caputo & Pelagagge, 2008; Cavalieri et al., 2004), and might provide important cost-engineering insights (Loyer et al., 2016). Further, machine learning can help to set sales prices as competitive as possible (Verlinden et al., 2008) and generate cost forecasts that are inexpensive in terms of the required type and amount of input data (Q. Wang, 2007). Machine learning-based cost modeling is particularly useful in the early stages of product development when optimizing the design of new products (Cavalieri et al., 2004).

3.2.2 Knowledge Gap

While the literature provides many examples of machine learning applications to predict the cost on part or component level, little is known about the effectiveness in the case of ***highly complex products***. Many modern products are very complex according to their number of parts and the application of commonality strategies resulting in a high variety of products. Secondly, since most cost estimation studies approach cost prediction from the perspective of a single generation, the product cost estimation literature lacks in studies considering ***multi-generational aspects*** of products (Cai & Tyagi, 2014). Third, cost estimation literature is limited when considering the potential of ***big data***. Most studies apply cost estimation methods on an intermediate size of data with few hundred training examples and few cost drivers (Caputo & Pelagagge, 2008; Chou et al., 2010; Loyer et al., 2016). Fourth, literature lacks qualitative research and ***field work***. Empirical cost estimation studies mostly focus on the comparison of statistical models according to their predictive performance based on archival data. However, little is known about the actual utilization in practice. In particular, the literature lacks on empirical analyses on the applicability of complex (*black-box*) machine learning models. Fifth, articles that also considers additional benefits for cost engineers and project management is scarce. One of the few studies that discusses ***additional benefits and insights*** is the work of Loyer et al. (2016), which, however, lacks in the empirical validation of insights.

3.3 Specific Aspects to look at

In the following, we introduce six research aspects to investigate the applicability of machine learning and big data techniques for the estimation of cost of complex and multi-generational products. The first aspect addresses the model selection problem of machine learning applications. Many machine learning algorithms have already been applied in cost engineering and were compared according to their performance. We intend to complement the existing literature by comparing various machine learning methods for the estimation of costs of highly complex products with a large variety of features. The process of finding the best-performing model from a set of models is called model selection.

1st aspect: *Which machine learning models are appropriate for the estimation of manufacturing costs of highly complex products?*

The second aspect deals with the potential of big data application for product cost estimation. Big data has not explicitly been applied in the context of product cost estimation. Currently, the number of training examples for product cost estimation is far from being considered *big* (i.e., 68 observations in Caputo and Pelagagge (2008), 519 in Chou et al. (2010), 254 in Loyer et al. (2016)). There is usually a positive association between the size of training data and the predictive performance of (complex) machine learning algorithms (Goodfellow et al., 2016; Grolinger et al., 2014). Training on more data usually leads to more accurate cost estimates, upon the condition of sufficient data quality. Therefore, we expect that using high-volume data will lead to more accurate cost estimates compared to intermediate amounts of data.

2nd aspect: *Does big data lead to a substantial increase in cost estimation accuracy compared to intermediate amounts of data?*

The third aspect addresses the problem of multi-generational product cost estimation. Using costing data from past generations for the prediction of a subsequent generation seems logical, but also raises concerns about the validity of historical data. There can be several years between two product generations (i.e., approximately seven years in the automotive industry). In the meanwhile, customer value, market, technology, and competitor environment can be substantially different. Consequently, data from past product generations are potentially not representative to predict subsequent product generations.

3rd aspect: *How to improve the predictive performance of multi-generational product estimation?*

The fourth aspect is related to the potential increase in predictive accuracy compared to manual calculations from cost experts. From the literature we know that state-of-the-art machine learning models can be more accurate than conventional statistical methods (Cavalieri et al., 2004; Verlinden et al., 2008). In addition, some research indicates that cost estimates from machine learning methods can outperform cost forecasts from domain experts, based on their experience and traditional cost estimation methods (Bendul & Apostu, 2017; Caputo & Pelagagge, 2008). However, the superiority of machine learning could only be confirmed for products with rather manageable complexity. When dealing with highly complex products, presumably human judgment becomes more important to produce accurate cost estimates. This leads to the question, whether machine learning methods produce more accurate cost estimates than manual calculations from cost engineers when dealing with complex products.

4th aspect: *Does machine learning technology lead to more accurate cost estimates than calculations from cost experts when dealing with highly complex products?*

The remaining aspects cover potential insights of machine learning and big data technology for cost management. Most cost estimation studies compare machine learning algorithms up to their predictive accuracy only. The work of Loyer et al. (2016) is one of the very few studies that also covers potential engineering insights that such technology might yield for cost experts. However, little is known whether such insights are valid and helpful in practice. The following two research aspects examine this question.

The fifth aspect concerns the reliability of cost driver identification based on machine learning. In the context of new product development it is crucial to understand what drives costs. Knowing which features cause the highest impact on cost is especially important for decision making at the early design stage. Cost driver analysis is the "examination, quantification and explanation of the cause-effect relationship of the cost drivers and total overhead costs of an operation" (Schniederjans & Garvin, 1997, p. 72). Understandably, the selection of cost drivers from a set of potential cost drivers can be difficult in the case of high product interdependencies and modular designs. Machine learning practitioners face a similar problem. The knowledge about important features is helpful to improve models, counteract unwanted behavior and build trust in the model (Hooker et al., 2018). Often only few features have a considerable impact on the target variable. The vast majority of features, however, are insignificant and could have also not been collected (Friedman & Meulman, 2003). Loyer et al. (2016) demonstrated the application of machine learning to select and quantify the most important cost drivers. Similar approaches have been applied in the cost driver detection for activity-based costing systems (K. Kim, 2003; Kostakis et al., 2008). However, due to the lack of empirical research it is mostly unclear whether machine learning can reliably unveil the underlying costing structure and produce reasonable results.

5th aspect: *Can machine learning models select the most important cost drivers reliably?*

The sixth aspect covers the relationship between product features and total manufacturing cost. The knowledge about the average cost and their sensitivity of features yields important insights for product planners in the early design stage (Farineau et al., 2001). Due to high product commonality and high interdependencies between parts, however, it is often difficult to estimate these relationships over multiple product configurations. Product features often occur only in combination with other features due to bundling and packaging approaches (i.e., equipment lines of cars). Therefore, the cost of one feature is often confounded with other features. Machine learning can be used to extract the relationships between product features and manufacturing costs by using the knowledge representation of machine learning models (Chan et al., 2018; Loyer et al., 2016). However, due to the lack of empirical research, little is known whether these approaches produce indeed meaningful and reliable information for cost planners.

6th aspect: *Can machine learning models reliably estimate the relationship between cost drivers and manufacturing cost?*

Accordingly, we explore the following research questions: *How can big data and machine learning technology be applied for complex product cost estimation? What are the actual benefits and insights for such technology for the product development process?*

3.4 A Case Study

We conducted a case study to investigate the research questions and the six research aspects. In the following, we shortly describe the case study approach and its adequacy to answer the research questions. Then, the case company and its products are introduced, and relevant background information of the industry is provided. Following this, the case study evidence and the data collection process are described. Then, the six research aspects are analyzed based on quantitative and qualitative data.

3.4.1 Case Study Research

In the following, we explain the reasoning behind choosing a case study approach. Case study research is appropriate in situations of asking *how* and *why* research questions, having no requirement of control over the behavioral aspects and a focus on present events (Yin, 2018). Since we are interested in the question of how machine learning and big data technology can be applied in practice, a case study is an appropriate method in contrast to methods relying only on quantitative data. In a case study, several sources of evidence are combined: archival documents, interview data, direct observations, and physical artifacts (Yin, 2018). The case study approach is often used to examine how techniques and processes from theory are applied in practice (Scapens, 1990). Otley and Berry (1994) furthermore highlight that case studies can provide a more comprehensive perspective by which theories can be generated and modified. To examine our research questions, highly sensitive data (quantitative and qualitative) is needed. In addition, the expertise of cost experts to evaluate and discuss the findings is required. Therefore, a case study is conducted that involved not just obtaining the data from the company but also working with company employees on site to obtain trust and a comprehensive understanding about processes and data sources. We select a research site on the basis of the following four requirements. Since we are interested in the applicability of machine learning and big data in the cost estimation task of highly complex products, sufficient complexity of the products is needed in terms of the number of parts and variety of product configurations. Second, the case company needs to handle multi-generational products. Third, the case company needs to have large volume of costing and technical data over at least two product generations. Fourth, the access to domain experts needs to be granted to discuss and evaluate findings.

3.4.2 Case Company

3.4.2.1 Research Site

The case study is conducted at a German car manufacturer operating in the premium segment. The case company is further referred to as *AutomotiveCompany*. The product line-up encompasses several vehicles ranging from small cars to luxury cars. The author was three years on site and participated in several cost management and new product development projects. The author supported project teams at the integration of costing data and data analysis. Therefore, extensive access to the company’s costing and technical databases was granted. The author mainly participated in the *parts controlling* department. Part controllers are responsible for the management of direct material costs of parts and participate in the *simultaneous engineering* (SE) teams. The estimation of direct material costs at *AutomotiveCompany* is mostly based on analytical techniques incorporating historical costing data adjusted by expert knowledge. So far, the controlling department of *AutomotiveCompany* has no prior experience of machine learning- and big data-based product cost estimation methods. When dealing with passenger cars, direct material cost represent the largest share of total cost (approximately 60%). Therefore, at

AutomotiveCompany most cost reduction effort and attention during new product development is placed on this cost type. **Figure 7** illustrates an exemplary breakdown of sales price and cost structure of a passenger car.

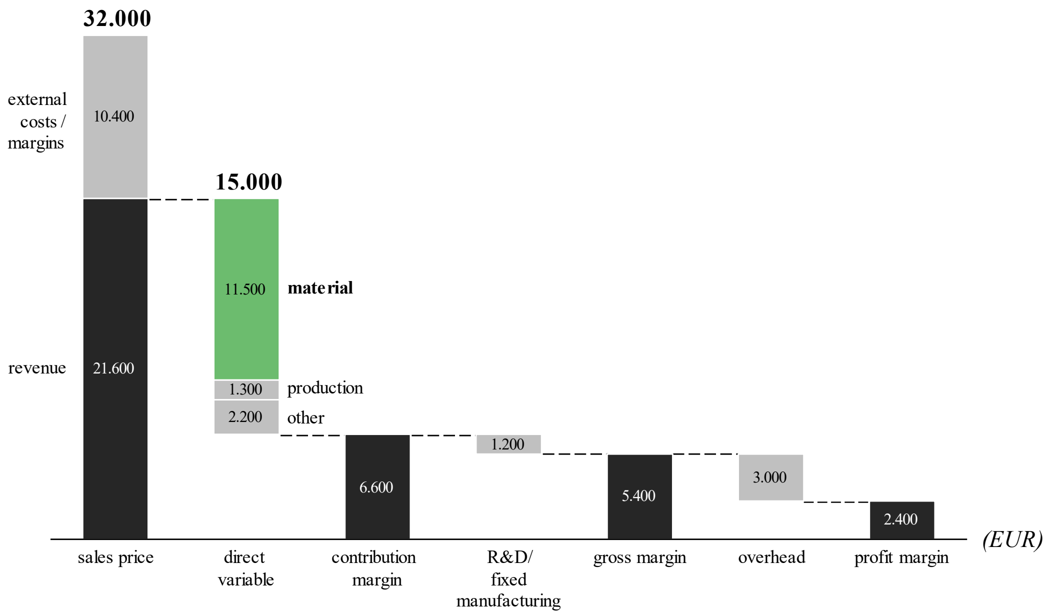
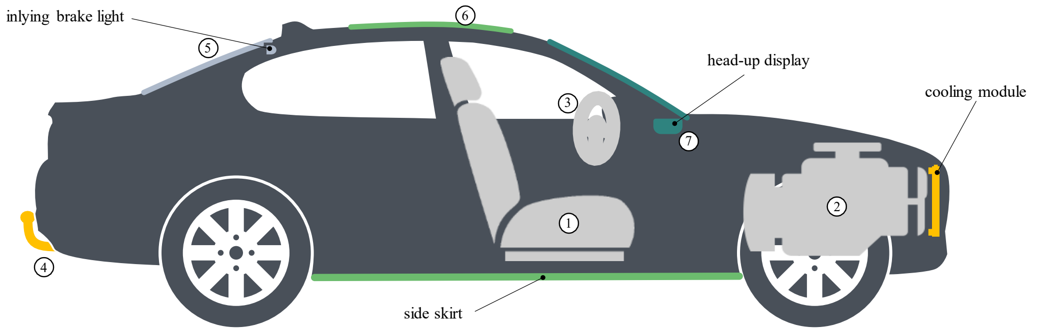


Figure 7: Breakdown of a passenger car's sales price into cost types. Disguised case company data.

In this case study, the unit of analysis are entire car models over thousands of configurations. A car model can be for example the Volkswagen Golf or the Volkswagen Passat. In total we analyze 120,378 unique configurations of four car models over two product generations. Modern passenger cars are highly complex products. At *AutomotiveCompany* a bill of materials consists of approximately 2,500 parts and components. On one side, components typically have a large technical variety, resulting in a wide range of component costs (i.e., car seats, steering wheel, combustion engine). On the other side, there are many interdependencies between product features, where one product feature requires or prohibit other features. **Figure 8** showcases some examples of the complexity of components and the complexity of the assembly of passenger cars.



High component complexity

1. Cost of **seats** depend on type (i.e.sports or comfort), covering, heaters, lumbar support, ventilation, head restraints, seat massage
2. Cost of combustion **engines** depend on performance, displacement, fuel type, emission standard, no. cylinders, hybridization, hot/cold climate requirement
3. Cost of **steering wheel** depend on covering, no. spokes, multifunction, shift paddles, design-elements (e.g. chrome or lacquered)

High assemply complexity

4. When installing a **coupling device** a more powerful cooling module is required.
5. When installing **privacy glazing** a brighter inlying brake light is required.
6. When installing a **panoramic roof** a stiffer side skirts is required.
7. When installing a **head-up display** a special windshield is required.

Figure 8: Complexity of modern passenger cars (graphic by the author)

3.4.3 Case Study Data

The case study data encompasses a mix of quantitative and qualitative data. The quantitative data is being used to analyze and test the research aspects. The qualitative data is used to evaluate the findings regarding the research aspects from the perspective of cost experts. The qualitative evaluation is based on the aspects: feasibility, usability, utility, and intention to use.

3.4.3.1 Quantitative Data Collection

The ***quantitative*** archival data includes complex bills of materials, production data, and descriptions of technical features. In the following, we explain the combination of data from *AutomotiveCompany*'s costing, production, and technical IT systems. First, we explain the data sources and the necessary pre-processing steps. Then, we explain the construction of the multi-generational data set and provide an overview of the direct material cost.

3.4.3.1.1 Data Sources and Pre-Processing

We employ three data sources: costing, production, and technical data. The ***costing data*** is based on the post-calculation system, which tracks the direct material costs of a car model in the form of complex bills of materials. A complex bill of materials covers all parts and components of a car model (i.e., 15 steering wheels, 10 combustion engines). The ***production data*** includes the configurations of product characteristics for all cars produced in the form of order codes. A product characteristic describes the specific version of equipment or technical part. An order code is the combination of all product characteristics that describe a certain configuration of a car model (i.e., with or without head-up display, the version of the car seats). The ***technical data*** encompasses short textual descriptions for each product characteristic. To create numerical representations for the product characteristics, we manually extract the individual ***product features*** from the descriptions. For example, a specific combustion engine can be described by the engine performance, number of cylinders, and the turbo specification. Ordinal dichotomous product characteristics are thereby encoded by 1 = with product feature, 0 = without product feature. Nominal categorical product characteristics are encoded by *one-hot encoding*. One-hot encoding creates a binary vector for each category. In total 461 individual product features are extracted from all product characteristics. 10 of which are cardinal (i.e., engine performance, rear brake performance), the majority however are binary features such as the advanced display system, advanced suspension system, and driver assistance system.

3.4.3.1.2 Data Combination

The product features and direct material cost of a produced car are combined based on the specific order code. The total direct material cost is calculated as the sum of all parts required to satisfy the order code. The product features can be mapped to each order code according to the corresponding product characteristics. Further, the cost is grouped into four categories or ***assembly groups***: *body*, *electrics*, *chassis*, and *engine*. The body group consists for example of the bodyshell, windows, heating, air conditioning, seats, bumper, interieur equipment, and trim. The electrics group contains the infotainment system, on-board power supply, electronic control units, and battery. The chassis encompasses axles, steering, brakes, and fuel system. The engine group includes the combustion engine, transmission, and clutch. Some parts cannot be attributed to any of the four assembly groups. Relative to the total cost the remaining costs range from 0.44% to 0.91%. For the sake of simplicity, we discard these parts when predicting cost at the assembly level. Duplicated order codes are kept in the database to capture the relative frequency of individual product features and feature combinations.

To analyze the multi-generational aspect of product cost estimation, we collect data for two periods containing two product generations: predecessor generation (three-month period) and successor generation (four-month period). The car models in both data sets are mutually exclusive. The data set includes four car models (*car a*, *car b*, *car c*, and *car d*) ranging from a small car to a luxury car. **Table 7** provides an overview of the data set. As an example, the product feature head-up display (HUD) is mapped to 1 (included) or 0 (not included), the product feature *steering wheel covering* includes among others the categories leather and polyurethane (PUR) coating, the product feature *combustion engine* includes among others the number of cylinders, engine performance, and a biturbo charging specification. The database covers four car models (*a*, *b*, *c*, and *d*) over two generations (I and II) with the corresponding product features and corresponding direct material costs (assembly-level and total cost).

Table 7: Overview of the combined data set. The data encompasses four car models for two generations with thousands of configurations according to different product features and the corresponding direct material costs (fictive data).

Car model	Product features								Direct material cost [€]		
	HUD	⋯	leather	PUR	⋯	cyl.	perform.	biturbo	electr.	⋯	sum
a_I	0	⋯	1	0	⋯	4	145	0	1,500	⋯	9,200
a_I	0	⋯	0	1	⋯	4	160	1	1,600	⋯	9,500
…											
d_I	1	⋯	1	0	⋯	6	170	1	6,200	⋯	31,000
…											
b_{II}	1	⋯	1	0	⋯	6	220	1	4,000	⋯	16,000
…											
d_{II}	1	⋯	1	0	⋯	8	270	1	7,600	⋯	38,000

3.4.3.1.3 Splitting the Data Set

To train and test machine learning models, the data is split into a training, validation, and testing set. The training set is used for fitting the models; the validation set is used to tune the hyperparameters; the testing set is used to evaluate the performance. We distinguish between predecessor generation data (generation I) and successor generation data (generation II). The predecessor generation data is employed for training, validation, and inner-generational testing. The data is split into 80% training set, 10% validation set and 10% testing set. To ensure that the duplicated order codes are mutually exclusive in each sample, we employ the following sampling process: For each car model, we randomly select one order code from the predecessor data set. Next, all observations that have the same order code are added to the training set and removed from the predecessor data set. The selection process is repeated until the training set exceeds 80% of observations. The same process is applied for the validation set (repeated until < 10%). The remaining observations are used for inner-generational testing. This sampling process leads to the following distribution of unique order codes: The training set captures in average 53% of the unique order codes over the four car models. The validation set contains in average 22% and the testing set 25% of order codes. The successor generation data is exclusively used for testing (inter-generational testing). **Figure 9** provides an overview of the data set configuration.

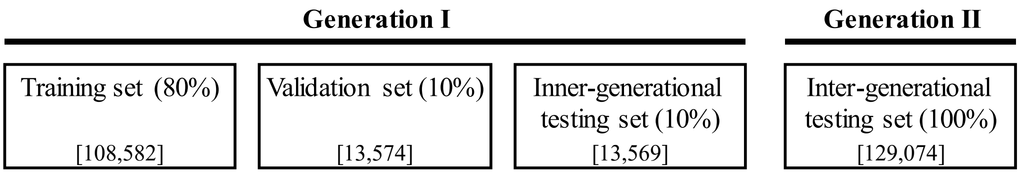


Figure 9: Overview of the multi-generational data set. Number of observations are depicted in brackets.

We employed a re-fit strategy to make use of the complete predecessor generation data. Therewith, the model for inner-generational testing is trained on the training and validation set. The inter-generational testing is based on the complete predecessor data set (training, validation, and inner-generational testing set). Thereby, the models are fitted to the additional observations without updating the hyperparameters. **Table 8** provides an overview of the costing data for the predecessor and successor generations. The number of individual order codes provide insight into the complexity of modern passenger cars. During the period of only three months, 32,880 unique order codes of *car* c_I have been produced. The direct material cost of *car* c_I ranges from €14,657 to €42,088. The standard deviation amounts to €4,406.

Table 8: Overview of the predecessor and successor product generations. The summary table incorporates the number of observations (Obs.), unique order codes (OC) and descriptive statistics of the direct material cost (DMC) in EUR.

Car	Obs.	OC	DMC^{min}	DMC^{max}	$\overline{DMC}$	DMC^{SD}
Predecessor product generation (I)						
car a_I	41,846	16,515	8,179	15,379	9,869	1,077
car b_I	23,044	11,875	12,090	21,649	16,136	1,991
car c_I	62,706	32,880	14,657	42,088	20,914	4,406
car d_I	8,129	2,061	26,549	50,808	31,991	5,544
Total	135,725	63,331				
Successor product generation (II)						
car a_{II}	24,034	9,312	8,340	12,981	10,099	1,196
car b_{II}	66,479	31,275	12,584	31,018	17,292	3,337
car c_{II}	32,177	15,403	17,768	32,567	23,591	3,221
car d_{II}	6,384	1,057	30,757	46,429	35,256	3,765
Total	129,074	57,047				

3.4.3.2 Qualitative Data Collection

To evaluate the results regarding *feasibility*, *usability*, *utility*, and *intention to use*, semi-structured interviews were conducted to collect ***qualitative*** data. We spoke with 14 cost experts, five from the complete vehicle controlling department (C1 – C5) and nine from the parts controlling department (P1 – P9) of *AutomotiveCompany*. All interviewed cost experts have work experience in the controlling department of at least three years. The approach to improve the predictive performance for multi-generational cost estimation (3rd aspect), the comparison of cost estimation accuracy (4th aspect) and reliability of cost driver selection (5th aspect) were discussed with complete vehicle controllers. The reliability of machine learning models to estimate the relationship between product features and manufacturing cost (6th aspect) was discussed with part controllers. The analysis of the 1st aspect (algorithm selection) and 2nd aspect (impact of big data) is based on quantitative data only.

The evaluation of research aspects is based on the measures feasibility, usability, utility, and intention to use (Agarwal & Prasad, 1998; Hua Tan et al., 2006; Platts, 1993; Tan et al., 2015). We formulate five criteria to assess the applicability of machine learning and big data in the context of cost management (**Table 9**).

Table 9: Evaluation criteria for case study results

Evaluation criteria		
C_1	*feasibility*	How valid would you rate the method/results?
C_2	*utility (support)*	Are the results helpful for your work? In what situation?
C_3	*utility (insight)*	Does/Do the method/results yield any new or unexpected insights?
C_4	*usability*	How easy could the method be integrated into your work?
C_5	*intention to use*	Do you intend to use the method in practice?

3.4.4 Results

3.4.4.1 1st aspect: Model Selection for the Estimation of Manufacturing Costs of Highly Complex Products

The first research aspect deals with the algorithm selection problem for product cost estimation of complex products. In this section, we compare the predictive accuracy of several machine learning models: *artificial neural network* (ANN), *case-based reasoning* (CBR), *decision tree regression* (DTR), *elastic net regression* (ELR), *gradient boosted regression* (GBR), *lasso regression* (LAR), and *linear support vector regression* (LSVR). To implement the models, each model has been tuned individually. The hyperparameter-setting procedure for each model is described in the Appendix B (**Table 19**). The models are trained on the training set (80%) and tuned according to the validation set (10%). The selected models were implemented in Python with the *scikit-learn* package (Pedregosa et al., 2011).

The predictive performance is measured with four metrics: *normalized mean absolute error* (NMAE), *mean absolute percentage error* (MAPE), *normalized root mean square error* (NRMSE), and *explained variance score* (EVS). The NMAE, MAPE and NRMSE are error metrics where lower values are preferable. The NMAE is calculated as the mean absolute error divided by the mean of the actual values. The MAPE is calculated by the mean relative absolute deviation between the actual value (y_i) and forecast value ($\hat{y}_i$) to the actual value. The MAPE measure is the most commonly used metric for evaluating forecasts in companies (Gneiting, 2011). However, one major drawback is that the metric puts more emphasis on negative deviations ($y_i < \hat{y}_i$). When used to select between alternative forecasting models, this metric systematically selects those which underestimate the actual values (Tofallis, 2015). Therefore, the NMAE and MAPE are both used to evaluate the relative error. The NRMSE puts more weight on larger errors and therefore additionally penalizes models with larger errors. The EVS measures the capability to explain variations in the data (higher values are preferable). The EVS almost equals the coefficient of determination (R^2) except that the EVS can handle the skewness of residuals (Guia et al., 2020). The four metrics are calculated as follows:

$$NMAE = \frac{\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|}{\frac{1}{n}\sum_{i=1}^{n} y_i} \tag{1}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{2}$$

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\frac{1}{n}\sum_{i=1}^{n} y_i} \tag{3}$$

$$EVS = 1 - \frac{Var(y - \hat{y})}{Var(y)} \tag{4}$$

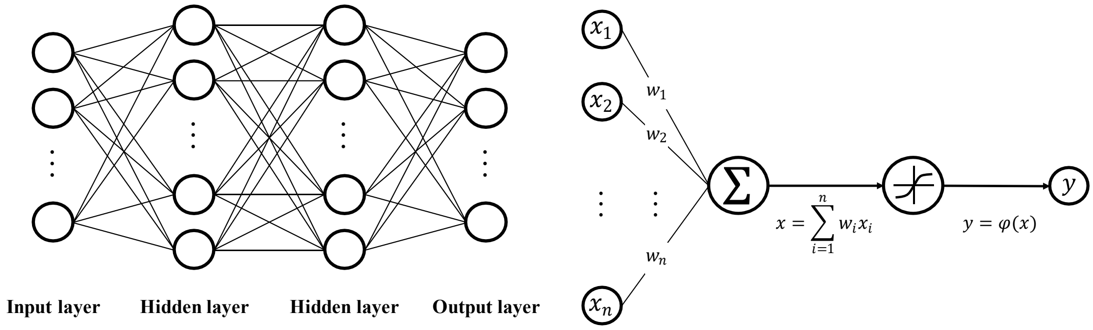
Table 10 compares the machine learning models according to the four metrics over the four cars (*car* a_I, *car* b_I, *car* c_I, *car* d_I). The highest predictive performances are attained by the GBR and ANN models. In the following, we further describe both algorithms and use the two models to analyze the remaining research aspects (2, 3, 4, 5, and 6). In doing so, we are also able to investigate the practical applicability of highly complex (black-box) machine learning models from the perspective of cost experts. Notably, the much simpler LAR model yields comparable predictive accuracy that suggests that not always high complex machine learning models are required when conducting product cost estimation within the same product generation.

Table 10: Comparison of machine learning models on the total cost prediction for the validation set. The table depicts the average performance over the four car models.

Model	NMAE [%]	MAPE [%]	NRMSE [%]	EVS [%]
GBR	**0.88**	**0.89**	**1.32**	**99.21**
ANN	**1.16**	**1.15**	**1.58**	**98.93**
LAR	1.26	1.27	1.67	98.76
LSVR	1.42	1.40	1.94	98.44
DTR	2.20	2.21	3.13	96.05
ELR	2.36	2.32	3.18	95.61
CBR	2.61	2.59	3.72	94.55

3.4.4.1.1 Artificial Neural Networks

Neural networks are graph models that mimic the function of neurons in the human brain (Shtub & Zimerman, 1993). The network consists of computational units, so-called neurons, that are organized in different layers. Each neuron obtains input signals from preceding units and passes a transformed signal to subsequent units if activated. The neurons on the input layer receive the input data, while the output layer combines all signals to produce the output data. In between, hidden layers allow for complex interactions and nonlinear behavior (**Figure 10a)**. The input value of a neuron is usually the simple weighted summation of all input signals, which is then modified by an activation function. In product cost estimation usually the back-propagation algorithm is used to determine the weights of the signals between the neurons (Niazi et al., 2006). The complexity of an artificial neural network is mainly determined by the number of hidden layers and the number of neurons. These parameters need to be adjusted carefully when designing the topology of neural networks. If the complexity of the network is too high, an overly complex function is generated to approximate the training data. Later, however, the network possibly fails to perform on the testing data set. This effect is called overlearning or overfitting. Determining the best topology for the neural network is usually done by trial-and-error where different networks are tested and the best one is selected.



(a) Schematic drawing of deep learning with two hidden layers

(b) Representation of the Rosenblatt's perceptron (Rosenblatt, 1958)

Figure 10: Neural network representations

In the following, we describe the learning procedure for multi-layer feed-forward neural networks based on the work of Rumelhart et al. (1986). Back-propagation neural networks are based on a supervised learning procedure, meaning that the network is constructed by presenting data with known input and output values. The aim is to find a set of weights (w) that reduces the differences between the generated output vectors (y) and the target output vector (t). **Figure 10b** represents a basic neural network design. The procedure starts by initializing the weights with small random numbers. The weight between neuron i and neuron j is denoted by w_{ij}. The network is successively fed by the training data and the output is calculated by a so-called forward pass. In doing so, the neurons in each layer determine their states by the input signals they receive from units in lower layers. The i-th input of neuron j is denoted as x_{ij}, the total input to neuron j (x_j) is calculated as $\sum_{i=1}^{n} x_{ij} w_{ij}$. The output of a neuron j (y_j) is then calculated by applying an activation function φ ($y_j = \varphi(x_j)$). The prediction error is then calculated by comparing the output layer and the target vectors (t_j). Derivatives of the error are propagated backwards from the output layer through the network. Each weight is changed by an amount proportional to the accumulated error, $\partial E / \partial w$. An important parameter for the back-propagation step is the learning rate (η) that adjusts the rate at which the weights are updated.

Algorithm 1: *Back-propagation algorithm*

1. Initialize weights with small random numbers.
2. Repeat:
   - (a) Selection of observation d from training set T.
   - (b) Compute *output* for d.
   - (d) Successive back-propagation of error on neurons

$$\delta_j \begin{cases} o_j(1-o_j)\sum_{k \in Downstream(j)} \delta_k w_{jk} & j \notin output \\ o_j(1-o_j)(t_j-o_j) & j \in output \end{cases}$$

   - (e) Update weights $\Delta w_{ij} = \eta \delta_j x_{ij}$

The number of hyperparameters of ANNs is vast. Typical parameters that are used in most implementations are the number and size of the hidden layers, the learning rate and the dropout rate (Diaz et al., 2017). Other hyperparameters that are often adjusted are activation functions, learning rates, number of training epochs, dropout for regularization, loss functions, and optimizers (Neary, 2018).

3.4.4.1.2 Gradient Boosted Regression

Boosting is an important approach in machine learning which was developed by Freund and Schapire (1997). The approach involves the combination of many simple models (so-called weak learners or base learners) to produce a single ensemble with high performance. This procedure is applicable for classification problems (categorical dependent variable) and regression problems (continuous dependent variable). In the boosting approach, the ensemble $F(x)$ is calculated as the weighted sum of base learners $f_m(x)$, $F(x)=\sum_{m=1}^{M}\beta_m f_m(x)$, where β_m denotes the expansion coefficient at each iteration. Thereby, an additive model is created by successively fitting base learners to the residuals of the current ensemble. This approach has been refined by Friedman (2001) and Friedman (2002) by the introduction of the *TreeBoost* algorithm that uses regression trees as base learners. Thereby, the TreeBoost algorithm combines the advantages of the boosting approach and regression trees such as conceptual simplicity and computational efficiency (Shin, 2015).

In the following, the TreeBoost algorithm for regression trees is described. Since the selection of the optimal base learner in each iteration is computationally infeasible, a steepest descent step is applied. In step 2(a), the so-called pseudo-residuals are computed for each observation i. In 2(b) a base learner is fitted to the pseudo-residuals (r_{im}) resulting in the terminal regions R_{jm}. The terminal regions are the decision nodes of the regression tree. In 2(c) the step size (γ_{jm}) is computed. For the commonly used squared-error loss, the solution is the regression tree that best predicts the current pseudo-residuals. Therefore, the step size γ_{jm} is the mean of the residuals in each corresponding terminal region. In 2(d) the ensemble is updated. The final model $F(x)$ consists of the aggregation of M base learners. Overfitting can be reduced by introducing a shrinkage parameter ν, $0 < \nu \leq 1$, and adjusting the number of base learners M. The algorithm is described in the following:

Algorithm 2: Gradient TreeBoost

---

1. Initialize $f_0(x) = \arg\min_\gamma \sum_{i=1}^{N} L(y_i, \gamma)$.
2. For $m = 1$ to M:
   (a) For $i = 1, 2, ..., N$ compute $r_{im} = -\left[\frac{\partial L(y_i,\ f(x_i))}{\partial f(x_i)}\right]_{f=f_{m-1}}$
   (b) Fit a regression tree to the targets r_{im} giving terminal regions R_{jm}, $j = 1, 2, ..., J_m$.
   (c) For $j = 1, 2, ..., J_m$ compute $\gamma_{jm} = \arg\min_\gamma \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i)+\gamma)$.
   (d) Update $f_m(x) = f_{m-1}(x) + \nu \sum_{j=1}^{J_m} \gamma_{jm} \mathbf{1}\ x \in R_{jm}$.
3. Output $F(x) = f_M(x)$.

The hyperparameter tuning of GBR models is mainly based on three parameters: the number of trees (M), the shrinkage parameter (ν), and the depth of the regression trees (Loyer et al., 2016). The depth of regression trees controls the number of binary evaluations in each tree. The shrinkage (or learning rate) parameter controls the weight of each decision tree in the aggregated model (Landry et al., 2016).

3.4.4.2 2nd Aspect: Impact of Big Data on Cost Estimation Accuracy

The second research aspect raises the question whether big data leads to a substantial increase in cost estimation accuracy compared to intermediate amounts of data. Big data usually refers to complex and very large data sets. It is often characterized by the 3Vs: volume, variety, and velocity (Laney, 2001; McAfee et al., 2012). To evaluate the impact of big data (in this case mostly defined by volume), we vary the amount of data in two ways. First, we reduce the number of observations in the training data set. Therefore, 300 unique order codes are randomly selected for each car model. Second, we take all observations but only select the top 20 most important features according to the GBR model.[2] Finally, we do both, which results in a sample of 1,200 configurations and 20 features. This yields three subsets of training data. **Figure 11** summarizes the results of the analysis for each performance metric. The usage of reduced observations and/or reduced number of features substantially decreases in predictive accuracy. The utilization of the reduced number of observations increases the NMAE score from 0.88% to 1.64%. The intermediate number of features leads to a NMAE score of 2.49%. In combination, the NMAE rises to 3.08%. Consequently, we can confirm the additional value of big data technology for product cost estimation.

[2] The selection of features is based on the impurity-based feature importance, see Section 3.4.4.5.

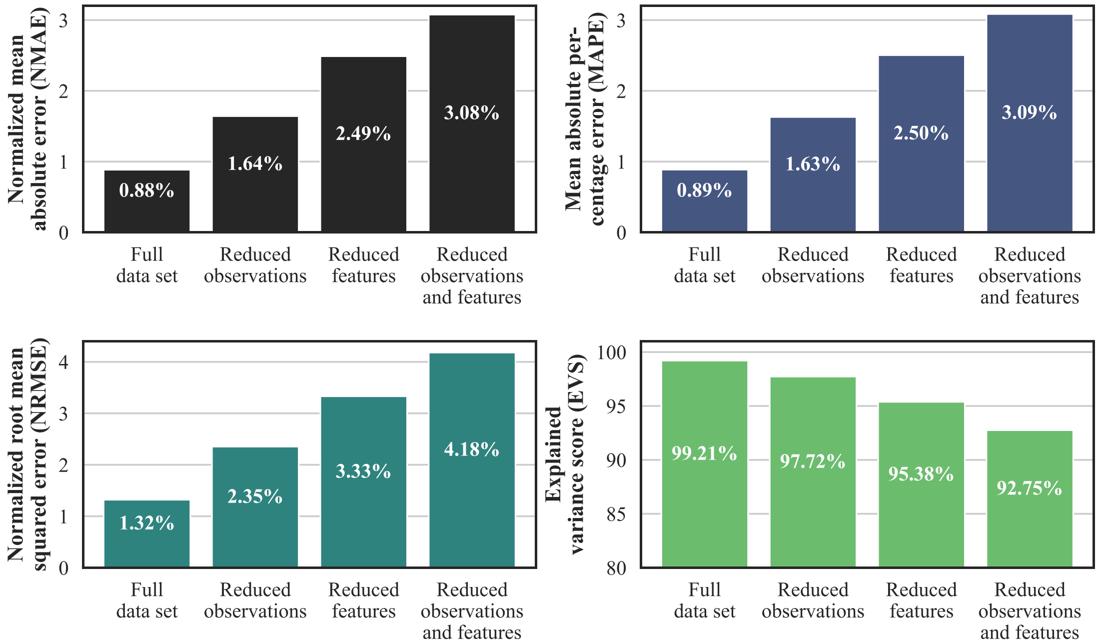


Figure 11: Impact of big data on cost estimation accuracy. The figure depicts a comparison of the predictive performance between the full data set, reduced number of observations (300 order codes for each car model), reduced number of features (top 20 most important features), and reduced number of observations and features.

3.4.4.3 3rd Aspect: Improving Predictive Performance for Multi-Generational Product Cost Estimation

The third research aspect deals with the improvement of predictive performance of multi-generational product cost estimation. At the case company, it is considered that direct material cost decrease by approximately four percent per year due to learning curves under the condition of equality in product design, quality, and other product properties. Therefore, some parts that remain unchanged will be cheaper in the subsequent product generation. To ensure the competitiveness of new product generations, new features, superior design, and better quality need to be introduced that again cause an increase in product costs. This increase in cost depends on several factors such as market environment and the targeted customer value. The change of cost between two product generations can therefore be approximated by the sum of the cost decrease due to learning curves and cost increase due to new and advanced product characteristics. It can be difficult to estimate this change of cost with cost estimation models relying on product features only. New features might be deployed in the successor product generation, which are unknown to the machine learning model. On the other hand, it is difficult to measure intangible properties such as quality, impression, and the design of parts that, however, also have considerable effects on product cost.[3] In the following, we introduce a *generation shift factor* to approximate the change of cost between two product generations.

To account for the cost changes between product generations, we adopt the target costing approach. Target costs are calculated as the difference between sales price and profit margin (profit percentage x sales price). We utilize this functional relationship between the sales price and cost to estimate the change in direct material cost between two generations. First, a linear regression model is fitted on the relationship between the average net earnings ($\overline{NE_I}$) and average direct material cost ($\overline{DMC_I}$) over all products from the preceding product generation. We use the regression model and the estimated average net earnings for a product i from the successor generation ($\overline{NE}^{e}_{II,i}$), to calculate the corresponding estimated average material cost ($\overline{DMC}^{e}_{II,i}$). The estimated generation shift factor for product i ($\hat{\delta}_i$) is defined as the ratio between the estimated average direct material cost of generation I and II. The final cost prediction of a predecessor generation product ($DMC_{II,i}$) is obtained by multiplying the generation shift factor ($\hat{\delta}_i$) with the prediction from the machine learning model for this product ($DMC^{m}_{II,i}$).

[3] In case of passenger cars, the interior trims can have more superior slush elements instead of plastic panels. A new generation of combustion engines might be equal according to its product features, but may differentiate by its running smoothness.

$$\overline{DMC}_I = \beta_0 + \beta_1 \overline{NE}_I + \epsilon \tag{5}$$

$$\overline{DMC}^e_{II,i} = \beta_0 + \beta_1 \overline{NE}^e_{II,i} \tag{6}$$

$$\overline{DMC}^e_{I,i} = \beta_0 + \beta_1 \overline{NE}_{I,i}$$

$$\hat{\delta}_i = \frac{\overline{DMC}^e_{II,i}}{\overline{DMC}^e_{I,i}} \tag{7}$$

$$DMC_{II,i} = \hat{\delta}_i \times DMC^m_{II,i} \tag{8}$$

where

I: predecessor project
II: successor project

In the following, the proposed approach is applied on company data from *AutomotiveCompany*. During the last two decades, the car manufacturer evolved from a volume manufacturer into a premium brand. As a result, net earnings and direct material cost increased during this transition. Accordingly, the mean direct material cost of the four car models increased from generation I to II. As an example, the mean direct material cost of *car c* increased from €20,914 in generation I to €23,591 in generation II (see **Table 8**). In any case, the distribution of direct material cost is shifted to the right. The shape of distribution, however, remained largely the same between both generations. Therefore, we can rule out distributive effects where more high-end configurations cause the increase in average direct material cost. The regression analysis is based on the four predecessor projects (*car* a_I, *car* b_I, *car* c_I, *car* d_I). Net earnings constitute the weighted mean for each car model. The results of the regression model are provided in Appendix B (**Table 20**). The OLS model exhibits a significant linear relationship (p-value < 0.01) between net earnings and direct material costs. Economically the model shows that an increase in net earnings by €100 leads to an increase in direct material cost by €48.81. The linear model leads to a R^2 of 0.993. The regression line and actual costs of car models for both generations are depicted in **Figure 12**. The scatter plot suggests a strong linear relationship between net earnings and direct material cost.

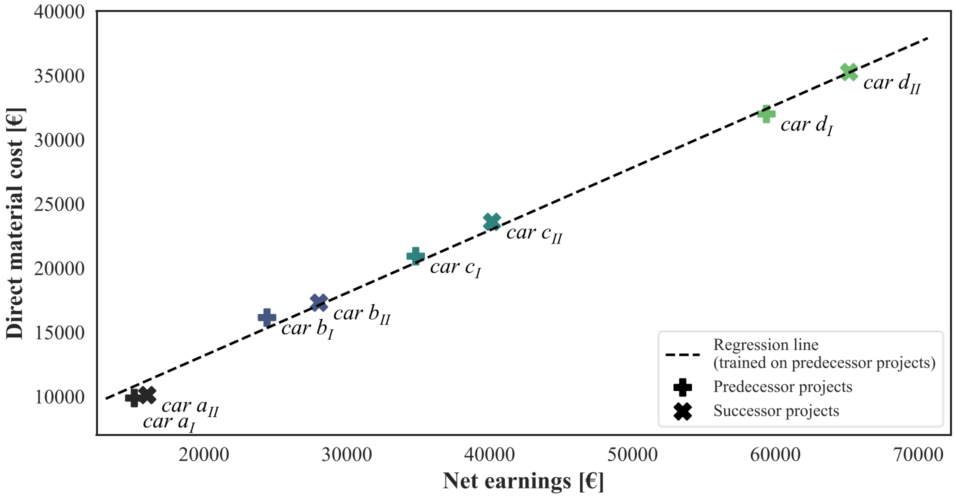


Figure 12: Linear relationship between the actual direct material cost and net earnings. The regression line is fitted on the predecessor data (R^2 0.993). The scatter plot and the regression line suggest a strong positive linear relationship between net earnings and direct material cost.

To validate the approach, we compare the predicted generation shift factor based on the regression formula ($\hat{\delta}$) with the actual change of direct material cost (δ). Therefore, we calculate the mean absolute error (MAE) between the actual change of cost and the predicted change of cost, $\frac{1}{4}\sum_{i \in \{a, b, c, d\}} |\hat{\delta}_i - \delta_i|$, which amounts to 0.0192 (SD 0.0184). The mean change of direct material cost over the four car models amounts to 1.0812 (SD 0.0450). Accordingly, a large proportion of the actual change between two generations can be estimated.

Finally, the approach is evaluated by cost experts from the complete controlling department of the case company using the evaluation criteria form **Table 9**. Overall, the method to adjust for the cost changes between two generations was perceived feasible, useful, and usable. Hence, the intention to use was rated high, provided that both generations are comparable according to certain market parameters such as sales volume and commodity prices. **Table 11** provides an overview of the evaluation with cost experts.

Table 11: Evaluation of the 3rd research aspect: Improving the predictive performance for multi-generational product cost estimation

Criteria		**Evaluation feedback from cost experts**
C_1:	*feasibility*	The approach was perceived as a valid method to roughly predict the costs of subsequent product generations. The ratio between direct material cost and net earnings in the regression analysis widely coincides with the rule of thumb at *AutomotiveCompany* of 50% (C1, C2, C3, C4).
C_2:	*utility (support)*	Overall, one was confident that the approach can help to adjust for cost changes between two generations in the very early phase when only little information about the subsequent generation is available (C1, C2, C5).
C_3:	*utility (insight)*	Most controllers have been surprised about the strength of the linear association between target net earnings and actual direct material cost. Normally, one would expect that the ratio between direct material cost and net earnings is lower in top-of-the-range segments (C1, C2).
C_4:	*usability*	One was optimistic about the potential integration into a machine learning-based cost estimation process, since all required information is available for almost any product in the early phase (C1, C3, C4).
C_5:	*intention to use*	Many controllers point out that the total cost of a car model depends on several factors such as project volume, exhaust emission standards, vertical range of manufacture, commodity prices, and manufacturing location. If subsequent generations are less comparable according to these factors, further adjustments will be required (C1, C2, C5).

3.4.4.4 4th Aspect: Comparison of Cost Estimation Accuracy: Machine Learning vs. Expert Judgment

The fourth research aspect raises the question of whether machine learning technology leads to more accurate cost estimations than calculations from cost experts. First, we evaluate the cost estimation accuracy of the GBR and ANN model on the inner- and inter-generational data set. Second, we compare the results with the predictive accuracy of cost experts. The testing data is based on the data set split depicted in **Figure 9**. During inner-generational forecasting, the cost of 13,569 new configurations within the same generation (I) are predicted. During inter-generational forecasting, the costs of 129,074 configurations of the subsequent generation (II) are estimated. For each car model an individual machine learning model was trained. Therefore, we obtain four distinct GBR models and four ANN models. To analyze the predictive accuracy of both testing sets, the accuracy metrics NMAE, MAPE, NRMSE, and EVS are applied. First, the accuracy metrics are calculated for each car model (*a*, *b*, *c*, and *d*) individually. As an example, the inter-generational cost prediction accuracy of *car a* is based on the 24,034 observations (see **Table 8**). Then, the accuracy scores of the four car models are averaged to obtain an overall mean accuracy score (*mean over cars*). As an example, the mean

NMAE of the total cost prediction is computed as $\frac{1}{4}\sum_{i \in \{a, b, c, d\}} \text{NMAE}_i$. By grouping the accuracy scores over the car models, we cancel out the odd number of observations (i.e., 66,479 *car b*, 6,384 *car d*). Otherwise, *car b* would have a 10-times greater impact on the overall accuracy score than *car d*. Moreover, the relationship between the predictive performance of direct material cost and the predicted generation adjustment factor ($\hat{\delta}$) can be analyzed. Finally, we can investigate the variation of accuracy scores over different car models.

The results of the comparison of the two machine learning models for the total cost estimation task are exhibited in **Table 12**. On the inner-generational prediction task the GBR model has an average NMAE of 0.88% over the four car models. The ANN performs slightly worse and results in an average NMAE of 1.75%. In the case of the inter-generational prediction task, the GBR model results in an average NMAE of 4.86% and the ANN results in a score of 5.97% respectively. Again, the GBR outperforms the ANN forecasting model. In the case of inter-generational cost prediction, the predictive accuracy appears to strongly depend on the error between the predicted and the actual adjustment factors ($\hat{\delta} - \delta$), as the highest (lowest) δ-*deviation* corresponds to the most (least) accurate cost prediction.

Table 12: Total cost prediction accuracy for inner- and inter-generational data. The mean accuracy measures are calculated as the average accuracy scores over the four car models (*car a*, *car b*, *car c*, and *car d*).

		Total cost prediction accuracy				
Model	**Metric**	**Mean over cars (SD)**	***car a***	***car b***	***car c***	***car d***
Inner-generational cost prediction						
GBR	**NMAE [%]**	**0.88 (0.26)**	**0.86**	**0.68**	**0.66**	**1.32**
	MAPE [%]	0.88 (0.26)	0.85	0.71	0.65	1.31
	NRMSE [%]	1.41 (0.43)	1.22	1.33	0.96	2.11
	EVS [%]	99.13 (0.44)	99.13	98.95	99.82	98.62
ANN	**NMAE [%]**	**1.75 (0.92)**	**1.07**	**1.77**	**0.93**	**3.24**
	MAPE [%]	1.75 (0.92)	1.06	1.78	0.90	3.24
	NRMSE [%]	2.18 (1.04)	1.45	2.13	1.27	3.89
	EVS [%]	98.92 (0.5)	98.81	98.76	99.74	98.39
Inter-generational cost prediction						
GBR	**NMAE [%]**	**4.86 (1.63)**	**3.67**	**6.97**	**2.93**	**5.87**
	MAPE [%]	4.57 (1.39)	3.62	6.33	2.85	5.46
	NRMSE [%]	7.02 (3.27)	4.27	11.88	3.80	8.11
	EVS [%]	79.39 (13.61)	91.63	62.58	93.87	69.46
ANN	**NMAE [%]**	**5.97 (1.36)**	**3.86**	**6.23**	**7.65**	**6.16**
	MAPE [%]	5.85 (1.34)	3.89	5.85	7.68	6.00
	NRMSE [%]	7.35 (1.65)	4.69	9.14	8.17	7.41
	EVS [%]	85.56 (8.12)	91.92	77.76	95.28	77.30
δ-*deviation*$_i = \hat{\delta}_i - \delta_i$			0.018	0.044	0.000	-0.015
Note:	$i \in \{car\ a, car\ b, car\ c, car\ d\}$					

Next, the cost estimation accuracy on assembly level is examined. Since *car a* has no assembly cost information, the analysis is based on the car models *b*, *c*, and *d*. First, the accuracy metrics are calculated for each car model and assembly group (body, electrics, chassis, and engine) individually. Second, for each assembly group the mean accuracy score is calculated by averaging the accuracy scores over the three car models. Then, the accuracy scores of the four assembly groups are averaged to obtain an overall accuracy score (*mean over assemblies*). For example, the mean NMAE on assembly level is calculated as $\frac{1}{4}\sum_{i\,\in\,\{body,\ electrics,\ chassis,\ engine\}}\frac{1}{3}\sum_{i\,\in\,\{b,\ c,\ d\}}\mathrm{NMAE}_i^j$. NMAE_i^j is the prediction error over all configurations of a car model (i) for an assembly group (j). **Table 13** exhibits the results of the model comparison on the assembly level. On the inner-generational prediction task the GBR model yields a mean NMAE of 1.14% over the four assemblies. The ANN performs slightly worse and results in 2.41% average NMAE. In the case of inter-generational prediction, the GBR model results in a NMAE of 13.98% and the ANN in a NMAE of 12.63%. The low accuracy scores on assembly level are mainly caused by the large deviations of the actual change of direct material cost between generation I and II over the assembly groups. For example, the actual average cost change between generation I and II of the *electrics* assembly amounts to 1.499, while the actual average cost change of the *chassis* group amounts to 0.9817. Since no sales prices on assembly level are available, the total cost adjustment factor of the corresponding car model must be equally used for each assembly group. Hence, the accuracy of the assembly cost estimation strongly depends on the error between the predicted and the actual adjustment factors for a given assembly ($\widehat{\delta_j} - \delta_j$). The highest (lowest) δ-*deviation*$_j$ corresponds again to the most (least) accurate cost prediction. As an example, the deviation between the *predicted* average cost change of the electrics group and the *actual* average cost change of the electrics group (δ-*deviation*$_{electrics}$) amounts to -0.389. The actual change of direct material cost of the electrics group is therefore, in average, underestimated. For the remaining assembly groups (body, chassis, and engine) the change of cost is, in average, overestimated.

Table 13: Assembly cost prediction accuracy for inner- and inter-generational data. The accuracy scores for each assembly group (body, electrics, chassis, and engine) are calculated as the mean accuracy score over the three cars *b*, *c*, and *d*. The mean assembly cost estimation performance is calculated as the average accuracy scores over the four assemblies.

		Assembly cost prediction accuracy				
Model	**Metric**	**Mean over assemblies (SD)**	**Body**	**Electrics**	**Chassis**	**Engine**
Inner-generational cost prediction						
GBR	**NMAE [%]**	**1.14 (0.56)**	**1.64**	**1.16**	**1.53**	**0.22**
	MAPE [%]	1.15 (0.57)	1.63	1.18	1.57	0.20
	NRMSE [%]	2.03 (0.34)	2.44	1.68	2.29	1.72
	EVS [%]	98.51 (0.49)	97.71	98.52	98.97	98.85
ANN	**NMAE [%]**	**2.41 (0.76)**	**2.25**	**2.61**	**3.46**	**1.33**
	MAPE [%]	2.42 (0.84)	2.21	2.58	3.60	1.27
	NRMSE [%]	3.35 (0.76)	3.17	3.27	4.53	2.42
	EVS [%]	97.54 (0.76)	96.67	97.11	97.68	98.69
Inter-generational cost prediction						
GBR	**NMAE [%]**	**13.98 (6.60)**	**6.99**	**24.82**	**12.25**	**11.85**
	MAPE [%]	13.70 (6.06)	6.50	23.30	12.23	12.75
	NRMSE [%]	16.96 (7.10)	10.52	28.97	14.45	13.88
	EVS [%]	62.87 (9.70)	72.31	46.63	65.69	66.85
ANN	**NMAE [%]**	**12.63 (8.43)**	**6.50**	**27.18**	**8.44**	**8.40**
	MAPE [%]	12.36 (7.79)	5.96	25.67	8.36	9.43
	NRMSE [%]	15.49 (8.84)	10.06	30.80	10.67	10.45
	EVS [%]	61.87 (8.47)	70.96	53.87	69.70	52.96
$\delta\text{-}deviation_j = \frac{1}{3}\sum_{i \in cars} \hat{\delta}_{ij} - \delta_{ij}$			0.046	-0.389	0.129	0.085
Note:	$i \in \{car\ b, car\ c, car\ d\}$, $j \in \{body, electrics, chassis, engine\}$					

The application of the generation shift factor ($\hat{\delta}$) proves to be an effective approach to improve the performance of inter-generational predictions. The usage of the total cost estimates from the GBR model without the application of the generation shift factor results in an average NMAE of 10.12% (+5.26 ppt). In the case of the ANN, the average NMAE without adjustment amounts to 11.97% (+6.00 ppt). Without adjustment, the assembly-level cost prediction of the GBR model leads to an average NMAE of 15.93% (+1.95 ppt) and in the case of the ANN to 16.64% (+4.01 ppt) respectively.

Further, we analyze the distribution of the prediction error between the estimated cost and actual cost. **Figure 13** depicts the actual and predicted total and assembly costs for the inter-generational prediction task of the GBR model. The total cost scatter plot shows an increasing prediction error (under-estimation) for *car b* and *car d* in higher-end configurations. The scatter plot on assembly level reveals lower performance of the *electrics* group especially at high-cost configurations, which presumably corresponds to the electric-specific innovations not included in the feature space.

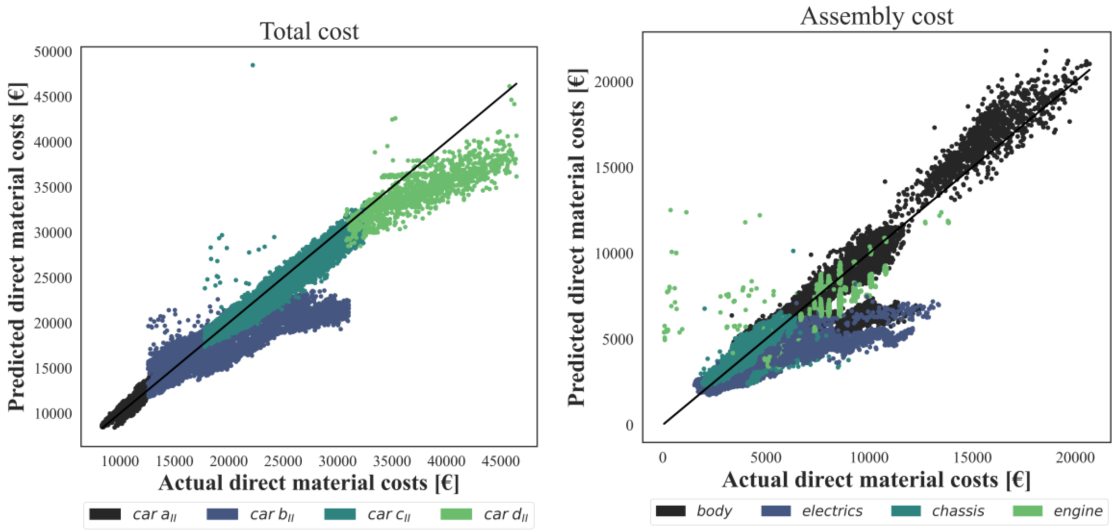


Figure 13: Scatter plot of actual versus predicted total and assembly costs for inter-generational prediction of the GBR model. The assembly costs are the mean costs per assembly over three cars.

Next, we compare the predictive performance with the manual estimations from cost experts. For this purpose, we rely on the corresponding *reference configurations* of the analyzed car models. *AutomotiveCompany* uses the reference configurations as proxy order codes for a car model to manage costs during product development. The manual cost estimates are based on the calculations from the controlling department at the early stage of new product development. The mean total cost estimation performance is calculated as the average NMAE over the four reference configurations of the four cars. The mean assembly cost estimation performance is calculated as the average NMAE over the four assemblies (body, electrics, chassis, and engine). The predictive performance for each assembly group is, again, calculated as the average NMAE score over the four reference configurations. The manual cost estimation results in an average NMAE of 7.29% (SD 4.05%) for complete cost prediction. Thus, both machine learning models outperform the manual estimations from the controlling department. In the case of the assembly-level cost estimation, the NMAE amounts to 10.47% (SD 7.46%). Accordingly, the manual calculations are more accurate than the machine learning models at more granular cost levels.

Finally, the machine learning-based cost estimation approach is discussed with cost experts from *AutomotiveCompany*. The results of the evaluation are summarized in **Table 14**.

Table 14: Evaluation of the 4th research aspect: Comparison of cost estimation accuracy: Machine learning vs. expert judgment

Criteria	**Evaluation feedback from cost experts**
C_1: *feasibility*	The ANN and GBR models are sufficiently accurate for providing total cost estimations for the early phase of new product development (C1, C2, C3, C4). The assembly cost prediction, however, has been rated as not sufficiently precise (C1, C2, C3, C4).
C_2: *utility (support)*	A serious concern was caused by the low transparency and interpretability of both machine learning models. A comprehensible cost estimation method was considered crucial in most situations of the cross-functional product-planning process (C1, C3, C4). Nonetheless, machine learning was considered a helpful tool for cost management, especially when the product design changes rapidly, few information is available, and only rough cost estimates are needed (C1, C3, C4, C5).
C_3: *utility (insight)*	The degree of cost increase of electrics was unexpectedly high. Yet, the increase could be explained by new innovations in infotainment, overall electrification of components and the shift of parts from other assemblies toward the *electrics* assembly group (C2, C4, C5).
C_4: *usability*	One was optimistic that the integration of machine learning-based cost predictions can help finding cost efficient product designs during the product-planning process. This design-to-cost approach yields great potential to improve the design phase (C1, C4, C5).
C_5: *intention to use*	Finally, one was positive about the intention to use and the integration into established product-planning tools (C1, C2, C3, C5).

3.4.4.5 5th Aspect: Reliability of Machine Learning-Based Cost Driver Selection

The fifth aspect raises the question whether machine learning models can identify and rank the most important cost drivers of highly complex products reliably. In the context of manufactured goods, cost drivers are described by any factor or activity that impact the total product cost. In this study, the 461 product features are used as cost drivers. We measure the impact of product features on direct material cost by the ***feature importance*** scores of the GBR model. The feature importance is a score that describes how useful a feature is for the machine learning model to predict the target variable. Therefore, the feature importance score quantifies the impact of any given feature on the predictive accuracy. One approach to derive the feature importance from statistical models is the utilization of linear regression coefficients, which is also commonly used for cost driver analysis. Since we apply more complex machine learning models to predict cost, we derive the feature importance directly from the trained models. Specifically, the ***impurity-based feature importance*** of the GBR model is used to select the most important cost drivers. For a single binary split tree T with $J - 1$ internal nodes, Breiman et al. (1984) proposed the squared relevance, $I_j^2(T)$, to measure the importance of features. The measure is based on the number of times a feature is used to split a tree, weighted by the squared improvement resulting from those splits (i_t^2). For regression trees, the squared improvement measures the effectiveness of a split to decrease the squared prediction error. Therefore, a high squared relevance corresponds to a strong cause-and-effect relationship between a product feature and the direct material cost. The feature importance metric can be applied to GBR models by calculating the mean over the M trees in the ensemble. Due to stabilizing effects, the aggregated measure turns out to be more robust than measures based on a single tree (Friedman & Meulman, 2003). Usually, the resulting squared relevance measures are proportionally scaled to a maximum value of 100.

$$I_j^2(T) = \sum_{t=1}^{J-1} i_t^2 \mathbf{1}(v_t = j) \tag{9}$$

$$I_j^2 = \frac{1}{M} \sum_{m=1}^{M} I_j^2(T_m) \tag{10}$$

where

i_t^2: reduction in squared error due to the split in node t
v_t: index of splitting variable in node t

The impurity-based feature importance calculation can be misleading for high cardinality features with many unique values. Therefore, the results of cardinal features (i.e., engine performance) are validated by comparing the results with an alternative measure, the ***permutation feature importance***. The permutation feature importance measurement was introduced by Breiman (2001) for random forest models. The permutation feature importance is a model agnostic measure that means that it can be applied to any machine learning method. First, the original model error is calculated ($error_{orig}$). Second, a feature-wise permutation step is conducted, where one feature is randomly shuffled, and afterwards, again, the model error is calculated ($error_{perm}$). The permutation feature importance (PFI) is then calculated as the ratio between the permutation error to the original error. If a feature is not important to the model, the PFI score is close to 1. If the PFI is considerably greater than 1, the feature is important.

$$PFI = \frac{error_{perm}}{error_{orig}} \tag{11}$$

The investigation for the 5th and 6th research aspects is based on a detailed analysis of *car c*. The car model is selected, since it yields the largest training data set and, as a top-of-the-range car, covers a high variety of potential product features. The impurity-based feature importance is derived from the GBR model, which was trained on the training and validation set of *car c* (n = 56,436). The permutation feature importance is also based on the GBR model, which was trained on the training and validation set and is calculated for the inner-generational testing set (n = 6,270). **Table 15** exhibits the 10 most relevant features and the importance measure according to the impurity-based importance. The most important feature is *engine power* with a relevance of 38%. In total, the top ten features have a combined importance of 81%. The 50 most important features cover 98% of the feature importance. When comparing the impurity-based and permutation-based importance scores, the results of the cardinal features are largely comparable: *engine performance* (equal rank), *rear brake performance* (rank 3 instead of 2), *engine size* (rank 2 instead of 4). The impurity-based and permutation methods identify the same cardinal features but in a slightly different order.

Table 15: Feature importance of the GBR model (impurity-based and permutation-based). The model was trained on the inner-generational training set (90% of the predecessor generation data). The feature scale is distinguished between binary features (B) and cardinal features (C).

Rank	Feature	Scale	Impurity-based importance	Permutation-based importance (rank)	
1	engine performance	C	0.380	7.75	(1)
2	rear brake performance	C	0.109	3.08	(3)
3	leather interior	B	0.092	1.64	(18)
4	engine size	C	0.062	4.16	(2)
5	camera system	B	0.036	1.39	(25)
6	driver assistance system	B	0.028	2.16	(5)
7	standard version (no high performance)	B	0.028	1.53	(21)
8	advanced display system	B	0.026	1.68	(15)
9	advanced suspension system	B	0.023	2.85	(4)
10	standard differential	B	0.020	1.42	(23)

To validate the selected cost drivers and the importance scores for direct material cost, the results are discussed with complete product controllers from *AutomotiveCompany*. The results are summarized in **Table 16**.

Table 16: Evaluation of the 5th research aspect: Reliability of machine learning-based cost driver selection

Criteria		**Evaluation feedback from cost experts**
C_1:	*feasibility*	Overall, the selection and ranking of the cost drivers could be confirmed by the cost experts. The GBR method reliably captured the features that cause a high proportion of direct material cost. Thereby, the relevance is not only based on individual component costs, but also on the influence on other parts (C4, C5, P6).
C_2:	*utility (support)*	The cost driver identification was considered to be helpful in several ways. First, the analysis helps to investigate the rather complex black-box model. Second, as the selection of features widely aligns with the ranking expected by the experts, it provides trust in the model and the cost predictions (C2, C3). Third, the analysis shows that despite the complexity of passenger cars, not hundreds of features are necessary to estimate material costs. This reduces the effort for data collection and improves the easiness to use in practice (C2, C3).
C_3:	*utility (insight)*	One controller explained that the high importance of engine performance might be caused by interdependencies with other parts. The combustion engine clearly accommodates a large proportion of direct material cost; however, the high relevance score can rather be explained by the effects on other parts (i.e., gearbox, motor cooling, overall powertrain) (C4, C5, P1). Still, it was perceived unplausible that engine performance explains 38% of total cost (C1, C4, C5).
C_4:	*usability*	The analysis helps to better understand the cost structure. Further, the cost experts can check what features need to be considered when managing cost and what features might be discarded (C1, C4, C5). The optimal number of features of a cost estimation model would cover about 70 input variables, which correspond to the number of features in the list of key product features at *AutomotiveCompany* (C1, C2). Others consider 20 cost drivers as acceptable to provide cost estimates in the early phase of new product development (C3, C5).
C_5:	*intention to use*	In general, the intention to use of the cost driver analysis was rated high (C1, C3, C4).

3.4.4.6 6th Aspect: Reliability of Machine Learning Models to Estimate the Relationship Between Features and Manufacturing Cost

The sixth aspect addresses the reliability of machine learning models to estimate the average costs of product features and the cost behavior of cardinal features over different feature values. The average cost of product features cannot be calculated from the data straightforwardly. In most cases, the total cost of a feature relies on many different parts and supplementary components. For example, the cost for a head-up display also depends on the bezel and the wiring harness. However, many of these parts serve multiple purposes and enable several features. As an example, the bezel of the head-up display might be covered in leather and therefore also corresponds to the feature *leather interior*. Therefore, the sum of cost over all parts that are connected to the head-up display would lead to an over-estimation of cost as they are attributed to the head-up display only proportionally. Alternatively, the cost of a certain product feature could be calculated by the difference between two configurations that differ only in this specific feature. For a binary feature the cost therefore can be calculated as the difference between the actual cost of a configuration with the product feature and without the feature. However, due to the high number of combinations between product features, these cases are limited and for some features they do not exist.

To solve this problem, we use the knowledge representation of machine learning models. In doing so, we utilize the machine learning models to simulate changing feature values and estimate the effect on total cost. To do this, the Morris approach is adopted (Morris, 1991). This method calculates the ***mean elementary effect*** (μ_a) of feature X_a over multiple differences (so-called e*lementary effects*) as a measure of global sensitivity. Thereby, the approach considers the average cost over various feature configurations and quantifies the range of costs. For the calculation of elementary effects (*EE*) we distinguish between binary (B) and cardinal (C) features. In the case of binary features, the difference between the total cost predictions with and without a feature is calculated. Therefore, for each observation, the original feature value is replaced by 1 or 0 respectively, all else being equal. To calculate the elementary effects of binary features, the more accurate GBR method is deployed. In the case of cardinal features, the differences between the total cost predictions of the initial feature value and the incremented value are calculated. For this purpose, the ANN is deployed. The GBR model could potentially classify initial and marginal changed feature values into the same terminal regions and therefore generally predict no cost changes. Finally, we calculate the mean elementary effect (μ_a) of feature X_a over n observations. In addition, the standard deviation of the elementary effects (σ_a) is computed, which provides information on the degree of interaction with other features (Pianosi et al., 2016).

$$\hat{Y} = f^m(X_1, X_2, \ldots, X_k) \tag{12}$$

$$EE_a(X) = \begin{cases} f^G(X_a{=}1, \cdot) - f^G(X_a{=}0, \cdot) & \textit{if } X_a \in \mathrm{B} \\ f^A(X_a{=}X_a{+}1, \cdot) - f^A(X) & \textit{if } X_a \in \mathrm{C} \end{cases} \tag{13}$$

$$\mu_a = \frac{1}{n}\sum_{i=1}^{n} EE_a(X^{(i)}) \tag{14}$$

$$\sigma_a = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} \left(EE_a\left(X^{(i)}\right) - \mu_a\right)^2} \tag{15}$$

where

f^m: machine learning model m := {A(NN), G(BR)}
B: binary features
C: cardinal features

To calculate the mean elementary effect, we use the GBR and ANN models, which were trained on the complete predecessor product generation data set of *car c* (n = 62,706). Thereby, we use actual data samples instead of applying a sampling procedure. The combination of product features underlies technical restrictions, which prohibit certain configurations (i.e., a high-performance engine together with a basic gear box). The application of a naive sampling method without accounting for these restrictions, would result in non-realistic samples, causing biases to the results. Instead, the complete predecessor product generation data set of *car c* is used. To treat each order code equally, we drop all duplicated observations, resulting in n = 32,880 individual configurations. **Table 17** depicts the results of the sensitivity analysis for the 10 most important features according to impurity-based measure of the GBR model (see **Table 15**). The table exhibits the mean and standard deviation of direct material cost for the selected features. Overall, the large standard deviations indicate high interaction effects among the product features.

Table 17: Mean elementary effects (*EE*). The mean elementary effects are based on the Morris method (values in EUR). The calculation is based on all unique observations of *car c*.

Rank	Feature (X_a)	Scale	Mean *EE* (μ_a)	Std. *EE* (σ_a)
1	engine performance	C	10.16	13.60
2	rear brake performance	C	63.39	4.66
3	leather interior	B	338.07	318.03
4	engine size	C	122.05	5.13
5	camera system	B	187.69	130.64
6	driver assistance system	B	535.23	158.70
7	standard version (no high performance)	B	-75.94	63.47
8	advanced display system	B	472.43	115.63
9	advanced suspension system	B	812.55	307.21
10	standard differential	B	51.82	252.90

Next, we analyze the cost behavior of cardinal features. Specifically, the change of direct material cost over changing feature values is investigated (i.e., different levels of *engine performance*). Therefore, ***partial dependence functions*** (Friedman, 2001) are deployed. The partial dependence function of feature X_a is defined as the expected value over the joint, marginal distribution over the remaining features X_R (Equation 16). The marginal distribution can be approximated by Equation 17, where $X_R^{(i)}$ are the observations of X_R in the training data. Partial dependence functions are usually centered to have a mean value of zero.

$$f_a(X_a) = E_{X_R}[f^m(X_a, X_R)] \tag{16}$$

$$f_a(X_a) \approx \frac{1}{n}\sum_{i=1}^{n} f^m(X_a, X_R^{(i)}) \tag{17}$$

To estimate the partial dependence functions, we use the GBR model of *car c*, which was trained on the complete predecessor product generation data set (n = 62,706). The calculation of partial dependence is based on all unique observations in the training data set (n = 32,880). The partial dependence plots for *engine performance*, *rear brake performance*, and *engine size* are depicted in **Figure 14**. Both engine-specific features produce step functions. The cost of *rear-brake-performance* is invariant for low performance levels and first starts to increase with a medium performance level.

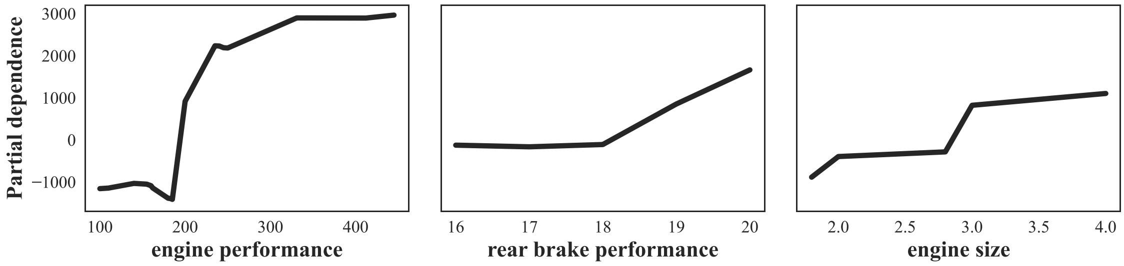


Figure 14: Partial dependence plots for cardinal features. The partial dependence measures the marginal impact of the cardinal features on the direct material cost. The three partial dependence plots suggest a non-linear relationship between the cardinal features and direct material cost.

Next, the elementary effects and partial dependence plots are evaluated by cost experts from the part controlling department (**Table 18**).

Table 18: Evaluation of the 6th research aspect: Reliability of machine learning models to estimate the relationship between features and manufacturing cost

Criteria		**Evaluation feedback from cost experts**
C_1:	*feasibility*	The mean elementary effects of the binary features could be approved in almost any case. The only implausible elementary effect was the *standard differential* where the estimated costs were far less than the actual component cost. This indicates that the analysis yields better results for additionally added components (i.e., camera system) or advanced versions of components (i.e., advanced suspension system) and is less valid for baseline versions of features. The elementary effects of the continuous features are also rated as plausible; however, the actual cost behavior is reflected far better by the partial dependence plots. The partial dependence plots were appealing to the cost experts. In the case of engine performance, the gradient boosted model exactly identified the well-known cost hike when passing a performance level 185. Here, the costlier V6 engine instead of the 4-cylinder in-line engine need to be deployed (P1). The cost behavior of the rear brake performance and engine size also correspond to the expected cost patterns (P3, P6).
C_2:	*utility (support)*	The results are considered helpful in many ways. First, rough cost indications about features can help optimizing product designs during the product-planning stage (P1, P5). Second, the analysis provides the mean cost over all configurations of a car model instead of relying only on one reference model. This enables a more holistic view on cost management (P4, C1). Third, the elementary effect analysis was considered helpful to increase trust into the overall machine learning approach (P8).
C_3:	*utility (insight)*	The standard deviation, however, was unexpectedly high (P2, P4). One controller explained that the variance of cost cannot only be attributed to different versions of components but also to varying pre-conditions. Hence, depending on the configuration, more or less components are added alongside with a given feature (P5). For example, advanced track rods are required for plug-in hybrid electric vehicles and vehicles with air suspension. Therefore, the cost of an air suspension system is lower for a plug-in hybrid vehicle as the costs for the track rod can be shared.
C_4:	*usability*	The usability was rated high for providing rough cost estimates during the early design stage and the management of cost during new product development (P2, P4, P8). In the case of detailed and granular cost calculations at later development stages, the approach was not considered usable due to the low accuracy and the lack of transparency and interpretability (P4, P8, P9).
C_5:	*intention to use*	The intention to use was rated high for the application at the design stage (P2, P3, P4, P8). The approach was well received by cost experts.

Additionally, ***two-way global sensitivity analysis*** was conducted to investigate interactions between product features. Campolongo and Braddock (1999) proposed an extension to the one-factor-at-a-time method, which enables the calculation of second-order effects. A second-order elementary effect provides a measure of the impact on the output due to the interactions between two input factors. A description of calculation of the *mean two-factor interaction effect* can be found in Appendix B. To evaluate the two-way interaction effects, a sample of feature pairs is selected and discussed with cost experts. The sample of feature pairs is based on the 10 most important features (**Table 15**) and the GBR model. The ANN model did not produce substantial two-factor interaction effects. Accordingly, only binary features are considered in the analysis. Furthermore, the feature *standard version* is discarded since no distinct association to parts and components is given. This selection results in six primary features, namely *leather interior*, *camera system*, *driver assistance system*, *advanced display system*, *advanced suspension system*, and *standard differential*. To select feature pairs with high interaction, Friedman and Popescu's *H*-statistic is calculated for all combinations between primary features (X_a) and secondary features (X_b). The statistic introduced by Friedman and Popescu (2008) measures how much variation of the prediction depends on the interaction between two features.[4] Next, for each primary feature the 10 secondary features with the highest *H*-statistic are selected and the mean two-factor interaction effect is calculated for each selected feature pair. The sample of 60 feature pairs is evaluated by the corresponding cost experts from the part controlling department. In most cases, the identified interaction effects could not be confirmed. However, for some combinations the GBR model produces reasonable results. For example, the advanced suspension system and the high-performance version of rear brake systems both need additional and more advanced parts that can be shared when both features are deployed (P5). However, only in the case of 7 feature pairs (12%) the interaction effects have been rated as plausible by the cost experts. Most interaction effects were rated as spurious and could instead be explained by price bundling strategies and the availability of car lines (P1, P4, P8). Accordingly, the GBR model might assign higher costs for certain feature pairs as they are usually deployed in fully equipped bundles or lines.

3.5 Discussion and Conclusion

This chapter investigates the applicability of machine learning and big data technology for the estimation of cost of complex products from a multi-generational perspective. Thereby, actual benefits and insights of such technology for cost managers were analyzed in a case study at a German car manufacturer. Two state-of-the-art machine learning algorithms, an ANN and a GBR model, were applied to estimate and analyze direct material cost of passenger cars. This in-depth case study is a response to the lack of field work of machine learning and big data applications for cost management during new product development. The results suggest that machine learning and big data can be

[4] The calculation of the *H*-statistic is based on the Python package *sklearn-gbmi 1.0.3*. Haygood, R. (2020, 4. Juni). *sklearn-gbmi 1.0.3*. Python Package Index (PyPI). https://pypi.org/project/sklearn-gbmi/, (accessed September 18, 2020).

useful to support and augment the management of cost and the decision-making process during new product development. Specifically, machine learning and big data technology can lead to more accurate cost estimations, is able to identify the most important cost drivers, and can help to determine the average costs of product features reliably. Cost experts from the case company found the results encouraging and believed that machine learning could improve the efficiency of their cost management process. In sum, machine learning and big data are both valid technologies to improve the estimation of direct material cost of passenger cars.

3.5.1 Research Implications

This chapter has several implications for research and practice. Specifically, the following six research aspects were analyzed to investigate the overall research questions: *How can big data and machine learning technology be applied for complex product cost estimation? What are the actual benefits and insights for such technology for the product development process?*

1st research aspect (*Which machine learning models are appropriate for the estimation of manufacturing costs of highly complex products?*): Several machine learning models have been applied in the context of product cost estimation on part level and products with intermediate complexity, whereby more complex models usually produced more accurate cost predictions (Deng & Yeh, 2011; Loyer et al., 2016; Stockton et al., 2013; Y. F. Zhang et al., 1996). Similarly, we find that machine learning models with higher complexity tend to perform better on the estimation of direct material cost of passenger cars. The most accurate models are the rather complex GBR and ANN algorithms. However, also much simpler models, such as the LAR model, obtained comparable predictive accuracy on the (inner-generational) validation set. This indicates that limited model complexity is sufficient to produce adequate results for the cost estimation within the same product generation. For the inter-generational testing set the LAR model yielded a NMAE of 7.86% on total cost estimation and 15.09% on assembly level. This corresponds to an increase in prediction error of +3.00 ppt (total) and +1.11 ppt (assembly) compared to the GBR model. We complement the existing literature on product cost estimation by confirming the complexity-accuracy relationship of machine learning models for highly complex products with a large variety of features when conducting inner-generational cost estimation. Notably, on inner-generational cost estimation, the superiority of highly complex models against simpler approach vanishes, indicating that not always high complexity is required to achieve sufficient levels of accuracy.

2nd research aspect (*Does big data lead to a substantial increase in cost estimation accuracy compared to intermediate amounts of data?*): The predictive accuracy of machine learning models usually increases with larger training samples, given data quality is sufficient (Goodfellow et al., 2016; Grolinger et al., 2014). In the cost estimation literature, the number of training examples for product cost estimation is merely of intermediate quality with a few hundred observations and only few features (Caputo & Pelagagge, 2008; Chou et al., 2010; Loyer et al., 2016). Our results also show that the utilization of high-volume data and high number of input features leads to a substantial increase in cost estimation accuracy. In particular, the relatively high number of features was found to be necessary to capture the complexity of passenger cars. When deploying training data with intermediate volume (300 order codes for each car model) and intermediate number of features (top 20 most important features), the NMAE increases from 0.88% to 3.08%. We add to the cost estimation literature by demonstrating the significance of the impact of big data on predictive accuracy. These results are important given the recent popularity of big data not only in practice, but also in research (Wlodarczyk & Hacker, 2014).

3rd research aspect (*How to improve the predictive performance of multi-generational product estimation?*): Many manufacturing companies incorporate multi-generational product development; however, most research approaches cost estimation primarily from the perspective of a single generation (Cai & Tyagi, 2014; Tyagi et al., 2015). This study shows that when applying multi-generational product cost estimation, it is important to consider the average cost changes from one generation to another. We propose a novel method incorporating the estimated net earnings of a product to improve cost estimates for multi-generational cost estimation, reducing the NMAE of the GBR model for the complete cost estimation from 10.12% to 4.86%. For the assembly-level cost, the NMAE of the GBR method decreases from 15.93% to 13.98%. We thereby add to existing cost estimation literature by introducing an easy-to-use method, which is especially useful when design engineers have only limited information about the cost structure of future products. Due to limited data availability many important factors that were suggested by cost experts, such as sales volume, technical standards, vertical range of manufacture, commodity prices, and manufacturing location, could not be tested.

Further research could expand this initial analysis of multi-generational adjustment by incorporating additional factors.

4th research aspect (*Does machine learning technology lead to more accurate cost estimates than calculations from cost experts when dealing with highly complex products?*): Most product cost estimation studies compare machine learning models with each other (Chou & Tsai, 2012; Deng & Yeh, 2011; Loyer et al., 2016) or with traditional statistical models such as regression models (Cavalieri et al., 2004; Verlinden et al., 2008). Research comparing machine learning methods with estimations from human experts based on experience (so-called expert judgment) is limited. Some research indicates that machine learning can be more accurate than manual estimates from experts for products of rather intermediate complexity (Bendul & Apostu, 2017; Caputo & Pelagagge, 2008). We contribute to this stream of literature by showing that state-of-the-art machine learning models can produce more accurate cost predictions for total product costs than cost experts, even for highly complex products such as passenger cars.

We also show that the performance depends on the level of product detail. On more granular cost levels (such as for the assembly levels of body, electrics, chassis, and engine), cost experts outperform machine learning methods. Our findings thereby challenge the often-perceived superiority of machine learning models in the field of product cost estimation (Caputo & Pelagagge, 2008; Cavalieri et al., 2004; Deng & Yeh, 2011; Loyer et al., 2016; Verlinden et al., 2008). On the assembly level, however, the cost prediction accuracy was strongly affected by the accuracy of the generation adjustment factors. On the assembly level, due to missing data availability, the adjustment factors were much more inaccurate and therefore the machine learning predictions also performed much worse. The comparison of different cost granularities (in terms of cost breakdown) with expert knowledge on more comparable ground is left for future research.

Based on the interviews with cost experts we find that complex machine learning models are considered most adequate during the design phase and early development phase of new product development when there are rapid changes of product designs, low integration of expert knowledge, and cost estimates are not used for cost goal setting. During the later stages of new product development, more detailed and comprehensible cost estimates are required. This is in line with the literature, which suggests that analogical cost estimation techniques (i.e., regression models, neural networks) are mostly applicable during the design stage of new product development, while analytical techniques (i.e., break-down approach, activity-based costing) are more adequate at later development phases (Chou et al., 2010; Farineau et al., 2001). In our interviews, we also find that the lack of interpretability and transparency is the most critical limitation for practical implementations. This corresponds with Verlinden et al. (2008), who find that cost engineers in most cases still prefer more comprehensible models despite being less accurate. It would be interesting to further investigate the role of interpretability when adopting machine learning models in practice. Future research could examine the trade-off between accuracy and interpretability in the context of product cost estimation or investigate what factors determine the relative importance of interpretability/transparency and accuracy when selecting machine learning models for cost estimation.

5th research aspect (*Can machine learning models select the most important cost drivers reliably?*): Machine learning can potentially be used to select and quantify cost drivers for product cost estimation and activity-based costing systems (K. Kim, 2003; Kostakis et al., 2008; Loyer et al., 2016). However, little is known about whether the machine learning approach actually produces valid and reliable results in this aspect. Our findings show that machine learning techniques can indeed capture the most important cost drivers reliably. However, the high importance scores indicate that models tend to oversimplify the cost driver structure and place too much importance on a small subset of features. So far, machine learning is primarily about the identification of patterns and correlations and less about the detection of causal relationships (Scholkopf et al., 2021). One problem of statistical analysis especially in big databases is thereby the pitfall of spurious correlations, where associations between variables are assumed, which are not causally justified (Calude & Longo, 2017). Thus, the extraction of knowledge from machine learning models yields serious epistemological problems when interpreting results and converting them back to conceptual information (Elish & Boyd, 2018). We add to the literature by a better understanding about the reliability of machine learning and big data technology to extract and quantify cost drivers.

6th research aspect (*Can machine learning models reliably estimate the relationship between cost drivers and manufacturing cost?*): Some research suggests using machine learning methods to get additional cost-engineering insights about the relationship between features and manufacturing cost (Chan et al., 2018; Loyer et al., 2016). We empirically analyze the reliability of machine learning methods to produce engineering insights about the cost structure of a product. We thereby answer the call by Verlinden et al. (2008) for additional research to identify ways to make complex machine learning models more appealing and user friendly to cost engineers. We find that machine learning methods can indeed reliably quantify the average feature costs in most cases. However, the evaluation with cost experts revealed that machine learning and big data in some cases produce misleading outcomes. This empirical study provides a better understanding on the validity of machine learning and big data technology for obtaining cost-engineering insights. Cost engineers can benefit from machine learning and big data with new and reliable insights about their products, however, need to be careful as misleading outcomes can also be produced. This is in line with Partovi and Anandarajan (2002) who state that artificial intelligence tools cannot and should not be depicted as a replacement for expert judgment. In this study, we trained machine learning models on real company data and evaluate results with costs experts, however, we did not operationalize the machine learning and big data techniques. Thus, the investigation of the operationalization in the development process, the usage in development teams, and the establishment of human-machine interfaces are left for further research.

3.5.2 Limitations

The results of this study must be seen in the light of some limitations. First, we implement a single-case design where the findings are neither statistically generalizable nor reproducible, potentially underlie biased qualitative data collection, and subjective interpretations (Cooper & Slagmulder, 2004b). Despite the common concerns of case study research, the limitations come along with the major strengths of the approach: We are able to obtain detailed insights into the applicability of such technology in an actual environment; we can qualitatively examine the properties of machine learning and big data usage; and we *analytically generalize* the cost estimation literature (Yin, 2018). Finally, the results are limited by only considering direct material costs. Due to limited data availability of other cost types, the analysis of total manufacturing cost was out of scope. Nonetheless, the study covers a large proportion of the total cost of passenger cars.

3.6 Appendix B

Table 19: Model implementation and hyperparameter tuning

Model	Implementation
ANN	The implementation of the (back-propagation) ANN is based on the libraries *keras* (Chollet & others, 2015) and *tensorflow* (Abadi et al., 2015). To tune the hyperparameters, we employed a combination of grid search and randomized search. First, the topology is selected, which defines the structure of the neural network by its layers and nodes. Second, we adjust the activation function, initialization method, optimizer, learning rate, dropout rate, number of epochs, and the batch size. The loss of the optimization algorithm is calculated by the mean squared error. To find an adequate topology, that sufficiently fits the complexity of the problem, we apply grid search. To compare different topologies, we first defined a baseline model. The baseline neural network is trained with the *Adam* optimizer (Kingma & Ba, 2014). Further, we used the default values of the Keras implementation: linear activation function, Glorot uniform (Xavier) initializer, no dropout regularization, 10 epochs, and a batch size of 32. We tested the combination of three different depths and widths of hidden layers. The width of each hidden layer may be 461 (input layer size), 1,000 and 2,000. An increasing width allows for high-order interactions between the features. The depth may be one (so-called *shallow* neural network), two or three hidden layers. The best results had been achieved with a two hidden layer topology with 2,000 nodes each. The remaining parameters had been tuned with randomized search. Due to the high number of combinations and extensive learning time, we chose randomized search for practical reasons. The parameter grid covers the activation function: <u>Linear</u> or rectified linear unit (ReLu). In the case of linear activation, the layer weight is initialized by <u>Glorot uniform</u> (Xavier). In the case of ReLu, the weights are initialized by the *He normal* method. The optimizer may be Adam or RMSprop. The learning rate may be 0.0001, <u>0.001</u> or 0.01. The dropout rate may be <u>0%</u> (no dropout regularization), 10% or 50%. We put a dropout layer between each fully connected layer. The batch size might be <u>32</u> or 64. The default settings are indicated by underline. The range of the number of epochs is set from 1 to 20. The randomized search was conducted for 20 iterations. The limited number of iterations is set due to practical reasons. The parameter settings that gave the best results on the validation data set are: Activation: ReLu, layer weight initializers: He normal, optimizer: Adam, learning rate: 0.01, dropout rate: 0%, batch size: 32, and number of epochs: 11.
CBR	The main hyperparameters of the CBR model are the distance metric, weighting scheme, and normalization of continuous variables. For the distance metric, we distinguish between the Manhattan or Euclidean distance metric. For the weighting scheme we distinguish between equal weights (Chou et al., 2010) and the *Binary-Dtree* method of Doğan et al. (2008). According to the Binary-Dtree method, the weight is 1 (selection) if a feature is present in a separate decision tree, otherwise, the weight is 0 (deselection). The weights are taken from the DTR model. Since the continuous variables have values above 1, they systematically carry more weight. To counter that effect, we normalize the continuous features by min-max transformation or standardize by removing the mean and scaling to the standard deviation. We performed grid search over all hyperparameters: distance metric (Manhattan or Euclidean), weighting scheme (equal weights or Binary-Dtree) and normalization of continuous variables (no adjustment, min-max scaling, or standardization). The parameter settings that gave the best results on the validation data set are: distance metric: Manhattan, weighting scheme: no and adjustment of continuous variables: no.

DTR	The DTR method is based on the *scikit-learn* package, which applies the CART (Classification And Regression Trees) algorithm. To tune the hyperparameters, we employed grid search. The hyperparameters are the maximum depth of the tree, the minimum number of samples in the leaf nodes, and the minimum number of samples required to split internal nodes. The maximum depth may be unlimited or limited to 10, 15, 20, 25, or 30 nodes. The maximum depth is set to prevent overfitting. The minimum number of samples in leaf nodes may be 1, 2, 4, 6, 8, or 10. The minimum sample split may be 2, 10, or 100. The default settings are indicated by underline. The parameter settings that gave the best results on the validation data set are: maximum depth of trees: 20, minimum number of samples in leaf nodes: 6, and minimum number sample for node splits: 10.
ELR, LAR	The ELR and LAR algorithms are based on the *scikit-learn* package. We compared the linear least squares regression with L1 (lasso) and L1&L2 (elastic net) regularization. Both methods are chosen over the regular OLS regression and ridge (L2) regression since the coefficients can be forced to be positive. The positive regression coefficients facilitate the economic interpretation of cost drivers. To tune the hyperparameters, we employed grid search. For both methods the hyperparameter grid contains the availability of an intercept (yes/no), alpha value (0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0), and normalization of regressors (yes/no). The alpha value is the constant that is multiplied with the penalty term. In the case of normalization, the regressors are normalized by subtracting the mean and dividing by the L2-norm. In the case of ELR, the parameter grid is extended by a mixing parameter that trades off the L1 and L2 penalty. The parameter ranges from 0 (L2 penalty) to 1 (L1 penalty). We set the mixing parameter to 0.3, 0.4, 0.5, 0.6, or 0.7. The default configurations are indicated by underline. The parameter settings of the LAR model, which gave the best results on the validation data set are: alpha value: 0.2, intercept: yes, and normalization: no. The parameter settings of the ELR model that gave the best results on the validation data set are: alpha value: 0.2, intercept: yes, normalization: no, and mixing parameter: 0.7.
GBR	The implementation of the GBR model is based on the *scikit-learn* package. The most relevant hyperparameters are the maximum depth of the tree, the number of boosting stages (M) and the learning rate (v) (Elith et al., 2008; Landry et al., 2016; Y. Xia et al., 2017). Y. Xia et al. (2017) state that the tree complexity is controlled by two alternative hyperparameters: the maximum tree depth (number of edges along the longest path) and the tree size (number of terminal nodes in a tree). To tune the aforementioned hyperparameters, we employed grid search. Grid search is important when tuning boosted regression tree, since the regularization requires the joint optimization of hyperparameters (Elith et al., 2008). The maximum depth restricts the number of decision nodes in the trees, which may be 2, 3, or 4. The boosting stages may be 100, 1,000, 2,000, or 3,000. The learning rate may be 0.05, 0.1, or 0.15. The default settings are indicated by underline. The parameter settings that gave the best results on the validation data set are: maximum depth of tree: 3, boosting stages: 3,000 and learning rate: 0.1. The GBR model of *scikit-learn* does not allow multi-target regression. To forecast the cost per assembly group, we trained four individual GBR models.
LSVM	The implementation of the (epsilon-insensitive) LSVM model is based on the *scikit-learn* package. To tune the hyperparameters, we employed grid search. The hyperparameters are the regularization parameter and the epsilon margin. The regularization parameter controls the penalty of observations outside the epsilon distance and prevents overfitting. Errors that are lower than the epsilon margin are ignored. The regularization parameter may be 0.1, 0.5, 1, 1.5, 10, 100. The epsilon parameter may be 0, 1, 10, 100. The default settings are indicated by underline. The parameter settings that gave the best results on the validation data set are: regularization parameter: 10, and epsilon margin: 0.

Table 20: The relationship between the average net earnings ($\overline{NE}$) and the average direct material cost ($\overline{DMC}$). The regression analysis encompasses four car models from the predecessor generation (I).

	$\overline{DMC}_I$
Const.	3408.5876 (3.080)
$\overline{NE}_I$	**0.4881**** (16.446)
Observations	4
R^2	99.3%
Note:	**p<0.01

Table 21: Background information and further explanation to the evaluations from cost experts.

3rd aspect: Improving predictive performance for multi-generational product estimation	
C_1: feasibility	*"The change of cost between two generations can be sufficiently explained with this approach quite easily. More accurate estimates can be produced by taking into account the sales volume. Other factors such as the vertical range of manufacture, location of site, and commodity are of secondary importance."* (C2)
C_1: feasibility	*"An important success factor for this approach is the target contribution margin when deciding on product characteristics. There will be very little product characteristics with a contribution margin less than 30%."* (C3)
C_1: feasibility	*"This correlation is not always correct for all features. In some cases, such as exhaust emission regulations, the product net earnings do not increase since there is no additional value for the customers. In such cases, manual adjustments to the cost estimates would be required. On the other side, one could argue that the customers will have more money in their pockets over time, which they again spend proportionally for mobility. Therefore, the additional cost can be compensated. This coincides with the approach from the sales department for the prediction of the sales prices at the early stage of new product development. According to the sales department, most customers allocate a fixed share for mobility and as the net household income increases, the net earnings for products will be higher as well."* (C1)
C_3: usability	*"The net earnings are available already in the early phase of new product development. Therefore, this approach is applicable to estimate the cost change. However, one need to make sure that the net earnings match the technical descriptions of a product. Sometimes at the early stages, the pricing by the sales department is conducted with little alignment and agreement with the technical department. As a result, the estimated direct material cost based on the technical descriptions can differ from the estimated direct material cost based on pricing information."* (C4)
4th aspect: Comparison of cost estimation accuracy: Machine Learning vs. expert judgment	
C_1: feasibility	*"The cost split on assembly level is not always consistent over the generations. In practice for each project and generation the composition of cost must be discussed and adjusted if required. The cost adjustment can be roughly conducted by a small number of key features."* (C1)
C_2: usability (support)	*"I need the resilience of the cost predictions and detailed information such as the underlying assumptions. In the decision-making process it is important to be persuasive and engage the team on a joint cost goal."* (C3)
C_2: usability (support)	*"In some situations, these analyses are helpful to quickly generate 'first sight' cost estimates. Especially, as we receive much more cost prediction requests, which cannot be handled manually anymore. However, one problem is the complexity of such tools. I am afraid that many people won't accept such estimates as they are less comprehensible. This is problematic, especially when the produced cost estimates place limitations on people such as cost goals."* (C4)
C_5: intention to use	*"Maybe we all must learn to trust such tools when they are sufficiently accurate without having to explain everything in detail again and again. It would be interesting to test such tools in practice. I could image that delivering cost ranges instead of precise cost values could increase acceptance."* (C5)

5^{th} aspect: Reliability of machine learning-based cost driver selection

C_1: feasibility	*"The engine performance is indeed the most important cost driver; however, I would not claim that this feature alone explains 38% of all cost. More likely, the model does not only consider direct effects but also indirect effects of engine performance such as gearbox, cooling system, brake system, tire, etc."* (C4)
C_2: usability (support)	*"This analysis can help increasing trust into machine learning models. However, the acceptance highly depends on who you are talking to. One individual may be fine with a black-box model while others want to have more information. In addition, it helps to conduct further prioritization of our list of key product features. This is important as the effort of data collection can be reduced for less important features."* (C2)
C_2: usability (support)	*"This prioritization helps focusing on the essential cost drivers during new product development. Moreover, such tools are easier to use when less cost drivers need to be adjusted."* (C3)
C_4: usability	*"This analysis can also be helpful for the cost optimization program of [AutomotiveCompany]. Thereby, the cost reduction focus can be placed on the right components: What components are highly important, and what are less worthwhile to examine in detail?"* (C5)

6^{th} aspect: Reliability of machine learning models to estimate the relationship between features and manufacturing cost

C_1: feasibility	*"An important input factor is the purchasing volume of components. It is possible that an 18-inch braking disk with high purchasing volume is less expensive than the smaller 16-inch disk with low volume."* (P6)
C_2: usability (support)	*"At the planning stage of product development, standard prices from the sales department are used to approximate the cost of combustion engines. (...) The utilization of actual costs per performance can help making better decisions at choosing the optimal engine performance for a given market."* (P1)
C_2: usability (support)	*"This global sensitivity analysis certainly is interesting. I have never conducted a business case with interdependencies between product features. This would be impossible to calculate manually due to the enormous complexity. Therewith, better product decisions can be made, since the complete product range can be considered and not only one reference configuration."* (P5)
C_2: usability (support)	*"This analysis also helps to identify cost outliers for certain parts. For some special parts with low take rates suppliers have above average margins. (...) Therefore, the reference model does not truly reflect the actual mean cost over the product range."* (C1)
C_5: intention to use	*"This analysis is helpful for the planning teams of product platforms. When discussing the engine performance and engine size of a new car, we immediately must point out all cost effects. In practice this is extremely difficult. (...) If due to a larger engine the installation space is getting too tight and new crash measures or pedestrian protection is required, the actual cost can be identified not until the very end of product design. But still, we need this discussion of cost at the early phase where the decisions are made."* (P3)

Calculation of the mean two-factor interaction effect

The calculation of the *secondary elementary effect* (SEE_{ab}) in the case of a binary primary feature X_a is denoted in Equation 18. The *two-factor interaction effect* (EE_{ab}) is calculated by subtracting the elementary effects of both features (Equation 13) from the secondary elementary effect. The *mean two-factor interaction effect* (μ_{ab}) can be considered as a global sensitivity measure for two-way interactions.

$$SEE_{ab}(X) = \begin{cases} f^G(X_a{=}1, X_b{=}1, \cdot) - f^G(X_a{=}0, X_b{=}0, \cdot) & \text{if } X_b \in B \\ f^A(X_a{=}1, X_b{=}X_b{+}1, \cdot) - f^A(X_a{=}0, \cdot) & \text{if } X_b \in C \end{cases} \tag{18}$$

$$EE_{ab}(X) = SEE_{ab}(X) - EE_a(X) - EE_b(X) \tag{19}$$

$$\mu_{ab} = \frac{1}{n}\sum_{i=1}^{n} EE_{ab}(X^{(i)}) \tag{20}$$

$$\sigma_{ab} = \sqrt{\frac{1}{n\text{-}1}\sum_{i=1}^{n} \left(EE_{ab}\left(X^{(i)}\right) - \mu_{ab}\right)^2} \tag{21}$$

where

f^m: *machine learning model* $m := \{A(NN), G(BR)\}$
B: *binary features*
C: *cardinal features*

4 Machine Learning in Product Cost Estimation: The Trade-Off between Accuracy and Explainability

Abstract

This chapter investigates the trade-off between accuracy and explainability of machine learning models in the context of product cost estimation during new product development. First, we empirically confirm the often-implied inverse relationship between explainability and accuracy from the perspective of cost experts. Second, we show that the importance of explainability relative to accuracy perceived by cost experts in different situations during new product development is an important factor when selecting between alternative machine learning models. Third, we identify several factors that determine the perceived relative importance of explainability to accuracy during new product development: phase of product development, information uncertainty, level of cost granularity, and the gap between target cost and planned cost. This experimental study shows that complex machine learning models, such as GBR (gradient boosted regression), are only adequate in few situations of product development. Machine learning models are adequate during the early development phase when dealing with low information uncertainty, high level of cost granularity, and a low gap between target cost and planned cost. In the majority of cases, more basic models such as CBR (case-based reasoning) and multiple linear regression (MLR) are preferred, although they are much more inaccurate.

Keywords: Machine learning; Cost estimation; Interpretability problem; Design of experiments

4.1 Introduction

The application of machine learning models for cost management during new product development yields new and enriching opportunities in many aspects (Caputo & Pelagagge, 2008; Loyer et al., 2016; Q. Wang, 2007). However, the applicability of complex machine learning methods is often limited by the ability to provide explainable results (Chou et al., 2010; Coussement et al., 2017). In the context of machine learning, explainability is the ability to demonstrate the reasons for a behavior or the ability to produce insights about the cause of an outcome (Gilpin et al., 2018). To achieve acceptance, machine learning systems must provide sufficient explanations of their decisions and predictions. There are machine learning models that are more explainable than others, and there is often a trade-off between accuracy and explainability. Accuracy in this context is defined as the predictive performance of an estimation model based on the difference between the predicted values and the actual results. Usually, the most accurate machine learning models are not very explainable, and vice versa (Adadi & Berrada, 2018). This inverse relationship between explainability and accuracy is known as the interpretability problem. As a consequence, state-of-the-art machine learning techniques with high accuracy are often referred to as *black boxes* (G.-H. Kim et al., 2004). This is problematic since explainability is often required to build trust and acceptance of machine learning models (X. Zhang et al., 2021). The trade-off between explainability and accuracy is well-known in the cost management literature (Cavalieri et al., 2004; Loyer et al., 2016; Verlinden et al., 2008). Cost management during new product development involves planning and reporting activities, including cost goal setting and decision making support (Wouters et al., 2021). For these tasks, explainability is needed to interact with cross-functional development teams. At the same time, sufficient predictive accuracy is necessary to make informed decisions with regards to the profitability of future products. This chapter empirically investigates the trade-off between explainability and accuracy in the context of product cost estimation depending on several factors during new product development.

Artificial intelligence research has attempted to solve the interpretability problem by introducing XAI, which aims to make artificial intelligence systems more understandable to users (Adadi & Berrada, 2018). However, improving explainability of complex machine learning models is difficult and usually only possible to a limited extent (Barredo Arrieta et al., 2020; Emmert-Streib et al., 2020; Linkov et al., 2020). At the same time, it is suggested that the applicability of complex machine learning models depends on the relative importance of explainability and accuracy for certain tasks and situations (Baryannis et al., 2019; Lee & Shin, 2020; Rey et al., 2017). Since the relative importance between situations can be different, machine learning models at different levels of complexity are required. During new product development however, it is mostly unknown which factors determine the relative importance between explainability and accuracy. Why is explainability (relative to accuracy) more important in one situation than in another? In this chapter, we raise the following research question: *Which factors determine the relative importance between explainability and accuracy for product cost estimation during new product development?*

An experimental study is applied to investigate this research question. A fractional factorial experiment with cost experts from a German car manufacturer (with the disguised name *AutomotiveCompany*) is conducted to analyze the impact of five factors. The participants are product controllers that are dealing with product cost estimation during new product development. Statistical tests are conducted to analyze the effects of different treatment conditions on the importance of explainability to accuracy. In addition, the findings are triangulated with qualitative data to validate the hypothesized argumentations.

This chapter contributes to the literature in several ways. First, we empirically confirm the often-implied inverse relationship between both attributes from the perspective of product controllers (Cavalieri et al., 2004; Loyer et al., 2016; Verlinden et al., 2008). Second, we empirically confirm the often-assumed significance of the relative importance of explainability to accuracy in the model selection process for product cost estimation (Cavalieri et al., 2004; Loyer et al., 2016; Verlinden et al., 2008). As a result, this experiment shows that complex machine learning models, such as GBR, are only adequate in few situations during product development. In most cases, more basic models such as the CBR and MLR are preferred unless being much more inaccurate. This challenges the current research trend of using complex machine learning technology for cost estimation (Caputo & Pelagagge, 2008; Cavalieri et al., 2004; Chou & Tsai, 2012; Deng & Yeh, 2011; Loyer et al., 2016; Verlinden et al., 2008). Third, we identify several factors that determine the relative importance of explainability to accuracy during new product development: *phase of product development*, *information uncertainty*, *level of cost granularity*, and the gap between target and planned cost (further referred to as *target cost gap*). Specifically, we show that explainability is more important than accuracy in the fuzzy front-end stage, when dealing with low uncertainty, low level of cost granularity, and high target cost gap. We thereby extend the XAI literature stream by introducing context-specific factors that significantly influence the relative importance of explainability to accuracy. So far, most research in this area is conceptual and the literature mostly considered general factors, such as the need for interaction with models (Frias-Martinez et al., 2005), the available trust (Alonso et al., 2015) or consequences of inaccurate predictions (Baryannis et al., 2019).

This chapter is structured as follows: In section 4.2 we provide an overview about the interpretability problem, explainable artificial intelligence, and factors that are expected to determine the trade-off between accuracy and explainability. In section 4.3, we provide an economic and technology acceptance perspective of the underlying problem and formulate five hypotheses that are based on the new product development literature. In Section 4.4, the fractional factorial experiment is described. Section 4.5 discusses the validity and reliability of the experiment and Section 4.6 reports on the significance of the factors. Section 4.7 evaluates the findings with qualitative data and Section 4.8 concludes the chapter.

4.2 Literature Review

In the following, we introduce the interpretability problem and define the terms interpretability and transparency in the context of machine learning. We then present the current approach to tackle the interpretability problem and specify factors that determine the trade-off between both attributes.

4.2.1 The Interpretability Problem of Machine Learning Algorithms

The ***interpretability problem*** describes the inverse relationship between accuracy and explainability of machine learning algorithms. In general, machine learning models can be evaluated by the three dimensions: accuracy, interpretability, and efficiency (Liu et al., 2016). High interpretability is required to make the reasoning of machine learning models more understandable to users and developers. As an example, explanations are needed for justification, improvement, and knowledge discovery (Adadi & Berrada, 2018). However, “the most accurate AI [artificial intelligence]/ML [machine learning] models usually are not very explainable” (Adadi & Berrada, 2018, p. 52145). This can be explained by the different levels of complexity of machine learning models. Complex models have higher flexibility than basic approaches, allowing for more complex problems to be solved (Barredo Arrieta et al., 2020). On the other side, highly complex models, such as deep neural networks, display poor interpretability, which makes validation by analysts or domain experts much more difficult (Huysmans et al., 2011). Rudin (2019, pp. 206–207) questions the "widespread belief that more complex models are more accurate, meaning that a complicated black box is necessary for top predictive performance". Moreover, when dealing with problems with reasonable features and structured data, there is often no substantial difference in accuracy between complex and simple machine learning models. Nonetheless, the highest accuracy scores for large data sets and complex problems are often only achieved by models with high complexity that even experts have problems to interpret, which then creates stress between interpretability and accuracy.

Gunning and Aha (2019) formalized the trade-off between explainability and predictive performance by arranging several machine learning models at different levels of complexity (deep learning, random forests, support vector machines, Bayesian belief nets, and decision trees) along the two dimensions. Thereby, deep learning was classified as having high accuracy but low explainability; and decision trees were specified with the highest explainability and lowest accuracy. Barredo Arrieta et al. (2020) extended the representation of the trade-off between model interpretability and accuracy, which is depicted in **Figure 15**.

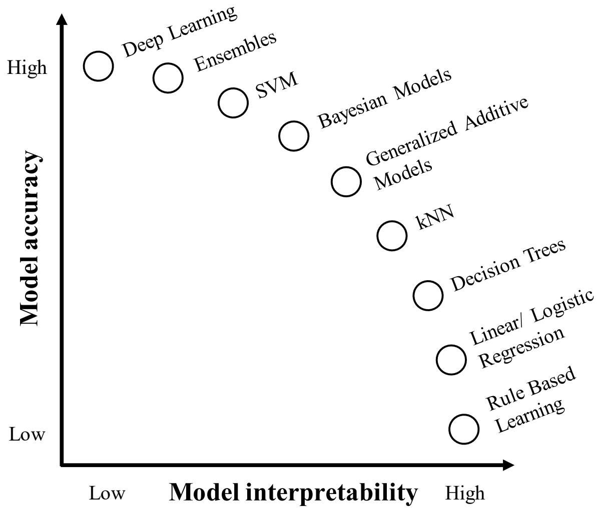


Figure 15: Trade-off between model interpretability and performance (representation from Barredo Arrieta et al. (2020))

4.2.2 The Difference Between Transparency and Interpretability

Gilpin et al. (2018) describe ***explainability*** as the ability to provide reasons for a decision or produce insights about the cause of an outcome to gain trust of users. Lipton (2018) considers explainability as an umbrella term for transparency and interpretability, dividing explainability into the two categories transparency (*how does the model work?*) and interpretability (*what can the model tell me?*). ***Transparency*** can be considered at the level of the entire model, of individual parameters, and at the level of the training algorithm. Gedikli et al. (2014) states that transparency facilitates the understanding of the reasoning of a model, where transparency can be distinguished between objective transparency (the model reveals the actual mechanisms) and user-perceived transparency (the model reveals the mechanisms important to users). ***Interpretability*** is a domain-specific and subjective concept; accordingly, there is no formal definition (Lipton, 2018; Rudin, 2019). An attempt had been made by Doshi-Velez and Kim (2017, p. 2), defining interpretability as the "ability to explain or to present in understandable terms to a human". Interpretability can be achieved by textual descriptions, visualizations, local explanations near the input point, and examples of decisions with similar inputs (Lipton, 2018). In addition, high interpretability can be achieved by linear associations and monotonic functions that enable intuitive reasoning (Hall & Gill, 2018).

4.2.3 Explainable Artificial Intelligence

Artificial intelligence research aims to solve the interpretability problem with ***XAI***. Adadi and Berrada (2018) describe XAI as a research field that aims to demystify black boxes and make results from artificial intelligence systems more understandable to users and developers. The objective is to develop models with higher explainability while maintaining high levels of accuracy. Thereby, explanations from artificial intelligence systems must use communicable representations such as visual and logical forms, to reduce the room for misinterpretation (Preece, 2018). Guidotti et al. (2019) present a classification scheme for XAI models, where a method needs to explain its results, provides representations to inspect the model, or provides a transparent solution for the problem. XAI tends to refer to a movement in reaction to the transparency and trust concerns of artificial intelligence systems, more than to a formal approach (Adadi & Berrada, 2018).

The main objectives of XAI are verification, improvement, knowledge discovery, and compliance to legislation (Meske et al., 2020; Samek et al., 2017). Further, developers can only improve machine learning systems if the model can be interpreted, and its weaknesses are reliably determined. Users of artificial intelligence systems primarily need explainability to compare the system's reasoning with own argumentations and analyze the reliability of results (Meske et al., 2020). Doshi-Velez and Kim (2017) argue that explanations are required because of the incompleteness in problem formalization and problem understanding. Finally, the data used to train models may contain human biases and prejudices, which poses a critical risk to the validity of machine learning applications. If trained on biased data, machine learning models can lead to unfair and incorrect behavior (Guidotti et al., 2019). XAI can help to detect biases and solve unintended behavior.

Explainable artificial intelligence encompasses several methods, which make opaque machine learning models more explainable. Hall and Gill (2018) present three steps for establishing explainable and trusting models, which are 1) achieving a deep understanding of data sets, 2) using models with interpretable inner workings, and 3) applying techniques that generate explanations for highly complex models. Another technique is *post hoc interpretability*, which involves the provision of explanations from the output of black-box models after training is completed (Peake & Wang, 2018). Thereby, the high accuracy levels of complex machine learning models can be maintained, while improving the comprehensibility of models through additional explanations. Nonetheless, the improvements of explainability through XAI are limited and these techniques usually cannot completely solve the interpretability problem (Alonso & Magdalena, 2011; Barredo Arrieta et al., 2020; Emmert-Streib et al., 2020; Linkov et al., 2020).

4.2.4 Factors that Determine the Importance of Accuracy and Explainability

Despite the attempts of XAI to improve interpretability without sacrificing accuracy, a fundamental trade-off remains. Therefore, data scientists need to choose between machine learning methods according to the specific accuracy and interpretability needs in a given situation and task (Boehm et al., 2019; Tripathi et al., 2021). Compromising between the accuracy and complexity of machine learning models can be subjective and the way to reach a balance is usually achieved through discussions (González et al., 2007; Rey et al., 2017). The assessment of models requires an agreement between data and business experts. Thus, looking for a good trade-off is one of the most challenging tasks in system modeling (Alonso & Magdalena, 2011). Therefore, task-adaptive and user-adaptive machine learning systems have been proposed to select models according to the interpretability needs in a given situation (Frias-Martinez et al., 2005; Lee & Shin, 2020). For example, Lee and Shin (2020) propose a task-dependent system, where white-box and black-box algorithms are compared according to their accuracy scores when interpretability is not required (i.e., when developing a chatbot for customer service). When, however, interpretability is required (i.e., when developing a recommender system) only white-box algorithms are evaluated. Then, the model is selected with the highest accuracy score and an acceptable level of interpretability.

Some research has considered factors that make explainability or accuracy relatively more important and thereby determine the machine learning model selection process. First, explainability is usually more important when the objective is to ***integrate*** knowledge of human experts in the machine learning system and to ***interact*** with models to improve predictions (i.e., add new or modify decision rules in decision trees) (Alonso et al., 2015; Frias-Martinez et al., 2005). An explainable model can be more easily improved when users know why a system produces certain results and the system can be adjusted accordingly (Adadi & Berrada, 2018). Second, significant ***costs of inaccurate predictions*** can place more importance on accuracy relative to explainability (Baryannis et al., 2019). Third, explainability is more important when a machine learning model needs to be ***validated*** against the prior knowledge of human experts (Alonso et al., 2015).

If a problem is analyzed in-depth and already validated in actual applications, explainability is less relevant (Doshi-Velez & Kim, 2017). Fourth, explainability is more important when there is ***low trust*** toward artificial intelligence systems and there is the need to convince recipients of outcomes (Alonso et al., 2015). Using explainable models justifies outputs and builds trust into models (Adadi & Berrada, 2018). Fifth, explainability is more important when the primary goal is to ***understand*** the data and discover relationships between variables (Adadi & Berrada, 2018; Baryannis et al., 2019). Sixth, in the case of ***legally*** and ***ethically*** relevant tasks exact explanations are required and black-box models are not acceptable (Ribeiro et al., 2016). Explainable models contribute to the auditability of artificial intelligence systems by providing explanations to regulatory stakeholders (Barredo Arrieta et al., 2020). Finally, explainability is more important when there is only ***limited time to understand*** outcomes from machine learning models. In more profound analyses and scientific fields, users of artificial intelligence are willing to spend much more time to understand results compared to simpler tasks (Doshi-Velez & Kim, 2017).

4.2.5 Knowledge Gap and Research Question

The selection of machine learning models in many practical applications depends on the relative importance of accuracy and explainability for a given task. Since the relative importance may be different under certain situations, more or less complex machine learning models are adequate. However, it is mostly unknown what determines the relative importance between explainability and accuracy in the context of cost estimation during new product development. Why is explainability (relative to accuracy) in one situation more important than in another? In this study, we raise the following research question: *Which factors determine the relative importance between explainability and accuracy for product cost estimation during new product development?*

4.3 Theory and Hypothesis Development

4.3.1 An Economic Formalization of the Problem

We can formalize the model selection problem under the trade-off between accuracy and explainability with the economic model of *production possibility curves* and *indifference curves*. The economic model describes the choice between two goods based on the personal preferences of an individual. The production possibility curve describes the transformation between two goods. In most applications a concave transformation is assumed. The indifference curve represents all combinations of goods with an equal level of utility to which an individual is indifferent. The shape of an indifference curve reflects the individual preference according to the relative importance between the goods. To find an optimal combination of two goods, the *marginal rate of transformation* is compared to the *marginal rate of substitution*. The marginal rate of transformation is defined by the slope of the production possibility curve, whereas the marginal rate of substitution is defined by the slope of the indifference curve. **Figure 16** formalizes the optimal choice of accuracy and explainability in two situations. Thereby, the production possibility curve reflects the inverse relationship between accuracy and explainability of machine learning models. The relative importance between explainability and accuracy in certain situations is described by indifference curves. Each situation shapes an individual indifference curve according to the specific preferences. In this illustration, situation B exhibits a higher relative importance of explainability to accuracy than situation A. In other words, in situation B an individual would give up more accuracy for additional explainability than in situation A.

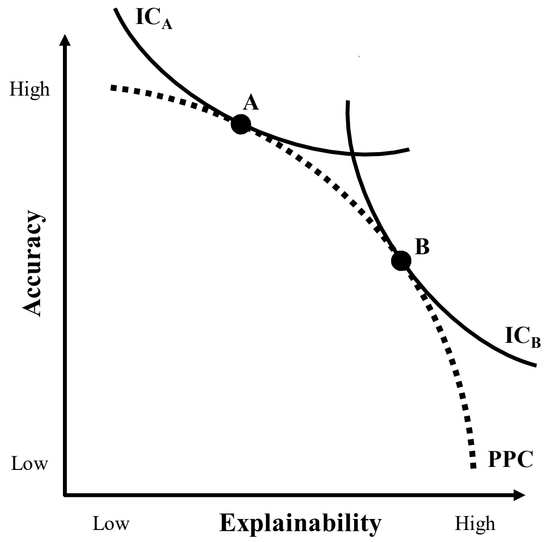


Figure 16: Illustration of the production possibility curve (PPC) between explainability and accuracy and indifference curves (IC). In situation B the importance of explainability to accuracy is higher than in situation A.

4.3.2 A Technology Acceptance Perspective on the Problem

Another theoretical perspective on the trade-off between accuracy and explainability is provided by the *technology acceptance model* (TAM) (Davis et al., 1989). The model introduces *perceived ease-of-use* and *perceived usefulness* as antecedents of *actual system use* of new technology. The factors of interpretability (or comprehensibility or completeness) and accuracy are commonly part of *information quality*, which determines perceived ease-of-use (Demoulin & Coussement, 2020; van der Linden & van de Leemput, 2015; Wixom & Todd, 2005; Zhu et al., 2012). In the model of Demoulin and Coussement (2020), *representational information quality* encompasses interpretability and *intrinsic information quality* includes accuracy. Additionally, the acceptance of new information technology often depends on the target users and the task environment (Moon & Kim, 2001). In order to incorporate task environment into the model, Dishaw and Strong (1999) extended the TAM by the *task-technology fit* (TTF) theory. In the context of information systems, task-technology fit is defined by “the matching of the functional capability of available software with the activity demands of the task” (Dishaw & Strong, 1998, p. 109). In the TAM-TTF model the assessment of users about perceived ease-of-use and perceived usefulness of information technology is not only determined by the characteristics of technology but also by the tasks for which they are deployed (Dishaw & Strong, 1999). The integrated model of TAM-TTF model is depicted in **Figure 17**.

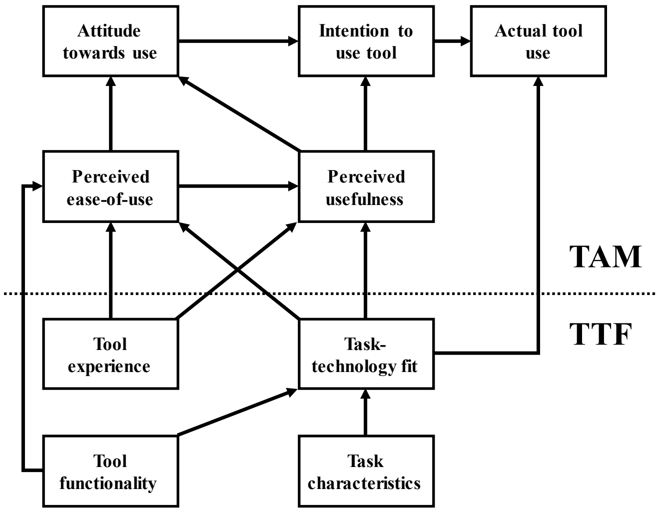


Figure 17: Integrated TAM-TTF model according to Dishaw and Strong (1999)

The integrated TAM-TTF model serves as a theoretical background concept for the underlying problem in this study. The model puts technology acceptance into perspective with interpretability and accuracy. Furthermore, the model describes the influence of task environment on technology adoption. Thereby, the TAM-TTF model describes the evaluation process of machine learning models depending on the relative importance of explainability and accuracy across different situations and tasks. In other words, when *representational information quality* is necessary to solve a task in a certain situation, white-box algorithms are more appropriate. When more importance is placed on *intrinsic information quality*, black-box algorithms are more likely to be chosen.

4.3.3 Hypothesis Development

In the following, we identify five factors which we expect to influence the relative importance of explainability and accuracy for cost management during new product development.

4.3.3.1 Development Phase Progress

The progress of a development project can be defined by the stage in the product development process in terms of timewise or conceptual maturity. It can be divided into the fuzzy front-end (FFE) phase and the development phase (J. Kim & Wilemon, 2002).

The FFE covers the period from opportunity identification until the start of development. The FFE includes product strategy formulation, ideation, product definition, and project planning (Khurana, 1998). During the early stages of new product development companies develop the product concepts and determine whether they should invest in the development of the product (Moenaert et al., 1995). This phase is characterized by complex information processing, ad hoc decision-making, and a dynamic and an interactive nature (Florén et al., 2018). J. Kim and Wilemon (2002) describe the information for decision making in the FFE phase as qualitative, approximate, and informal. In regard to cost estimation accuracy, there is often just a *first sight* estimate, which is mostly based on the experience of cost experts (Rush & Roy, 2000). In the later development stages, detailed cost estimates based on precise cost calculations are available. Thereby, companies tend to use computer-based tools to calculate manufacturing costs in greater detail. The development phase is characterized as quantitative, precise, and formal (J. Kim & Wilemon, 2002). While the FFE phase is about creating a blueprint and deciding on product feasibility, the development phase is about creating the product and *making it happen*. Accordingly, we assume that for the ***conceptual work*** of the FFE phase explainability is more important than accuracy due to the complex information processing and the qualitative and approximate nature. On the other side, we expect that for the ***implementation work*** of the more detailed development phase, accuracy is more important than explainability due to the focus on concept realization and its precise nature.

Another characteristic of the FFE phase is that decision-making is often based on assumptions. The selection of cost estimation methods is largely dependent on the progress of new product development and the available amount of information (Curran et al., 2007). This creates a dilemma for R&D managers. As new product development projects proceed, the quality of cost estimates increases due to the amount of information available. However, the influence on total cost decreases as more and more product characteristics are already determined. Reliable cost information is often obtained only in the late phase of product development (Bode, 2000). During the development phase, decisions are increasingly based on actual information and facts. In the early phase, development teams build on several assumptions that determine interpretations and actions (Hey et al., 2007). Planning too detailed in the very early phase is a common mistake in the front end of engineering programs (Lucae et al., 2014). Too much detail in the early phase increases the dependence on more assumptions that can result in rework and lower performance during later stages. In general, we expect that the ***assumptions*** for the cost estimates during the early phase need to be explained in greater detail, compared to the ***facts*** available in the development phase. In sum, we expect that the relative importance of explainability to accuracy is higher during the FFE phase of new product development (lower progress) than in the development phase (higher progress).

H1: *The importance of explainability relative to accuracy is higher in the FFE design phase of new product development than for the development phase.*

4.3.3.2 Information Uncertainty

Information uncertainty is an important property of new product development projects, which can be defined as the ambiguity or lack of information caused by the “lack of definition, lack of knowledge or lack of trust in knowledge” (Wynn et al., 2011, p. 4). Information uncertainty during new product development stems from various sources such as environment, technology, consumer, competition, and resources (Song & Montoya-Weiss, 2001). High task uncertainty of new product development projects places more importance on intensive communication in teams and results in greater information processing requirements (Tushman & Nadler, 1978). Further, perceived technological uncertainty moderates the relationships between cross-functional integration, synergy of resources, and efficiency in the development process (Song & Montoya-Weiss, 2001). Therefore, projects with high uncertainty should be executed using an organic approach (i.e., autonomous team structure) that promotes the ability of teams to process information (Patanakul et al., 2012). Thereby, the perception of high uncertainty increases the positive impact of cross-functional teamwork on the performance of projects. For more unpredictable and uncertain settings, product development is characterized by experimental and iterative problem solving approaches (Brown & Eisenhardt, 1995; Eisenhardt & Tabrizi, 1995). In the context of XAI applications, explainability is more important when the objective is to *integrate* knowledge of human experts in the machine learning system and to *interact* with models in order to improve predictions (Alonso et al., 2015; Frias-Martinez et al., 2005). All in all, we expect that ***expert knowledge*** is more important under uncertain conditions, which corresponds with a higher ***need for manual adjustments*** of cost estimates.

One way to cope with uncertainty is to create ***assumptions*** about unknown future events and variations of assumptions to engage in contingency planning (Allaire & Firsirotu, 1989). During new product development, there are often many assumptions toward the benefits of new products (Berchicci & Bodewes, 2005). Further, cost estimates often require many assumptions to create leeway to satisfy technical requirements and organizational objectives (Tyebjee, 1987). We expect that consumers of forecasts have lower expectations about accuracy when dealing with information uncertainty. In these cases, rough cost estimates are more likely to be accepted. On the contrary, we expect that accuracy is more important if there is certainty about information. In this case, cost experts can rely on ***facts*** rather than assumptions. Due to the importance of expert knowledge and dependence on assumptions, we expect explainability to be more important than accuracy when dealing with uncertainty.

H2: *The importance of explainability relative to accuracy is higher in the case of high information uncertainty than for low information uncertainty.*

4.3.3.3 Cost Granularity

Costs can be estimated at different levels of granularity, ranging from total cost of products to individual parts. Cost granularity therefore defines the level of detail or level of aggregation of cost with which a product can be described. At each of these levels, a specific (management) team is likely responsible for achieving the cost goals. At a more granular level, development teams primarily deal with realizing detailed tasks of few components, while at more aggregated level, people primarily deal with management tasks and overall product idea generation. For example, members of SE teams need to concrete solutions for their specific parts, whereas the team responsible for the overall product achieves goals by managing lower-level teams. Thereby, development teams at higher cost granularity (i.e., SE teams) mainly work with ***homogenous sets of few components***. As no direct lateral coordination between SE teams is required, explainability is less important. The main objective is to find an optimal solution for the specific set of components and, therefore, accuracy is needed. A detailed decomposition of cost allows for more accurate cost control (Filomena et al., 2009). Product development at more aggregated level (low-cost granularity) involves a ***heterogeneous set of many components***, which involves the management of multiple and diverse teams and the optimization of the overall concept of a product. For example, the overall team must understand the reasons for high costs so they can make trade-offs and discuss actions with SE teams. This coordination task requires explainability for the main part. Tseng et al. (2005) state that in the conceptual design work breadth of data is more important than precision, while during the detailed design stage data needs to be precise. Thus, we expect a higher importance of explainability to accuracy when managing costs on low-cost granularity (i.e., complete product level) compared to high cost granularity (i.e., component-level).

H3: *The importance of explainability relative to accuracy is higher in the case of low-cost granularity than for high-cost granularity.*

4.3.3.4 Product Novelty

The technological innovation literature often describes technology novelty by the familiarity with a given technology (Tatikonda & Rosenthal, 2000). Product novelty is the "uniqueness and newness of a product" and the "infrequency and rarity of the product design" (Horn & Salvendy, 2009, p. 228). Henderson and Clark (1990) distinguish between *radical innovation* and *incremental innovation*. A radical innovation establishes a new set of core designs that are assembled in a new architecture. An incremental innovation improves and extends established product designs. Thereby, the individual components are refined, while the core designs of components and the overall architecture remains the same. Incremental innovations adopt existing products and are targeted toward existing markets (Reid & Brentani, 2004). Pullen et al. (2009) state that radical and incremental innovation need different strategies, processes, and organizational factors for product development. Moreover, radical and incremental innovation can be distinguished by the level of market and technology uncertainty (Herstatt et al., 2004). Radical innovations involve high technical uncertainty, resulting from the use of non-existing technologies. Furthermore, radical innovations involve high market uncertainty where the needs and requirements of customers are mostly unclear (Patanakul et al., 2012). Incremental innovations, on the other side, have low technical and market uncertainty. Since we control for information uncertainty as a separate factor in the experiment, we develop the hypothesis independently from the uncertainty aspect. Incremental innovation is characterized by the change of existing products (Abdul Ali et al., 1993; A. Ali, 1994). Thereby, incremental innovations often rely on internal information (Herstatt et al., 2004). Project teams must be familiar with the work done in the past, to successfully establish products with incremental innovation (Patanakul et al., 2012). Moreover, managers use evaluation criteria more thoroughly when making project continuation or termination decisions for incremental projects in contrast to radical ones (Schmidt et al., 2009). Thereby, technical criteria is primarily used to review incremental projects, while financial criteria is mostly used for radical ones (Schmidt et al., 2009). Consequently, we expect that in the case of incremental innovation it is important to be knowledgeable about the technical background of the cost estimates. Accordingly, we expect that development teams are asked to ***conduct in-depth comparisons*** with the predecessor product. When dealing with radical products, explainability is less important since there is ***no direct predecessor*** that can be used as a reference. In conclusion, we expect that the relative importance of explainability to accuracy is higher for incremental innovation than in radical innovation.

H4: *The importance of explainability relative to accuracy is higher in the case of incremental innovations (low novelty) than for radical innovations (high novelty).*

4.3.3.5 Target Cost Gap

In the early product development process, when first cost estimates are being made, the difference between the cost estimate and the cost target becomes clear. We refer to this difference as the target cost gap, and we expect this to matter for the relative importance of explainability and accuracy. Target cost gap is defined by the difference between the cost goal of a product and the planned cost of the current technical solution. The target cost gap thereby refers to how demanding development teams perceive their task of developing a product that meets the cost goals as well as all other product requirements. After realizing the gap between target cost and planned cost, the next step is to establish measures on decreasing or even closing the cost gap. Often, this is a complex task leading to trade-off discussions on product characteristics. Based on the requirements of product features, trade-offs between costs and functionalities are conducted, where non-essential functionalities are sacrificed in order to afford more important components (Thore Olsson et al., 2018). We expect that an increasing cost gap, will demand more explanations and discussions due to the higher ***cost pressure*** on the project.

The development of a product that attains the cost goal and also satisfies the requirements of customers involves a number of cost management methods such as value engineering, functional analysis, quality function deployment, and design for manufacture and assembly (Ax et al., 2008). Value engineering, for example, examines the relationship between the functions and costs of a product and validates alternative product designs (Al-Qady & El-Helbawy, 2016). The effective implementation and application of these methods requires detailed information about the product and costs (Everaert et al., 2006). We expect that an increasing cost gap, will demand more explanations due to the ***need for detailed information*** to operate the cost management methods. In sum, we expect that the importance of explainability is higher, when there is a large overshoot of the estimated costs over the cost goal.

H5: *The importance of explainability relative to accuracy is higher in the case of high target cost gaps than for low target cost gaps.*

Figure 18 provides an overview of the five factors which are expected to determine the relative importance of explainability to accuracy of cost management during new product development. The relative importance between both attributes is expected to determine the selection of machine learning models.

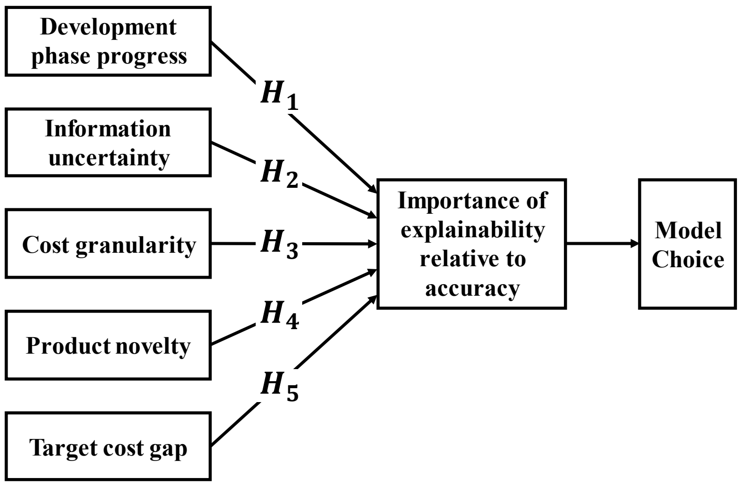


Figure 18: Model overview of the factors and corresponding hypotheses

4.4 Research Method

4.4.1 Experimental Design

The ***factorial within-subject experiment*** employs a 2 (FFE/early development phase) x 2 (low/high information uncertainty x 2 (complete/assembly cost granularity) x 2 (incremental/radical product novelty) x 2 (low/high target cost gap) design. Since the participants of this study come from one organization, *AutomotiveCompany*, the experimental task is tuned to the specific case company context. This makes the experiment realistic and practical in its application.

The importance of explainability to accuracy is ***measured directly*** and ***indirectly***. For the indirect measure, we use several cost estimation models with different levels of explainability and accuracy. In the experiment, the participants first rate multiple models according to their perceived explainability and accuracy score. In a second step, the participants are asked to select a model they considered appropriate for solving a common cost estimation task under different conditions. The experimental conditions are based on the five factors from the prior developed hypotheses. Since each factor is varied over a level of two, a full factorial design would result in a rather impractical amount of $2^5 = 32$ treatment combinations. Considering that high-order interaction effects are less relevant to answer our research question, we apply a ***fractional factorial design*** using half a fraction of the original design. Due to the factorial within-subject design, the manipulation is obvious to the participants. However, the potential change of behavior (so-called carry-over effect) during the experiment is less critical in this experiment and is minimized by randomization. Instead, we use the awareness of the manipulation to obtain additional information during the subsequent discussions. At the end of each experiment, we discuss the manipulated factors with each participant and thereby gather additional information about the impact of each factor. Thus, qualitative data is used to validate the reasoning in our hypotheses and obtain additional arguments. Qualitative data is also gathered to explain any unexpected results.

4.4.2 Sample

Participants were cost experts from three product controlling departments of *AutomotiveCompany*, who are responsible for the management of direct material costs: *complete* vehicle controlling, *assembly* controlling, and *parts* controlling. Complete vehicle controllers are responsible for the overall cost management of full costs from the FFE phase until the end of production. At *AutomotiveCompany* the FFE phase launches around 70 months before the *start of production* (SOP). The assembly controllers are specialized on the management of direct material costs and are responsible for managing the *product design phase*, which starts about 60 months before SOP. The assembly controlling is split into five sub-assembly groups: exterior body, interieur body, electrics, chassis, and engine. Parts controllers are specialized on individual components and are responsible for the management of the *product development phase*, which starts about 45 months before SOP. The *early product development* (EPD) phase covers the implementation of concepts, while the late product development phase encompasses sourcing and homologation. Further, part controllers participate in the SE teams. In these teams, representatives from various business areas jointly develop the parts and components. A tentative and simplified representation of the product development process at *AutomotiveCompany* is showcased in **Figure 19**.

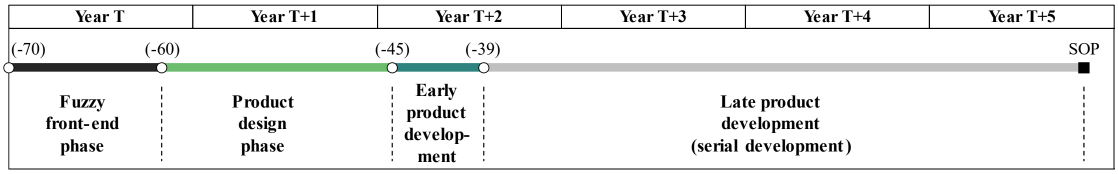


Figure 19: Approximate timeline of the product development process at *AutomotiveCompany*. Months before SOP are in parentheses.

To select participants for the experiment, *stratified convenience sampling* is applied. Since we expect different results over the five sub-assembly groups, an equal quota is utilized to prevent biases. Using convenience sampling is no limitation, since we don't aim at generalizing to a full sample and the assignment to the conditions is not affected by the selection process. In total, 40 product controllers participated in the experiment. To select suitable participants, we only included controllers with a total work experience of at least three years at the three sub departments. For example, a controller that spent two years in part controlling and one year in assembly controlling would be included in the sample. **Table 22** provides an overview of the sampling process. Machine learning users can be divided into three groups: artificial intelligence novices, data experts, and artificial intelligence experts (Mohseni et al., 2018). Artificial intelligence novices have no or only little know-how on artificial intelligence systems. Therefore, artificial intelligence novices will probably have the most difficulties understanding the functionality and results of artificial intelligence systems. The participants of this study can mostly be regarded as members of this group.

Table 22: Participant sampling

1. Sample universe		2. Selection criteria		3. Sample strategy	
Complete	**20**	**Complete**	**17**	**Complete**	**10**
Assembly	**23**	**Assembly**	**17**	**Assembly**	**10**
engine:	6	engine:	4	engine:	2
chassis:	5	chassis:	4	chassis:	2
ext. body:	5	ext. body:	2	ext. body:	2
int. body:	4	int. body:	4	int. body:	2
electrics:	3	electrics:	3	electrics:	2
Parts	**73**	**Parts**	**45**	**Parts**	**20**
engine:	15	engine:	10	engine:	4
chassis:	10	chassis:	7	chassis:	4
ext. body:	17	ext. body:	9	ext. body:	4
int. body:	17	int. body:	12	int. body:	4
electrics:	14	electrics:	7	electrics:	4

4.4.3 Experimental Task and Conditions

In the following, the *experimental task* and the *conditions* are explained. It is important to introduce a task which all participants from the three sub departments are familiar with. In general, the responsibilities and the spectrum of tasks of the three departments are quite different. A common denominator between the three departments is, however, the cost estimation of product features (i.e., head-up display, panoramic roof). At *AutomotiveCompany* the impact on cost for each change in design of a development project is evaluated and reported to project management. Accordingly, the participants are provided with the following task: *The product planning team requests a change of product design for a car model during development. Your objective is to predict the updated costs for the new design.* This task is repeated several times under various conditions.

The manipulations of the independent variables are summarized in **Table 23**. The treatment levels and descriptions were discussed with the head of the part controlling department of *AutomotiveCompany* beforehand, to ensure that the manipulation was sufficient.

Table 23: Coding of the manipulations of independent variables

Factor	Code	Level	Description
Development phase progress	-1	FFE	The task is conducted 70 months before SOP.
	+1	EPD	The task is conducted 45 months before SOP.
Information uncertainty	-1	low	There is much certainty about the underlying technology, costumer, and competitor environment. The properties and requirements of the product are determined and approved.
	+1	high	There is little certainty about the underlying technology, costumer, and competitor environment. The properties and requirements of the product are vague and not yet approved.
Cost granularity	-1	complete	Your objective is the estimation of complete vehicle costs.
	+1	assembly	Your objective is the estimation of costs on assembly level[5].
Product novelty	-1	increm.	Your objective is to estimate the cost of a successor car with a direct predecessor.
	+1	radical	Your objective is to estimate the cost of a new car without any predecessor.
Target cost gap	-1	low	The estimated costs of the current design surpass the target costs by 5%.
	+1	high	The estimated costs of the current design surpass the target costs by 20%.

The conditions are based on the fractional factorial design with half a fraction, which results in $2^{5-1} = 16$ treatment combinations. The design provides a V-resolution where no main effect is confounded by three-factor interactions (or higher). Further, two-factor interaction effects are unconfounded by other two-factor interactions. The experimental design matrix is depicted in **Table 24**.

[5] The assembly cost granularity assigns the product costs to four groups of components: engine, chassis, body, and electrics.

Table 24: Experimental design matrix: Treatment combinations (conditions) of the 2_V^{5-1} fractional factorial design

Conditions	Development phase progress	Information uncertainty	Cost granularity	Product novelty	Target cost gap
1	-1	-1	-1	-1	+1
2	+1	-1	-1	-1	-1
3	-1	+1	-1	-1	-1
4	+1	+1	-1	-1	+1
5	-1	-1	+1	-1	-1
6	+1	-1	+1	-1	+1
7	-1	+1	+1	-1	+1
8	+1	+1	+1	-1	-1
9	-1	-1	-1	+1	-1
10	+1	-1	-1	+1	+1
11	-1	+1	-1	+1	+1
12	+1	+1	-1	+1	-1
13	-1	-1	+1	+1	+1
14	+1	-1	+1	+1	-1
15	-1	+1	+1	+1	-1
16	+1	+1	+1	+1	+1

4.4.4 Dependent Variable Measurement

The dependent variable of this research is the relative importance of explainability to accuracy (*EtA*), which is measured directly with a Likert scale (S) and indirectly with the aid of three cost estimation models (M). For the direct measurement, we ask participants about the perceived relative importance of explainability to accuracy with a 10-point Likert scale ($\boldsymbol{EtA^S}$). The bipolar Likert scale ranges from 10 (explainability is most important) to 1 (accuracy is most important). For the indirect measurement ($\boldsymbol{EtA^M}$), participants first rate three machine learning models with different levels of complexity according to their perceived explainability and accuracy. Then, each time they consider the cost estimation task, the participants were asked to choose between these models to solve the experimental task. The relative importance of explainability to accuracy in any given condition can then be derived from the selected model and the rating of that particular model's perceived explainability and accuracy. The model selection is based on the *intention to use* construct of Agarwal and Prasad (1998). Venkatesh et al. (2003) introduces three determinants for intention to use: the degree to which the system improves job performance, the degree of ease-of-use, and the extent of social implications by using the system. The applied explainability evaluation can be considered as *application-grounded*, as domain experts and a real application were involved (Doshi-Velez & Kim, 2017)

$$EtA^S = \quad \textit{Explainability to Accuracy based on Likert scale} \tag{22}$$

$$EtA^M = \quad \textit{Explainability to Accuracy based on model choice} \tag{23}$$

An OLS regression analysis is conducted for both *EtA* measures to test the main effects of the five factors on the relative importance of explainability to accuracy. In the regression analysis, we investigate 640 observations (40 participants x 16 treatment combinations). We standardize both variable measures for each participant, since the mean explainability to accuracy score and their variance over the conditions is highly subjective.

In the following sections, we further describe the indirect approach for measuring explainability to accuracy. First, we describe the evaluation of models according to the perceived explainability and accuracy. Then, we select three models with high expected variation according to both attributes. Following this, we describe each model and explain the model implementation. At the end, we describe the archival costing and technical data that is used to train and test the models.

4.4.4.1 Model Evaluation According to Accuracy and Explainability

Explainability is the combination of transparency and interpretability. Both sub-attributes are evaluated individually and combined by average. The ***transparency measure*** is based on the items measures by Gedikli et al. (2014) and Berkovsky et al. (2017). Gedikli et al. (2014) aimed to compare different explanations in recommender systems. In their questionnaire they ask to rate “to which extent the different explanation interfaces were suited to increase the transparency of the system” (Gedikli et al., 2014, p. 379). The question was followed by a short definition of transparency: “Transparency means that the explanation interface helps you to understand how the recommendation system works” (Gedikli et al., 2014, p. 379-380). The construct was measured on a seven-point scale from *not at all* to *very much*. The work of Berkovsky et al. (2017) investigates trust as a major factor for the success of recommender systems. User-trust is compared with the constructs of competence, transparency, intention to re-use, and overall model trust. The transparency construct is phrased as “I understand the best the [*sic*] reasons for the suggestions provided ...” (Berkovsky et al., 2017, p. 291). We measure transparency on three seven-point scale items ranging from *I strongly disagree* to *I strongly agree*.

- **T_1**: “I can fully understand how the cost estimation model works.”
- **T_2**: “I can understand the best reasons for the results provided by the model.”

- **T3**: "I understand how the model comes to its solution."

The ***interpretability construct*** is based on the work of Piltaver et al. (2016) and Demoulin and Coussement (2020). In the survey of Piltaver et al. (2016), task users are requested to give subjective opinions on several classification trees. Thereby, respondents rate how comprehensible several variations of decision trees are. An easy-to-comprehend classification tree is described as follows: "I can use the knowledge represented by the classification tree as soon as I read it for the first time; I can easily remember it; I can explain it to another person without looking at the figure." (Piltaver et al., 2016, p. 337). To analyze factors that determine the usage of text mining tools, Demoulin and Coussement (2020) measure interpretability by three items: "It would be easy to interpret what text-mining outputs mean. Text-mining outputs would be easily interpretable. The measurement units for text-mining outputs would be clear." (Demoulin & Coussement, 2020, p. 7). We measure interpretability on two seven-point scale items ranging from *I strongly disagree* to *I strongly agree*.

- **I1**: "The explanation of results is obvious to me."
- **I2**: "I can explain the results to another person."

In the context of cost estimation, ***accuracy*** measures the ability of a method or model to predict the actual cost for a project or activity. In this study, the construct of accuracy is based on the definition of Dysert (2007) describing accuracy as the degree to which an estimate differs from the actual values. Thus, cost estimation accuracy indicates the degree to which the estimated cost may vary from the final cost for a project. We measure accuracy on a seven-point scale ranging from *not accurate* to *very accurate*.

- **A**: Please rate the accuracy of the following models. Accuracy of a cost estimate is defined as the deviation of the forecasted cost from the actual cost of a project.

4.4.4.2 Selectable Cost Estimation Models to the Participants

We select three models with different expected levels of explainability and accuracy: CBR, MLR, and GBR. In the following, we evaluate each model individually according to both attributes and subsequently provide short descriptions of the machine learning methods.

For the purpose of explainability, **CBR** shows its core strength. Neither complex algorithms are deployed, nor complex associations between dependent and independent variables are assumed. The CBR method functions in a highly transparent manner where the source of the estimated cost and the underlying solution is known at any time (Duverlie & Castelain, 1999). However, the effectivity of CBR highly depends on the similarity toward past cases. In case of high innovations, the predictive performance might be limited due to the lack of sufficiently similar cases from the past (Roy, 2003). Further, the method is limited by the ability to inter-/extrapolate cost and consider interdependencies between product features. In sum, we expect *high* explainability but *low* accuracy for the CBR approach in the context of product cost estimation.

Due to the linear design and the non-existing interdependencies between features, the results of **MLR** are easily interpretable based on the regression coefficients (James et al., 2013). With the regression model, one can easily comprehend the cost behavior of the model. Thereby, the prediction is credible on a term-by-term basis (Smith & Mason, 1997). The cost prediction performance of MLR methods is robust, yet often lower than more complex state-of-the-art machine learning models, failing to capture nonlinearities and interdependencies between input variables (Loyer et al., 2016). We expect *medium* explainability and *medium* accuracy for the MLR model.

While individual regression trees are easy to interpret visually, it is not the case for a combination of trees such as **GBR**. Due to the difficulty in obtaining explanations from complex ensembles, gradient boosting is often referred to as a black box (Hatwell et al., 2021). Yet, GBR is often the most accurate model for estimating product costs compared to other models (Loyer et al., 2016; Shin, 2015). Therefore, the explainability is expected to be *low*, but the accuracy to be *high*. The expected explainability and accuracy scores of the three models are summarized in **Table 25**. In the following, we provide the main characteristics and describe the implementation of the three models.

Table 25: Selection of machine learning models and the expected levels of explainability and accuracy

Model	Expected accuracy	Expected explainability
CBR	*low*	*high*
MLR	*medium*	*medium*
GBR	*high*	*low*

4.4.4.3 Case-Based Reasoning (CBR)

CBR systems are inspired by the processes of remembering in human reasoning (Chen & Burrell, 2001). The method attempts to solve a problem by remembering and comparing it to similar cases from the past. The system then adopts the solution gained from prior cases to solve a new case. The CBR cycle consists of four sub-phases (Hu et al., 2016):

(1) To solve a new case/problem, the system retrieves the most similar case from a set of previous cases (case base) according to a similarity measure (*retrieve*)
(2) In case of sufficient similarity, the new case is solved by adapting the solution of the retrieved case (*reuse*)
(3) In case of insufficient similarity, the solution of the retrieved case is revised. (*revise*)
(4) The confirmed solution is retained and stored in the case base (*retention*)

In the context of product cost estimation, the actual cost of past products is used as prior cases. Each instance in the case base is described by k input variables ($x_1,...,x_k$) and a cost label y. To predict the cost for a new case, the historical cost of the most similar case is adopted. First, the similarity scores for all cases in the case base are calculated. Then, the case with the highest overlap is selected. In case of sufficient similarity, the cost of the selected case is used to predict the cost of the new case. The similarity between a new case N and an old case S can be calculated as the weighted sum of similarities of each input variable $f(N_i,S_i)$. The similarity f for the i-th variable can be calculated by the Manhattan or Euclidean distance.

$$Similarity(N,S) = \frac{\sum_{i=1}^{k} w_i f(N_i,S_i)}{\sum_{i=1}^{k} w_i} \tag{24}$$

4.4.4.4 Multiple Linear Regression (MLR)

A common approach to estimate the relationship between several independent input variables ($x_1,...,x_k$) and a continuous dependent variable y is the MLR model. The cost estimate y can be expressed as a linear combination of the input features x in addition to a stochastic error term ϵ.

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \epsilon \tag{25}$$

The regression coefficients β are obtained by the least-squares-method. By minimizing the sum of squared prediction errors, an optimal solution for the parameter vector β can be derived. Thereby, the coefficients express the impact of the input factors on the product cost. One assumption of regression models is the independence of independent variables, whereas otherwise, the cost estimates can be inaccurate (S.-G. Wang et al., 1990). To account for this effect, we apply the *Least Absolute Selection and Shrinkage Operator*, which allows for parameter shrinkage during regression and automatically conducts variable selection (Chan et al., 2018). Less important variables for estimating the cost are automatically discarded and multicollinearity can be reduced.

4.4.4.5 Gradient Boosted Regression (GBR)

Boosting is an important approach in machine learning, which was developed by Freund and Schapire (1997). The approach involves the combination of many simple models (so-called *base learners*) to produce a single powerful ensemble model. The ensemble $F(x)$ is calculated as the weighted sum of multiple base learners $f_m(x)$. The expansion coefficient at each iteration is denoted as β_m.

$$F(x) = \sum_{m=1}^{M} \beta_m f_m(x) \tag{26}$$

Thus, an additive model is created by sequentially fitting base learners to the current residuals at each iteration. This approach has been refined by Friedman (2001) and Friedman (2002) by introducing the *TreeBoost* algorithm using regression trees as base learners. Thereby, it combines the advantages of the boosting approach and regression trees, which are computational efficiency and conceptual simplicity (Shin, 2015). In each iteration, the residuals between the current ensemble and actual values are used to train an additional decision tree. To update the ensemble, the predicted values of the additional decision tree are added to the corresponding results of the current ensemble. Further, a learning rate can be deployed to increase the generalizability of the ensemble model.

4.4.4.6 Implementation of Models

The three models were trained on archival costing data from *AutomotiveCompany*. Four car models over two generations were used to train and test the supervised machine learning models. Each model was trained on data from the predecessor generation (in total 135,725 observations of different car configurations) and tested on the subsequent generation (in total 129,074 observations). The models were used to estimate the direct material cost of passenger cars. The costs can be expressed at two levels: complete vehicle level and assembly level (body, electrics, chassis, and engine). The costing data is described by multiple product characteristics. Most characteristics are categorical (i.e., polyurethane, leather, Alcantara leather coating for steering wheels) or on a binary scale (i.e., head-up display available or not available). The categorical variables are one-hot encoded. 10 product characteristics are cardinal (i.e., engine performance). Overall, the costing data is labeled with 461 product features. For a more detailed description of the data set, see chapter 3.4.3.1. Hyperparameters were tuned on a small subset of the training data from the predecessor generation. A detailed description of the implementation of models can be found in **Table 19** in Appendix B. The selected models were implemented in Python with the *scikit-learn* package from Pedregosa et al. (2011).

4.4.5 Procedure

The experiment consists of four phases: introduction, model rating, evaluation of conditions, and discussion. The four phases are accompanied by a questionnaire (Appendix C). In the ***introduction phase***, we provide background information on machine learning. Further, we gather supplementary data about the participants, such as prior experience in product controlling and machine learning, as they might be influencing factors on the relative importance of explainability to accuracy.

In the ***model-rating phase***, the models are rated according to their perceived transparency, interpretability, and accuracy. First, the participants are provided with standardized descriptions for each model. As an example, the description of the CBR method is depicted in **Figure 20**. The description of MLR and GBR can be found in Appendix C (**Figure 28** and **Figure 29**). Subsequently, the participants are asked to rate the perceived transparency according to the items T_1, T_2, and T_3.

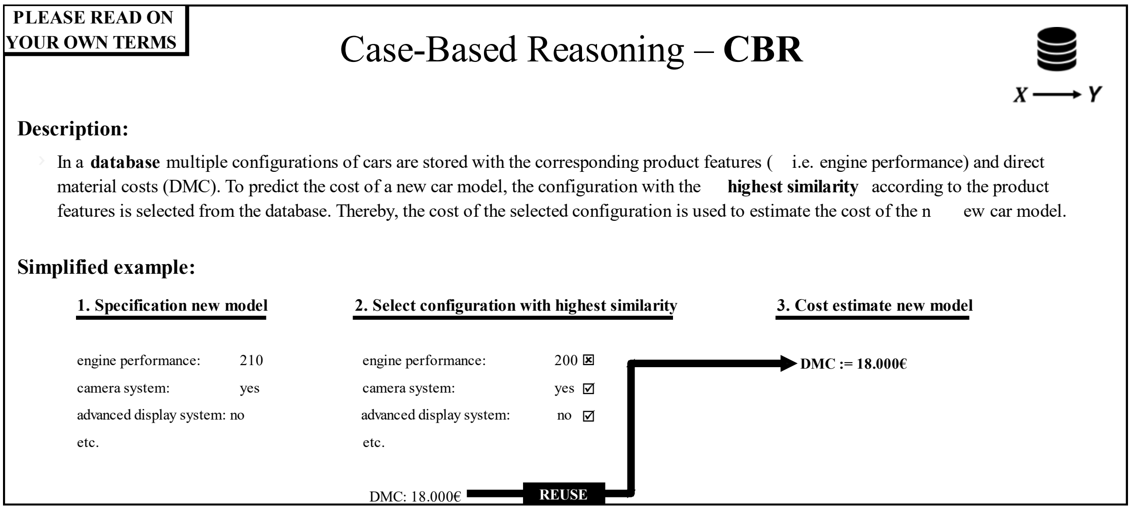


Figure 20: Description of the CBR method

To evaluate the interpretability of the models, participants are further familiarized with the models by watching short video clips from a *demonstration tool*. For each machine learning model, the demonstration tool estimates the direct material cost based on a common set of product features. Thereby, each model provides individual explanations for the estimated costs. We created short video clips about the usage of the demonstration tool to ensure comparability among participants. In each video clip, the machine learning models carry out the same procedure: First, the entry form of product features is presented to the participants (**Figure 30** in Appendix C) and the cost of a given product configuration is calculated. Second, two features are changed, and the new costs are calculated. Third, the specific explanations for the updated costs are showcased. The length of each video clip amounts to approximately 1 minute. The explanation of the CBR model is threefold. In the first step, the identification key of the most similar case is presented. In the second step, a list of differences between the target feature values and the feature values of the retrieved case is provided. In the third step, the bill of materials of the retrieved case is showcased, to obtain costs on part level and additional information such as part descriptions and sourcing information. The demonstration tool of the CBR model is displayed in **Figure 21**. The explanation of the MLR model depicts the regression formula of the linear model. Therewith, the cost for each feature can be retrieved from the corresponding regression coefficients (**Figure 31** in Appendix C). The

GBR model is intentionally presented as a black box (**Figure 32** in the Appendix C), where neither calculation steps nor explanations are provided. In this case, the participants are provided only with the estimated costs. After presenting the three video clips, the interpretability of each model is evaluated according to the items I_1 and I_2.

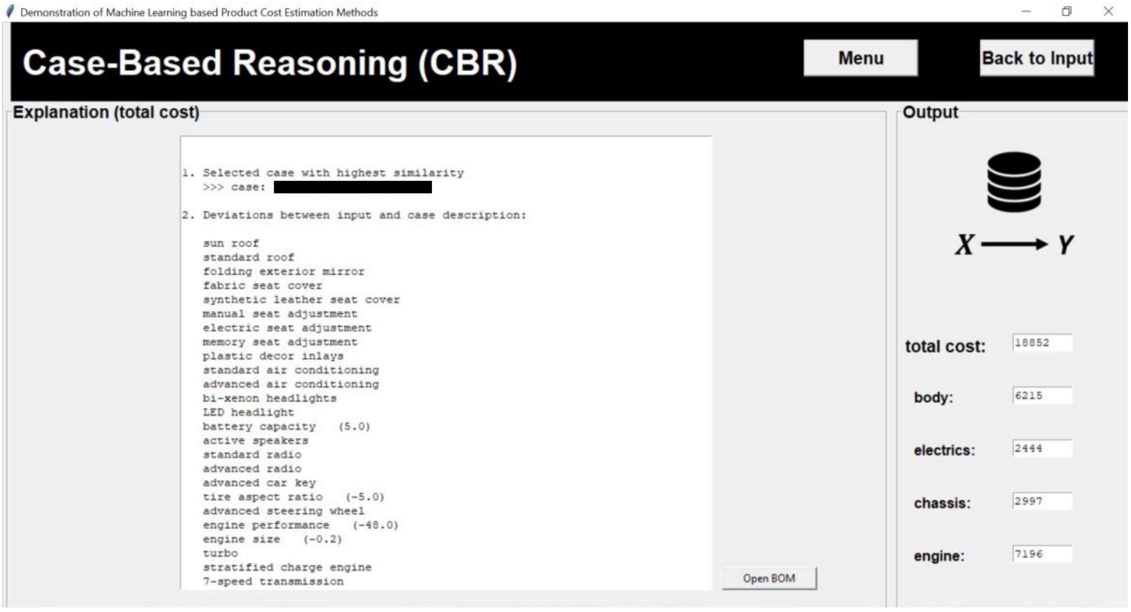


Figure 21: Demo Tool: CBR

To rate the accuracy of models, the participants are provided with the actual results for the inter-generational testing set for total- and assembly-level direct material cost prediction. To evaluate the predictive performance, the three models are compared according to the NMAE metric, where lower values are preferable. The NMAE is defined by $\frac{MAE}{\frac{1}{n}\sum_{i=1}^{n} y_i}$, where MAE denotes the mean absolute error and y_i are the actual costs. The overall total cost estimation performance is calculated as the average NMAE over the four car models. In the case of total vehicle cost estimation, the GBR model performed best (NMAE of 4.86%). The NMAE of the MLR model amounts to 7.86%. The CBR model corresponds to the highest error (10.16%). The same order applies to the accuracy scores on assembly level. In the case of assembly-level cost estimation, the mean NMAE is calculated by the average NMAE of the four assembly groups (body, electrics, chassis, and engine) over three car models. The multi-target prediction is conducted for *car b*, *car c*, and *car d*. *Car a* is discarded, since no assembly-level costing data was available. The most accurate prediction is delivered by the GBR model (NMAE of 13.98%), second by MLR (15.09%) and third by the CBR model (18.75%). **Table 26** summarizes the actual cost prediction accuracy of the three models. The perceived accuracy measure is based on item A.

Table 26: Prediction of the total- and assembly-level cost for the subsequent product generation. The predictive accuracy is measured by the NMAE metric. Standard deviation over cars and assembly groups are provided in parenthesis.

	NMAE [%]	
Model	**Total cost**	**Assembly cost**
CBR	10.16 (3,68)	18.75 (8.96)
MLR	7.86 (3.44)	15.09 (9.41)
GBR	4.86 (1.63)	13.98 (6.60)

In the ***evaluation phase***, the participants evaluate the perceived relative importance of explainability to accuracy for each experimental condition. First, the direct measurement on the 10-point Likert scale is applied ($\boldsymbol{EtA^S}$). Second, participants are asked to select a model to solve the experimental task. Based on the model choice in each condition, the corresponding explainability and accuracy score can be derived from the prior rated scores of the models ($\boldsymbol{EtA^M}$).

In the ***discussion phase***, each participant is asked about the effect of each factor on the importance of explainability and accuracy. For each situation of a given factor (i.e., FFE and EPD phase of development phase progress) we ask whether explainability or accuracy would be more important. We are also interested in the reasons, examples, and notions behind those decisions. The gathered qualitative data serves three purposes. First, we use the data to validate the arguments from our hypotheses. Second, the qualitative data enables us to identify additional arguments behind the expected results not included in our hypotheses. Third, in the case of insignificant or unexpected results we can find out why a variable may not be effective or results in a different outcome.

Throughout the experiment, the treatment combinations are randomly ordered for each participant. To prevent confusion during the presentation and evaluation of models, the rating of transparency, interpretability, and accuracy are not ordered randomly. We use a fixed order: CBR, MLR, and GBR. The experiments are conducted by screen shared audio conferences. As the participants are German native speakers, all questions and instruction are translated into German. The mean duration of the experiments amounts to 60.92 min (SD 8.51). The average work experience amounts to 10.57 years (SD 6.80). The average machine learning experience amounts to 2.00 (SD 1.13) on a 7-point Likert scale. Accordingly, most participants can be regarded as machine learning novices (Mohseni et al., 2018).

4.5 Validity and Reliability of Variable Measures

In the following, the validity and reliability of the variable measures and assumptions are analyzed. Specifically, we analyze six underlying assumptions and aspects of the interpretability problem. First, the three models are compared according to their predictive performance. We expect that the three models have an increasing level of accuracy (CBR < MLR < GBR), see **Table 25**. For total- and assembly-level cost prediction, the order of accuracy is in line with the expected order, see **Table 26**. Consequently, we can confirm the assumption of increasing predictive accuracy over the three models.

Second, we analyze the perceived explainability scores from the perspective of product controllers. We expect a decreasing level of explainability over the three machine learning models (CBR > MLR > GBR), see **Table 25**. The average scores of perceived explainability for the three models are summarized in **Table 27**. The mean score of each model is calculated as the mean over all participants. The transparency measure is based on the average of the three items T_1, T_2, and T_3. The interpretability measure is based on the average of the two items I_1 and I_2. Explainability is calculated as the average of transparency and interpretability scores for each participant individually. The *Cronbach's alpha* of the three items of transparency amounts to 0.964. The *Cronbach's alpha* of the two items of interpretability amounts to 0.957. Due to the two-item measurement, we additionally calculated the *Spearman–Brown coefficient*, which equally amounts to 0.957. The statistics are based on 120 observations (40 participants x 3 models). Accordingly, we observe excellent ***internal consistency (reliability)*** for the transparency and interpretability measures. On the other side, the high alpha values indicate redundancy of items. Redundancy makes the test instrument less efficient since little additional information is obtained by the extra items (Cronbach, 1951). The acceptable range of alpha values varies across scholars. Some authors consider alpha values up to 0.96 to be sufficient, while other regard maximum alpha values of 0.98 as acceptable (Taber, 2018). However, the results indicate that a combination of several constructs might not be necessary when measuring interpretability and transparency. Transparency and interpretability are highly positively correlated. The correlation coefficient amounts to 0.84 over the 120 observations. The CBR model achieves the highest scores for explainability, second highest is MLR and least explainable is the GBR model. These results are consistent with the expected order of explainability over the three models. Further, high ***content validity*** of the explainability measure is achieved since both sub-attributes (transparency and interpretability) are introduced and evaluated separately. Furthermore, there is high consistency in the assessment of explainability across participants: 29 participants agreed on the explainability order of CBR > MLR > GBR.[6] The consistency of the assessment over several raters indicates high ***inter-rater reliability***. Overall, we can confirm the expected order of explainability of the three machine learning models.

Table 27: Comparison of the perceived scores of accuracy and explainability for the three machine learning models. The table depicts the mean and standard deviation over all participants. The values are based on the items T_1, T_2 and T_3 (transparency), I_1 and I_2 (interpretability) and A (accuracy).

Model	Accuracy	Explainability	Transparency	Interpretability
CBR	3.25 (1.26)	6.69 (0.38)	6.78 (0.45)	6.60 (0.55)
MLR	4.58 (1.01)	6.08 (0.79)	6.29 (0.77)	5.86 (1.06)
GBR	5.85 (0.95)	2.67 (0.95)	3.46 (1.41)	1.88 (0.96)

Third, we analyze the assumption about the inverse relationship between accuracy and explainability of the three models. Based on the interpretability problem, we expect a negative correlation between both attributes. The relationship between the evaluated accuracy and explainability scores of the three models is depicted in **Figure 22**. The perceived scores yield a significant negative correlation between accuracy and explainability over the three models. The correlation coefficient over the 120 observations (40 participants x 3 models) amounts to $r = -0.586$, p-value < 0.001. Moreover, the association between explainability and accuracy follows the concave transformation function depicted by Barredo Arrieta et al. (2020) (see **Figure 15**). Therefore, we can confirm that some models perform primarily on accuracy, while others perform mainly on explainability and that this trade-off is actually perceived by people. The consistency toward the interpretability problem ensures ***construct validity***.

[6] 37 participants agreed on the order CBR ≥ MLR > GBR

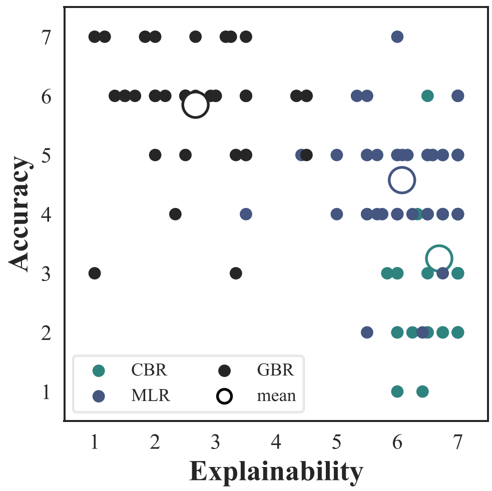


Figure 22: Perceived explainability and accuracy over the three models: gradient boosted regression (GBR), multiple linear regression (MLR), and case-based reasoning (CBR). The scatter plot suggests an inverse relationship between perceived explainability and accuracy.

Fourth, we expect that the relative importance of explainability to accuracy is important when choosing between alternative machine learning models in different situations during new product development. The frequency distributions of the selected models over the perceived importance of explainability to accuracy for each participant and treatment condition (40 x 16 observations) is depicted in **Figure 23**. The three distributions show that the model selection strongly depends on the perceived relative importance of explainability to accuracy. In situations primarily requiring explainability, CBR is selected; whereas the GBR model is used in situations which demand mainly accuracy. This is in line with the expected association between the perceived importance of explainability to accuracy and the selection of machine learning models. To further analyze the strength of the impact of the relative importance of explainability to accuracy for model selection, we used *analysis of variance* (ANOVA). The ANOVA yields significant differences, F-statistic = 608.72, p-value < 0.001, among the means of perceived importance of explainability to accuracy (EtA^S) over the three models. We can confirm that the perceived importance of explainability to accuracy plays a significant role when cost experts select between alternative cost estimation models.

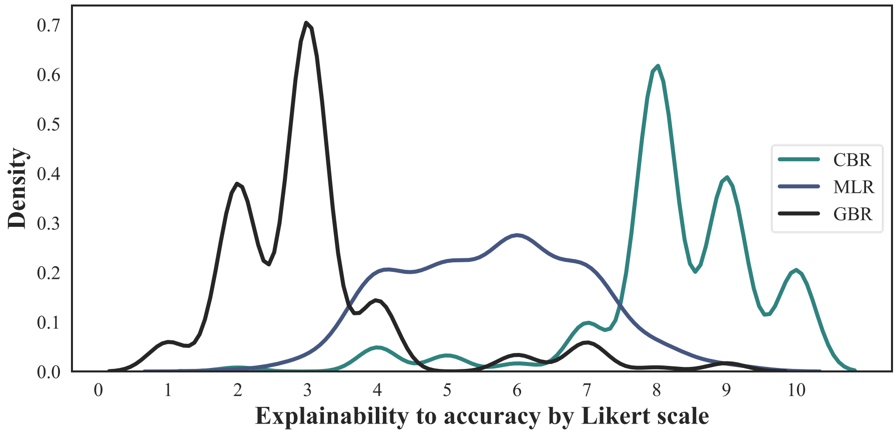


Figure 23: Model selection over perceived importance of explainability to accuracy in different situations. The gradient boosted regression (GBR) model is mostly selected in situations with low levels of perceived *explainability to accuracy* (accuracy is more important), whereas the case-based reasoning (CBR) model is mainly chosen in situations with high levels of perceived *explainability to accuracy* (explainability is more important).

Fifth, we expect a high correlation between the direct and indirect measurement since both proxy the perceived importance of explainability and accuracy in a given situation. The correlation coefficient between both standardized measures over the 640 observations (40 participants x 16 conditions) amounts to $r = 0.819$, p-value < 0.001. Both variable measures are standardized for each participant individually. **Figure 24** depicts the scatter plot between both standardized measures and the linear regression line. Accordingly, we observe ***strong internal consistency (reliability)*** between both dependent variable measurements.

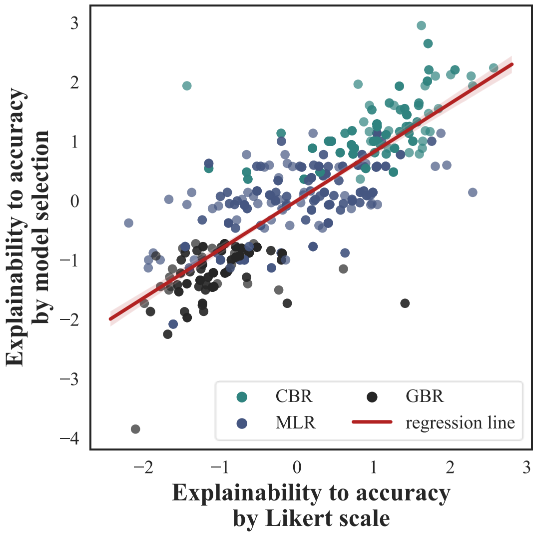


Figure 24: Scatter plot between the direct and indirect measurement of the relative importance between explainability and accuracy. Both measures are standardized for each participant individually. The scatter plot and the regression line indicate a strong linear relationship between both variable measures.

Sixth, we analyze the variation of the relative importance of explainability to accuracy for different situations during new product development. We expect that the relative importance of explainability to accuracy depends on various conditions and task characteristics. To analyze this assumption, we depict boxplots for both *EtA* scores over the 16 conditions (**Figure 25**), which suggest strong variation of *EtA* scores. To statistically test the differences, ANOVA is conducted. The ANOVA yield significant differences among the means over the 16 conditions (F-statistic = 6.994 for EtA^S and F-statistic = 5.266 for EtA^M). For both *EtA* measures the p-value is less than 0.001. Therefore, we can confirm the expected variation of explainability to accuracy over different situations during new product development.

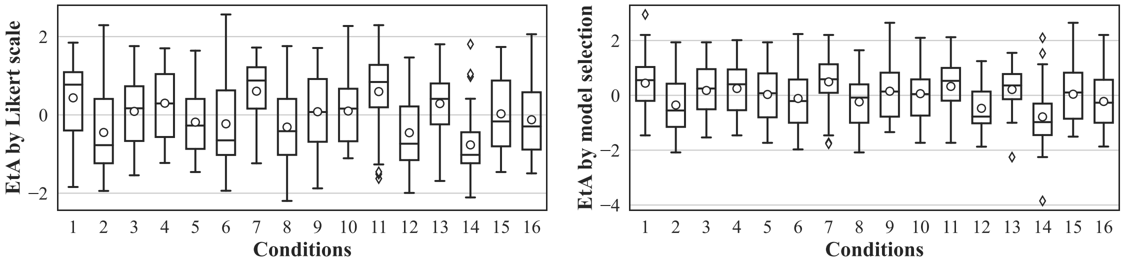


Figure 25: Boxplots of the importance of explainability to accuracy scores (*EtA*) over the experimental conditions. Mean *EtA* scores are indicated by circles. For both variable measures, by Likert scale and by model selection, the boxplots suggest different importances of explainability relative to accuracy (different means and medians) over the 16 treatment combinations.

4.6 Results

In the following, the effect of the five factors on the relative importance of explainability to accuracy in the context of cost estimation is analyzed. Since we use an OLS regression analysis to test the main effects of the five factors, we first evaluate the assumptions of linear regression with binary independent variables. First, we checked the assumption of normality with the aid of quantile-quantile-plots where we find no abnormalities (**Figure 33** in Appendix C). Additionally, we conducted a Kolmogorov-Smirnov normality test on the residuals. For both *EtA* measures the normality assumption can be confirmed (p-value = 0.138 for EtA^S and p-value = 0.222 for EtA^M). Second, we used residual plots to check for homoscedasticity (**Figure 34** in Appendix C). The residual plots exhibit equal variance for both dependent variables. Third, we checked the homogeneity of variance between treatments. In doing so, we conducted a Levene's test over the 16 treatment combinations. We can confirm equal variances between the groups for both *EtA* measures (p-value = 0.427 for EtA^S and p-value = 0.740 for EtA^M).

Table 28: Main effects of the five factors on the relative importance of explainability to accuracy. The importance of explainability to accuracy is measured by scale (*EtA^S*) and model selection (*EtA^M*). Both *EtA* measures are standardized for each participant individually.

	EtA^S	*EtA^M*
Constant	0 (0)	0 (0)
Development phase progress	**-0.2421***** (-6.559)	**-0.2345***** (-6.232)
Information uncertainty	**0.0906*** (2.454)	0.0441 (1.172)
Cost granularity	**-0.0867*** (-2.348)	**-0.0729**$^+$ (-1.937)
Product novelty	-0.0320 (-0.866)	**-0.0862*** (-2.291)
Target cost gap	**0.2462***** (6.670)	**0.1809***** (4.807)
Observations	640	640
R^2	13.6%	10.2%

$^+p<0.1$; $^*p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

The results of the regression analysis are summarized in **Table 28**. Supporting **H1,** we observe a significant negative correlation between *development phase progress* and both *EtA* measures. During the EPD phase of new product development (high progress), more importance is placed on accuracy relative to explainability. In the FFE stage (low progress), explainability is more important. As hypothesized in **H2**, *information uncertainty* is positively correlated with the relative importance of explainability to accuracy. The factor equally produces a positive effect on both *EtA* measures. However, the effect is only significant for the direct *EtA* measurement. We address the differences between the measures in the discussion section (Section 4.7). Supporting **H3**, the *cost granularity* is negatively correlated with the relative importance of explainability to accuracy. The higher the cost granularity (i.e., assembly level cost), the more importance is placed on accuracy relative to explainability. The effect is significant for *EtA^S* and marginally significant and *EtA^M*. As expected in **H4**, *product novelty* is negatively correlated with the relative importance of explainability to accuracy for both *EtA* measures. However, the result is only significant for the indirect *EtA* measurement. Finally, we can confirm the positive association of *target cost gap* and the relative importance of explainability to accuracy (**H5**). The main effects of the five factors on the direct measurement are depicted in **Figure 26**.

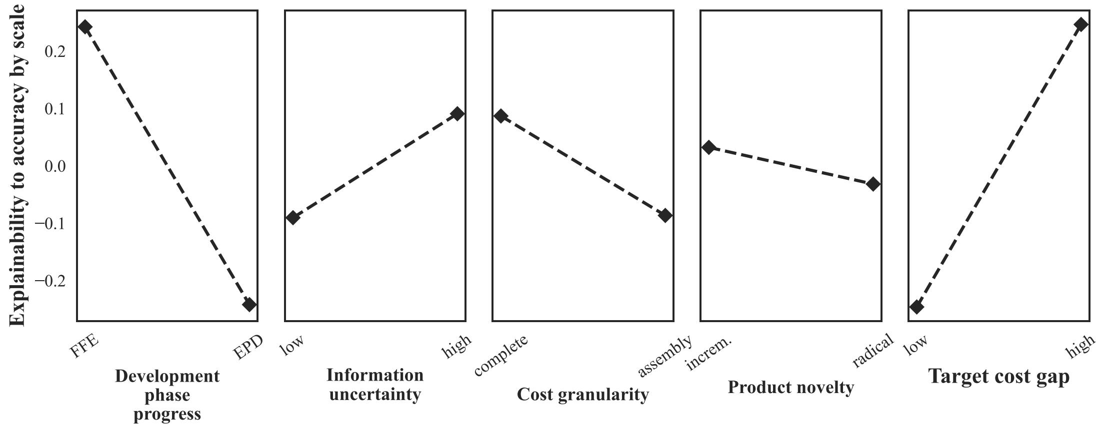


Figure 26: Main effects of the five factors on explainability to accuracy by scale. The figure illustrates that explainability is more important than accuracy (higher EtA^S value) for the *FFE* development phase, *high* information uncertainty, *complete* cost granularity, *incremental* product novelty (slightly), and *high* target cost gaps.

The mean EtA^S over the treatment combinations and participants amounts to 5.68 (SD 2.32) on a 10-point Likert scale indicating that in average explainability is slightly more important than accuracy. Accordingly, participants tend to select simpler cost estimation methods in most situations. Despite being much more inaccurate, cost experts mostly chose either the CBR or the MLR model in almost 74% of the cases (**Figure 27**). In only 26% of the cases the more complex GBR model was considered to be adequate.

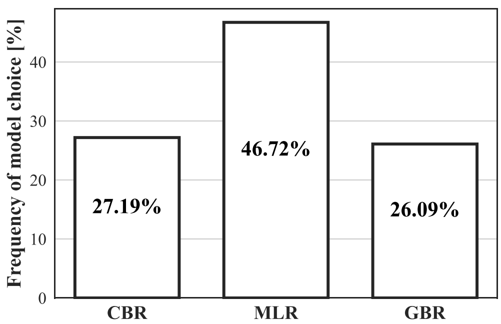


Figure 27: Frequency of models selected by the participants to solve the cost estimation task in several conditions (40 x 16 observations)

We further tested for interaction effects between the independent variables. However, no significant second-order effects (or higher) could be identified. The five factors explain 13.6% of the variance of EtA^S and 10.2% of EtA^M. We assume that the explained variance is limited by the difficulty of processing five factors with different levels. Research has shown that the capacity of processing information is limited to approximately 7 cognitive entities at once (Miller, 1956). We expect that the scenario evaluations with 10 entities (5 factors x 2 situations) are prone to errors due to intuition and spontaneity. The high complexity can lead to evaluations that might contradict prior assessments and therefore reduce the explained variance of the dependent variables. Moreover, there was no significant difference in the mean *EtA* scores between the three controlling departments (complete, assembly, and parts). Further, we find no significant associations between work experience or machine learning experience from participants and the average explainability to accuracy ratings.

4.7 Discussion

For the discussion of results, we will draw on the qualitative information gathered during the experiment. We asked participants to articulate some of the reasons why one attribute, explainability or accuracy, would be more important in both situations of a given factor. We also asked follow-up questions to make sure that we understand the arguments correctly and their context in which they are effective.

We conducted a content analysis, which combines a deductive and inductive approach. During the deductive phase, we assigned the participants' statements to the arguments from our hypotheses. We thereby compare the explanations provided by the participants with the arguments from our hypotheses. Some of the participants' reasons reiterate our theoretical arguments, which are collected in Panel A of **Table 29-33**. In the inductive phase, we obtain new arguments from the remaining participants' statements. Some reasons mentioned by the participants suggest other arguments supporting our hypotheses (Panel B). We further develop these arguments, drawing on new literature. Furthermore, reasons mentioned by the participants might help to understand unexpected results (Panel C). We develop possible theoretical explanations for the results, drawing on the participants comments and new literature.

4.7.1 Development Phase Progress

The results in the experiment supported H1: We found a negative relationship between *development phase progress* for both *EtA* measures. The importance of explainability relative to accuracy is higher in the FFE phase than in the EPD phase of new product development. The arguments we developed for H1 are summarized in **Table**

29, together with several quotes from the experiment. These provide support for the theoretical arguments and provide more confidence about the validity of the results of the experiment. The participants' statements did not suggest other arguments for H1.

Table 29: Evaluation of the factor development phase progress (FFE or EPD)

Panel A: Quotes supporting original arguments for the hypothesis *(deductive)*	
1) FFE: Explainability is more important due to the focus on **conceptual work** (*doing the right things*).	*"Explainability is important while accuracy is less important, since the project framework needs to be developed first." (Parts controller 14)* *"The FFE phase is about product ideas where I decide based on rough estimates." (Complete vehicle controller 24)* *"In the FFE phase, I have big levers [for cost improvement], which do not need to be calculated in detail." (Complete vehicle controller 15)*
2) EPD: Accuracy is more important due to the focus on realization and **implementation work** (*doing the things right*).	*"In the EPD phase I need accuracy. I need to demonstrate profitability, since I substantiate and materialize the concepts." (Complete vehicle controller 24)* *"At the latest at the stage of pricing a feature it is extremely important to know the exact costs, otherwise there is the danger of producing at a loss." (Assembly controller 22)*
3) FFE: Explainability is more important since more **assumptions** are required (*less decisions are made*).	*"Here in the FFE phase it is important to be clear about the assumptions and why you come to certain decisions. At later stages I need to know on what assumptions the decisions were based." (Assembly controller 22)* *"In the FFE phase much is undecided, thereby I do not need to calculate to the last penny. I have many unknowns of the technical solutions available that €100 or €200 more or less do not matter." (Parts controller 5)* *"In the FFE phase I need explainability. Especially when innovative parts need to be calculated, such as new active roll stabilization concepts, we have dozens of assumptions for the cost calculation. When the selling price does not cover the estimated costs, the controlling and sales department are at loggerheads and must explain their assumptions." (Parts controller 18)*
4) EPD: Accuracy is more important, since more **facts** are available to rely on (*more is decided*).	*"At a later time, orders are partly placed, and specific information is already known. In addition, I have actual costs and precise calculations of the product. When I have a circuit diagram in the late phase, I can precisely calculate the cost [of the wiring]." (Parts controller 3)* *"The product is already defined as such and therefore will not be questioned." (Parts controller 8)*
Panel B: New arguments and quotes supporting the hypothesis *(inductive)*	
Panel C: Arguments and quotes that help understand unexpected results *(inductive)*	

4.7.2 Information Uncertainty

The results of the experiment supported H2: The importance of explainability relative to accuracy is higher in the case of high information uncertainty than in low information uncertainty. The arguments we developed for H2 are summarized in Panel A in **Table 30** alongside with several statements from the experiment.

One participant's statement stimulated us to consider another theoretical argument for H2. As indicated in Panel B of **Table 30**, the participant mentioned that the assumptions may change more often, making accuracy less important. Due to the fast-moving nature of product development at high uncertainty, the calculation of highly accurate cost estimates is not necessary. This argument is in line with the increasingly agile product development processes, where every modification of a product that affects the costs must be considered when tracking product costs (Germani et al., 2011). Short duration of validity could be another theoretical argument why higher information uncertainty makes accuracy less important than explainability.

One participant helped us to come up with a theoretical explanation for the unexpected result of why the indirect measure is not significant. Some participants stated that the more accurate GBR model was often more useful in the case of high information uncertainty as it provides an easy and quick way to calculate many product designs. Manual adjustments would often be required when applying the CBR and MLR models to achieve high accuracy, which are not always necessary in the case of the GBR model. If many scenarios or different product designs need to be evaluated, a complex machine learning method is therefore more adequate. This corresponds to Deng and Yeh (2010), who state that machine learning is beneficial when conducting design-to-cost approaches, since machine learning models don't rely on human-based judgment and they can quickly update cost when the design of a product changes during new product development.

Table 30: Evaluation of the factor information uncertainty (high or low)

Panel A: Quotes supporting original arguments for the hypothesis *(deductive)*	
1) High uncertainty: Explainability is more important, due to **manual adjustments** to integrate expert knowledge.	*„In case of information uncertainty explainability is more important. The results of such models often must then be further adjusted, this has to be somehow explained." (Parts controller 10)* *„When dealing with information uncertainty explainability is more important. I need to know how numbers can be influenced" (Assembly controller 22)*
2) High uncertainty: Explainability is more important since more **assumptions** concerning technology and market environment are required.	*"In case of uncertainty, explainability is more relevant as there are more active discussions. Furthermore, the assumptions are not set and need to be fixed. These assumptions are then often questioned along with the [cost estimation] model." (Parts controller 2)*
3) Low uncertainty: Accuracy is more important since more **facts** and certain information is available to rely on.	*"In case of high information certainty, the reference is fixed. When the reference is accepted explainability is less important. Consequently, accuracy can be higher." (Parts controller 5)* *"In case of information certainty, the assumptions are agreed upon, 'That is how it is…'. Therefore, I need to give less explanations." (Parts controller 17)* *"In case of high certainty, the goal is often to reach the key competitor. To do so I need accuracy." (Parts controller 9)*
Panel B: New arguments and quotes supporting the hypothesis *(inductive)*	
4) High uncertainty: Accuracy is less important due to the low **duration of validity**.	*"There is no need to calculate down on the last penny when everything will look different in two months." (Assembly controller 28)*
Panel C: Arguments and quotes that help understand unexpected results *(inductive)*	
5) High uncertainty: Complex machine learning models are more adequate when the cost of **many different designs and scenarios** need to be estimated.	*"Complex models such as GBR can also be used to optimize product designs. [...] The calculation of many different product designs by hand would otherwise take too long." (Parts controller 1)*

4.7.3 Cost Granularity

The results of the experiment supported H3: The importance of explainability relative to accuracy is higher in the case of aggregated, low cost granularity (total cost) than in the case of detailed, high cost granularity (assembly-level costs). The arguments we developed for H3 are summarized in Panel A of **Table 31** together with several statements from the experiment.

The participant’s statements also suggest two more theoretical arguments for H2, which are summarized in Panel B of **Table 31**. First, participants explained that the relative cost prediction performance of total- and assembly-level cost is important to consider when selecting a machine learning model. The larger prediction error of the assembly level was considered by many participants in the model selection process to be critical.[7] In the case of total cost prediction, accuracy is therefore less important due to the relatively high level of accuracy. A minimum level of predictive accuracy is required to ensure the applicability of cost estimates in practice (Verlinden et al., 2008). The minimum level of predictive accuracy of cost estimates depends on several factors, such as the phase during new product development (Li Qian & Ben-Arieh, 2008) and the error compared to alternative methods (i.e., manual calculations) (Caputo & Pelagagge, 2008). This leads to another theoretical argument where in the case of assembly cost prediction, accuracy is more important, because of the relatively low accuracy, which did not exceed an acceptable threshold. Second, one participant mentioned that on a higher cost granularity more specialized knowledge is necessary. In such cases, it is usually difficult for non-experts to join discussions and challenge cost forecasts. Understanding specific jargon and asking the right questions is difficult for non-experts when interacting with domain expert teams (Markus, 2001). Therefore, less explainability is needed when specialized knowledge is required. The necessity of expert knowledge on more granular cost levels could be another argument why explainability is more important on more detailed cost levels.

[7] In the case of total cost estimation, the NMAE ranges from 4.86% (GBR) to 10.16% (CBR). On assembly level, the NMAE ranges from 13.98% (GBR) to 18.75% (CBR).

Table 31: Evaluation of the factor cost granularity (complete or assembly level)

Panel A: Quotes supporting original arguments for the hypothesis *(deductive)*	
1) Complete: Explainability is more important due to the **heterogeneous set of many components**.	*"The complex settling and prioritization of product features between the assemblies requires explainability: 'The chassis assembly gets the air spring; therefore, we will have an inferior interior [in the body assembly]'. The total cost matters in accordance with the target properties." (Complete vehicle controller 13)* *"Explainability is important for the reconciliation and comparison with other projects with internal and external benchmarks." (Parts controller 36)*
2) Assembly: Accuracy is more important due to the **homogenous set of few components**.	*"Since the granularity is higher on assembly level there is the expectation of a higher accuracy due to the reduced complexity." (Complete vehicle controller 7)*
Panel B: New arguments and quotes supporting the hypothesis *(inductive)*	
3) Assembly: Accuracy is more important due to the **relatively low accuracy of multi-target predictions**.	*"The inaccuracy is much higher on the assembly level. Since the roughly 20% accuracy score of the CBR method is by far too inaccurate, I inevitably need to use more accurate models such as MLR or GBR." (Parts controller 8)* *"I am automatically more accurate on the complete vehicle level." (Assembly controller 11)*
4) Assembly: Explainability is less important due to **specialized knowledge**.	*"On a certain level of detail and cost granularity there are less experts that are knowledgeable. For example, in the case of calculations of highly innovative production processes, where you need to become acquainted with a highly specialized topic first." (Parts controller 31)*
Panel C: Arguments and quotes that help understand unexpected results *(inductive)*	

4.7.4 Product Novelty

The results of the experiment only partly supported H4: The importance of explainability relative to accuracy is higher in the case of incremental innovation than for radical innovation. As the direct measurement is not statistically significant, we cannot completely confirm the significance of this factor. The arguments we developed for H4 are summarized in Panel A of **Table 32** together with several participant statements.

The participant's statements also stimulated another argument for H4, which is summarized in Panel B of **Table 32**. Two participants explained their choice of machine learning model by the difficulty to provide accurate estimates in the case of radical innovation compared to incremental innovation. When dealing with radical innovation, accuracy is more important because of the relatively low level of accuracy as cost estimations must be conducted from scratch. In the case of incremental innovation, accuracy is less important, because of the relatively high level of accuracy due to the availability of existing cost references. This corresponds to the nature of incremental products, which commonly involves familiar technologies, known markets, and high predictability of outcomes (Schmidt et al., 2009). Accordingly, the statements suggest another theoretical argument where the high accuracy of incremental innovation makes accuracy relatively less important compared to radical innovation.

We also used participant’s statements to come up with an explanation why the direct measure is not significant. Three distinct explanations were suggested. First, one participant mentioned that a similar approach to the CBR method is normally used for incremental innovation and therefore represents the current practice at *AutomotiveCompany*. The fit of the CBR method to the approach commonly applied when estimating the cost of incremental innovations could be one explanation why more importance is placed on explainability in the case of the indirect measure but not for the direct measure. Literature considers that the behavior of individuals is not just determined by their beliefs and attitudes, but also by their habits (Burton-Jones & Hubona, 2006). In most cases, the frequency in which a behavior has been carried out in the past is also a reliable predictor of future action (Ajzen, 2002). In the context of technology acceptance, the experience with information technology has a positive impact on system usage through habit formation (Burton-Jones & Hubona, 2006). The longer an individual has used a system, the more likely it will become a habitual tool. Second, a participant explained that in the case of radical innovation there is more focus on conceptual work where more explainability is needed. The conceptual work could be one theoretical explanation why more importance is placed on explainability during radical innovations. Radical innovation involves the identification of new concepts, while incremental innovation improves a given solution by modifying available concepts (Norman & Verganti, 2014). Third, two participants stated that in the case of radical projects more assumptions are required that need to be explained. The larger number of assumptions could be another explanation why radical innovation makes explainability more important than accuracy. Radical innovations build on several assumptions about the future customer value and the company's solutions to satisfy the customer needs (Gudem et al., 2014). In addition, radical innovation redirects a company to new markets, which implies turning away from existing assumptions (Herrmann et al., 2007).

Table 32: Evaluation of the factor product novelty (radical or incremental)

Panel A: Quotes supporting original arguments for the hypothesis *(deductive)*	
1) Increm.: Explainability is more important, due to in-depth **comparisons** with the **predecessor**.	*"With existing predecessor projects explainability is more important. Everyone is raising questions about how it was with the predecessor project." (Parts controller 3)* *"Everyone would ask why, if it costs €10 more compared to the predecessor project." (Parts controller 31)* *"Here the acceptance is much more important. One is always more critical in the case of successor projects." (Complete vehicle controller 15)*
2) Radical: Accuracy is more important since **no comparisons** with the **predecessor** are requested.	*"In radical projects accuracy is more important. Nobody can challenge that I might be wrong in this case." (Complete vehicle controller 15)* *"I don't need to explain myself, since there is no real predecessor project. I have a larger playing-ground with radical projects. The more possibilities I have, the more artificial intelligence tools can be applied." (Parts controller 1)* *"In case of radical projects there is no predecessor so the necessity for explainability is reduced." (Parts controller 3)*
Panel B: New arguments and quotes supporting the hypothesis *(inductive)*	
3) Increm.: Accuracy is less important due to the relatively **high level of accuracy**.	*"One is already more accurate in case of successor projects. Explainability is more important here." (Parts controller 16)*
4) Radical: Accuracy is more important due to the relatively **low level of accuracy**.	*"In case of radical projects, a more or less similar reference is first needed, which means there will be lower accuracy. Consequently, I need a model that increases this accuracy." (Parts controller 6)*
Panel C: Arguments and quotes that help understand unexpected results *(inductive)*	
5) Increm.: Explainability is more important since the CBR model fits the **current the practice**.	*"I'd rather choose CBR in successor projects since we do it like this when dealing with incremental projects. Therefore, explainability is more important." (Parts controller 25)*
6) Radical: Explainability is more important due to the focus on **conceptual work**.	*„In radical projects I need more explainability. Thereby, I do not need to calculate down to the last penny. Moreover, I need to know the design configuration of the new car. What features does it have? How is the car conceptualized?" (Parts controller 14)*
7) Radical: Explainability is more important since more **assumptions** are required.	*"Explainability is more important in radical projects since I have many assumptions that I need to explain in this situation." (Parts controller 18)* *"In radical projects I need more explainability since I need to estimate costs for new components. Often, I need to explain the calculation steps and the underlying assumptions." (Parts controller 4)*

4.7.5 Target Cost Gap

The results of the experiment supported H5: The importance of explainability relative to accuracy is higher in the case of high target cost gap than in low target cost gap. The arguments we developed for H5 are provided in Panel A of **Table 33** together with several statements from the participants.

Some participants' statements stimulated us to consider two more theoretical arguments for H5. As indicated in Panel B of **Table 33**, some participants mentioned that when dealing with high cost gaps explainability becomes more important due to the low acceptance of target costs. Goal commitment is negatively related to goal difficulty (Erez & Zidon, 1984). Therefore, extreme levels of goal difficulty limit the effectivity of goal setting (Yukl & Latham, 1978). In the field of XAI, explainability is more important when there is low trust into a machine learning model (Alonso et al., 2015). The low acceptance of target costs could be another theoretical argument why higher target cost gaps make explainability more important than accuracy. Still, it should be mentioned that understanding and trust are linked but not mutually dependent: One can trust a model in its predictions without being able to explain it, and vice versa (Hall & Gill, 2018). Furthermore, participants stated that accuracy is more important to validate the cost gap when dealing with lower target cost gaps. High prediction errors could mistakenly discard a profitable product when costs are overestimated or mistakenly include low-margin products when costs are underestimated. When comparing cost forecasts and cost goals, a high accuracy level of cost forecasts is important to correctly select products with high margins and discard loss-making products (Joseph & Vetrivel, 2012). This suggests another argument for H5, where the validation of the cost gap makes accuracy more important when dealing with low target cost gaps.

Table 33: Evaluation of the factor target cost gap (high or low)

Panel A: Quotes supporting original arguments for the hypothesis *(deductive)*	
1) High target cost gap: Explainability is more important due to a **higher pressure** on the project.	*"The project is in greater focus in the case of 20% target cost gap. It is likely that the project will not pass the next development milestone and will be halted." (Complete vehicle controller 7)* *"In the case of 20% I need explainability. There are cost workshops where each part and even special equipment is questioned." (Parts controller 18)*
2) Low target cost gap: Explainability is less important due a **lower pressure** on the project.	*"There are less questions with 5%. In other words, I need less explainability. The purchasing department can then catch up to the remaining 5% with purchasing performance." (Parts controller 17)* *"I am more in a comfort zone in a situation with a 5% deviation. No one is bothered since other problems of other projects are more important." (Complete vehicle controller 19)*
3) High target cost gap: Explainability is more important due to the requirement of **detailed information**.	*"With deviations of 20% I need explainability. I need to identify the drivers and reasons why I am that far away from the target." (Parts controller 14)* *"I need to know where I can find opportunities for improvement. Therefore, I need additional information, such as sourcing data or technical drawings." (Parts controller 1)*
Panel B: New arguments and quotes supporting the hypothesis *(inductive)*	
4) High target cost gap: Explainability is more important due to a **low acceptance of target cost**.	*"In a situation with 20% cost gap, I must bring people on board. Otherwise, the target cost won't be accepted." (Assembly controller 23)* *"There are many questions of how the target development is formed and what is behind that target cost with such deviations." (Parts controller 3)*
5) Low target cost gap: Accuracy is more important as the target deviation **needs to be validated**.	*"With cost gaps of 5% I need accuracy since I need to make sure the cost gap is indeed only 5%." (Parts controller 9)* *"A prediction error of 5% is more critical in the case of 5% cost gap then in the case of 20% cost gap." (Parts controller 31)*
Panel C: Arguments and quotes that help understand unexpected results *(inductive)*	

4.8 Conclusion

This chapter investigates factors influencing the relative importance of explainability to accuracy in the context of cost estimation during new product development. By conducting a within-subject fractional factorial experiment, we statistically test five factors that are expected to determine the relative importance between explainability to accuracy: development phase progress, information uncertainty, cost granularity, product novelty, and target cost gap. We show that the adequacy of machine learning methods for cost estimation significantly depends on the specific situations and conditions during new product development.

4.8.1 Research Implications

This paper provides several contributions to the literature. First we empirically confirm that people perceive the trade-off between accuracy and explainability. The cost estimation literature often assumes an inverse relationship between accuracy and explainability without empirical evaluation (Cavalieri et al., 2004; Loyer et al., 2016; Verlinden et al., 2008). Our experiment shows that both attributes are indeed significantly negatively correlated (r = -0.586). Moreover, the relationship matches the transformation curve depicted in Barredo Arrieta et al. (2020).

Second, we provide evidence that the relative importance of explainability to accuracy matters for the selection of machine learning methods for cost estimation. In the integrated technology acceptance model, accuracy and explainability are important factors expressing the information quality of a system and thereby positively influence the use and value of information technology (Wixom & Todd, 2005). In the machine learning model selection process, accuracy and explainability are expected to be main determinants for the adequacy of a model (Boehm et al., 2019; Tripathi et al., 2021). In the cost estimation literature, interpretability is expected to be a limiting factor of machine learning methods (Cavalieri et al., 2004; Loyer et al., 2016). Often more simpler cost estimation models are used despite being less accurate compared to machine learning models (Verlinden et al., 2008). We contribute to this stream of literature by empirically confirming that the relative importance of explainability to accuracy is a significant factor in the model selection process. However, we find that other factors, such as the availability of common practices of solving a cost estimation task at a company and the ability to calculate many different product designs, are also important when selecting between machine learning models for cost estimation. We show that despite having much higher perceived accuracy (GBR 5.85 in contrast to 4.58 and 3.25 of MLR and CBR on a 7-point Likert scale), the complex machine learning method was considered adequate in only few situations of new product development. The GBR was selected in 26.09% of the cases, while the much simpler MLR and CBR methods are perceived adequate in most cases (73.91%). We add to the cost estimation literature by demonstrating the significance of the relative importance of explainability to accuracy in the machine learning model selection process. We confirm that the lack of interpretability is indeed a major limitation for product cost estimation in practice when applying machine learning. Our findings therefore challenge the current research trend of using machine learning within the environment of product cost estimation (Caputo & Pelagagge, 2008; Cavalieri et al., 2004; Chou & Tsai, 2012; Deng & Yeh, 2011; Loyer et al., 2016; Verlinden et al., 2008).

Third, we identify several factors that determine the relative importance of explainability to accuracy of machine learning methods for product cost estimation during new product development. The XAI literature identified several general factors based on theoretical approaches and conceptual frameworks that influence the selection of machine learning models such as the need for interaction with models (Alonso et al., 2015; Frias-Martinez et al., 2005), the available trust (Alonso et al., 2015), consequences of inaccurate predictions (Baryannis et al., 2019), and the need to validate outcomes (Alonso et al., 2015). We add to this literature stream by empirically testing the importance of context-specific factors. Especially for product cost estimation during new product development, we investigated five factors that determine the relative importance between explainability and accuracy:

- *Development phase progress*: Explainability relative to accuracy is more important in the FFE design phase than in the EPD phase. This factor matters due to the difference between conceptual and implementation work and because of the different number of assumptions required.
- *Information uncertainty*: Explainability relative to accuracy is more important in the case of high information uncertainty than for low information uncertainty. This factor is important because of the requirement of manual adjustments, the necessity of assumptions, and different durations of validity of product designs.
- *Cost granularity*: Explainability relative to accuracy is more important in the case of low cost granularity (i.e., total cost) than for high cost granularity (i.e., assembly cost). The homogeneity of components, the difference of accuracy levels of cost estimates, and the requirement of specialized knowledge for different cost granularities make this factor important.
- *Product novelty* has a significant impact on machine learning model selection but has no significant impact on the importance of explainability relative to accuracy. On the one side, the need for comparisons with predecessors and the low level of accuracy *make accuracy more important* in the case of radical innovation; on the over side, the focus on conceptual work and the requirement of many assumptions *make explainability more important* in radical innovation. Overall, we cannot

confirm the statistical significance of this factor on the relative importance of explainability to accuracy.

- *Target cost gap*: Explainability relative to accuracy is more important in the case of high target cost gaps than for low target cost gaps. This factor matters since it affects the pressure on the project, the required level of detail of information, the acceptance of cost goals, and the need for validation of the current cost gap.

In sum, we contribute to the literature by identifying several factors that significantly determine the importance of explainability relative to accuracy and the model choice for cost estimation during product development. These findings are relevant given the popularity of machine learning in product cost estimation (Caputo & Pelagagge, 2008; Chou et al., 2010; Deng & Yeh, 2011; Loyer et al., 2016) and the often-discussed interpretability problem in this field (Cavalieri et al., 2004; Loyer et al., 2016; Verlinden et al., 2008).

4.8.2 Limitations

The generalizability of findings is limited by certain aspects. First, the experimental task covers the estimation of cost during the rather early phase of new product development. For the estimation of costs at later phases other factors might be relevant. Second, the experiment is conducted at a case company with very little experience with machine learning applications. As the participants can be regarded as artificial intelligence novices (average machine learning experience of 2.00 on a 7-point scale), the selection between models might be different in other environments with more experienced machine learning users. Third, as this study incorporates a within-subject design, it is subject to the typical limitations regarding fatigue and carry-over effects. Carry-over effects are mitigated by randomization, while the awareness of the changing factors was used on purpose to get insights into the underlying reasoning behind the selection of machine learning models. Finally, the within-subject design with five factors was challenging for participants to distinguish between treatment combinations. In some cases, the participants mentioned that they focused mainly on a subset of factors and used the remaining factors to adjust for nuances. This might lead to inconsistent results, which also reduces the coefficient of determination of the regression analysis.

4.8.3 Future Research

This experiment opens some promising avenues for further research. On the one hand, the discussions at the end of the experiment introduced two new potential factors that could also play a significant role for the relative importance of explainability to accuracy: the *level of hierarchy* and the *complexity of parts*.

> *"The information need of recipients does highly depend on the project environment and the* ***level of management hierarchy****. (…) The higher the hierarchy level, the broader and less detailed the results should be. In lower levels of hierarchy, there are more experts and therefore explainability is needed." (Complete vehicle controller 7)*

Top managers usually prefer more aggregated information for economic decision-making, while middle management needs information in greater detail (Odar et al., 2015). Aggregated information is required for strategic planning tasks, while more detailed information is necessary on operation control tasks (Gorry & Scott Morton, 1989). Accordingly, we expect that the importance of explainability relative to accuracy is higher in the case of interactions with middle/low management than in the case of top management.

> *"When dealing with* ***technical black boxes,*** *such as electronic control units, there are anyhow less questions. In the case of simple and tangible parts everyone has a say, and it gets complicated." (Parts controller 26)*

It is expected that the need for explainability is lower in the case of highly complex components, where mainly domain experts are involved. Non-experts struggle with the jargon, finding the right questions to raise when interacting with experts (Markus, 2001). Moreover, non-experts require information in an accessible way. Therefore, we expect that the importance of explainability relative to accuracy is higher when dealing with low part complexity than in the case of high part complexity.

On the other hand, it would be interesting to analyze antecedents of relative importance of explainability to accuracy in other areas during new product development. As this experiment covers only the aspect of cost estimation, other aspects of cost management or decision making are left for future research.

4.9 Appendix C.

Questionnaire and experimental procedure

1. **Introduction**
   a. Department: ____________
   b. Work experience (in years): ______
   c. Experience with machine learning
   very low ☐–☐–☐–☐–☐–☐–☐ *very high*
2. **Introduction of the three machine learning models**
3. **Model rating (CBR, MLR, GBR): Transparency**
   a. I can fully understand how the cost estimation model works.
   I strongly disagree ☐–☐–☐–☐–☐–☐–☐ *I strongly agree*
   b. I can understand the best reasons for the results provided by the model.
   I strongly disagree ☐–☐–☐–☐–☐–☐–☐ *I strongly agree*
   c. I understand how the model comes to its solution.
   I strongly disagree ☐–☐–☐–☐–☐–☐–☐ *I strongly agree*
4. **Presentation of video clips of the three machine learning models**
5. **Model rating (CBR, MLR, GBR): Interpretability**
   a. The explanation of results is obvious to me.
   I strongly disagree ☐–☐–☐–☐–☐–☐–☐ *I strongly agree*
   b. I can explain the results to another person.
   I strongly disagree ☐–☐–☐–☐–☐–☐–☐ *I strongly agree*
6. **Presentation of empirical test results for each machine learning model**
7. **Model rating (CBR, MLR, GBR): Accuracy**

   Please rate the accuracy. Accuracy of a cost estimate is defined as the deviation of the forecasted cost from the actual cost of the project.
   Not accurate ☐–☐–☐–☐–☐–☐–☐ V*ery accurate*
8. **Experimental task**

   The product planning team requests a change of product design for a car model during development. Your objective is to predict the updated costs for the new design.

9. Condition evaluation (16 treatment conditions)

a. Relative importance of explainability to accuracy in this scenario.
Accuracy ☐–☐–☐–☐–☐–☐–☐–☐–☐–☐ *Explainability*

b. In this scenario I intend to use the following method.
☐ *GBR* – ☐ *MLR* – ☐ *CBR*

10. Discussion of each of the five factors

When comparing both situations of a given factor, what is more important: explainability or accuracy? Why?

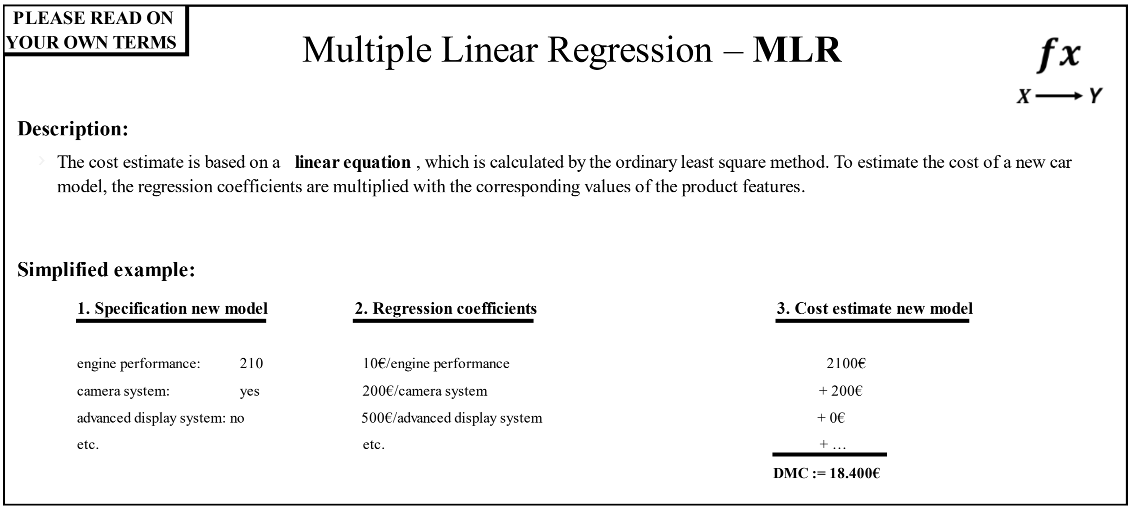


Figure 28: Description of the MLR method

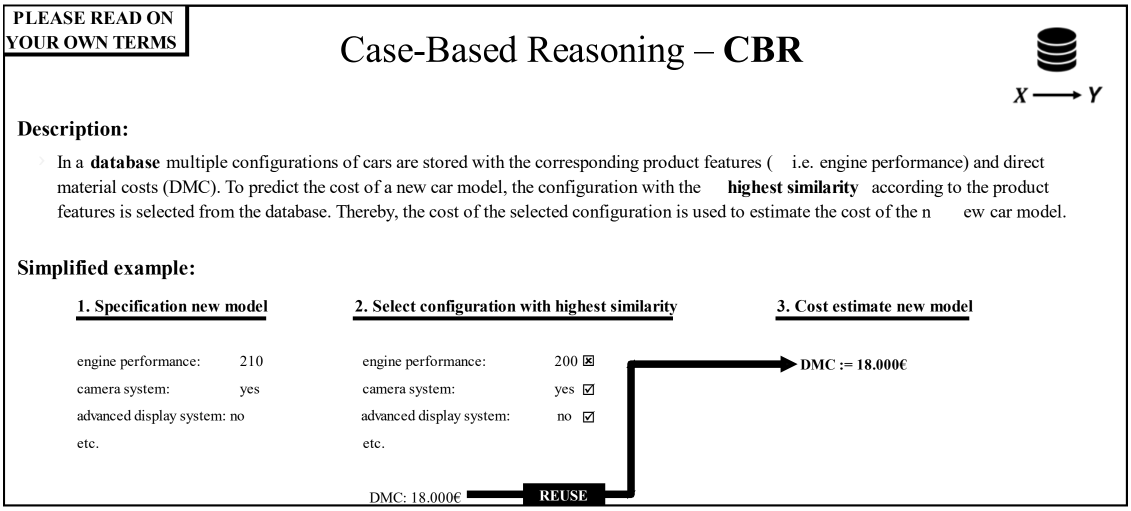


Figure 29: Description of the GBR method

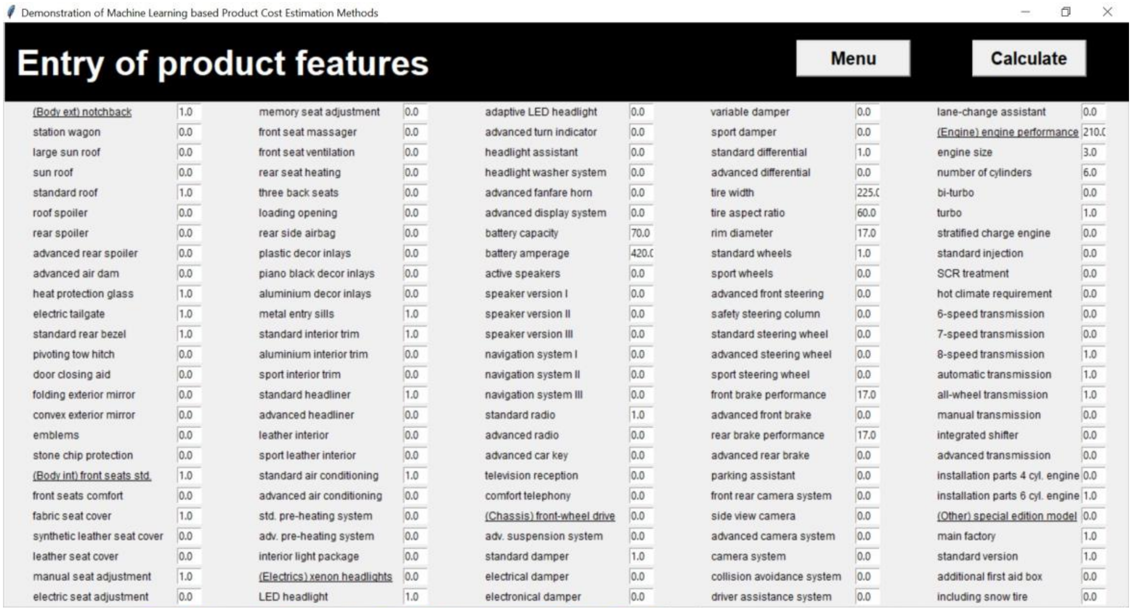


Figure 30: Demo Tool: Entry of features. The entry form consists of the 125 most important features according to the GBR model. The same features for each model are used, to make sure that the models are only distinguishable by their explanations.

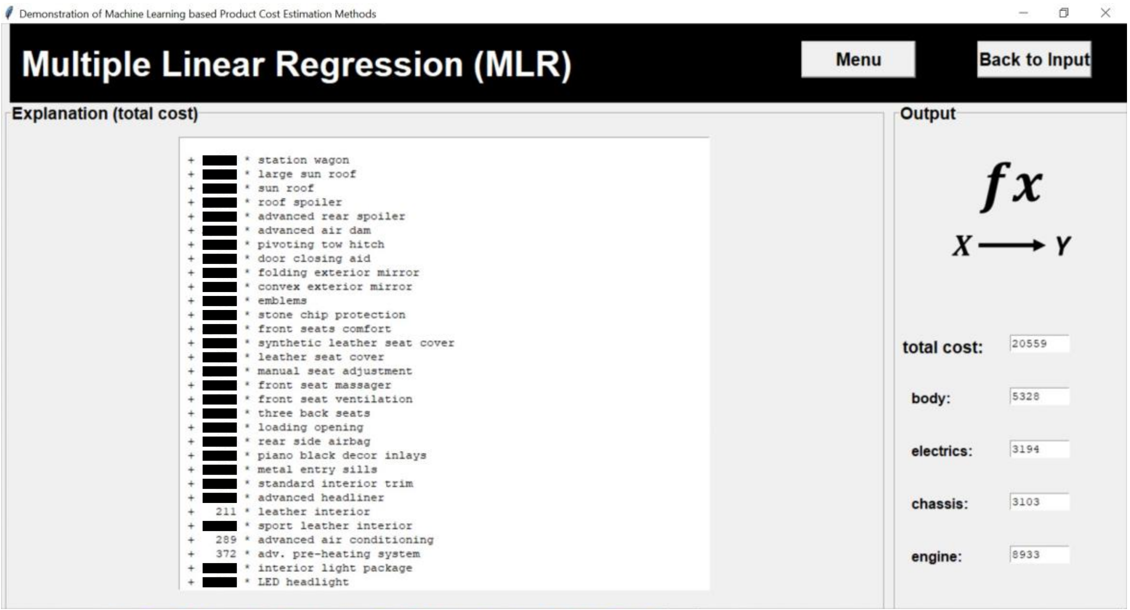


Figure 31: Demo Tool: MLR

Figure 32: Demo Tool: GBR

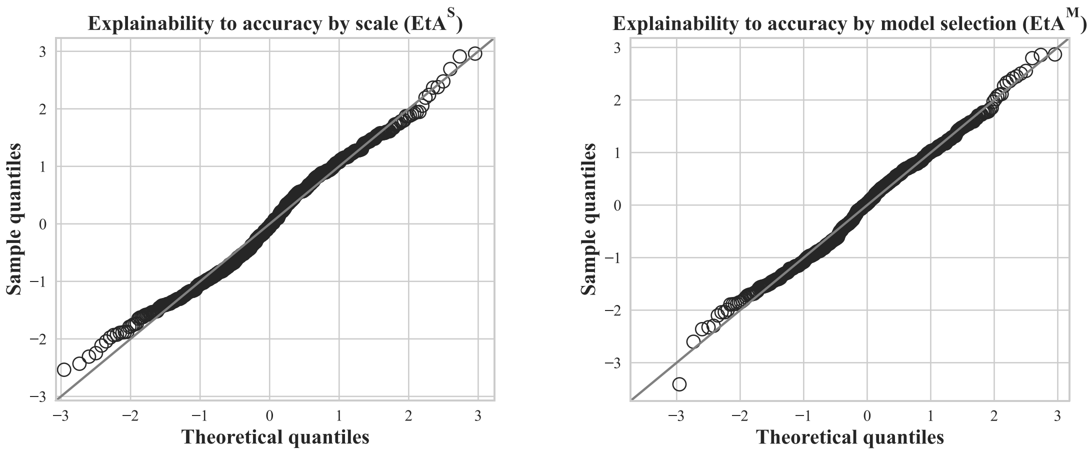


Figure 33: Quantile-quantile plots for explainability to accuracy by scale (EtA^S) and by model selection (EtA^M). The quantile-quantile plots of both variable measures suggest that residuals are normally distributed.

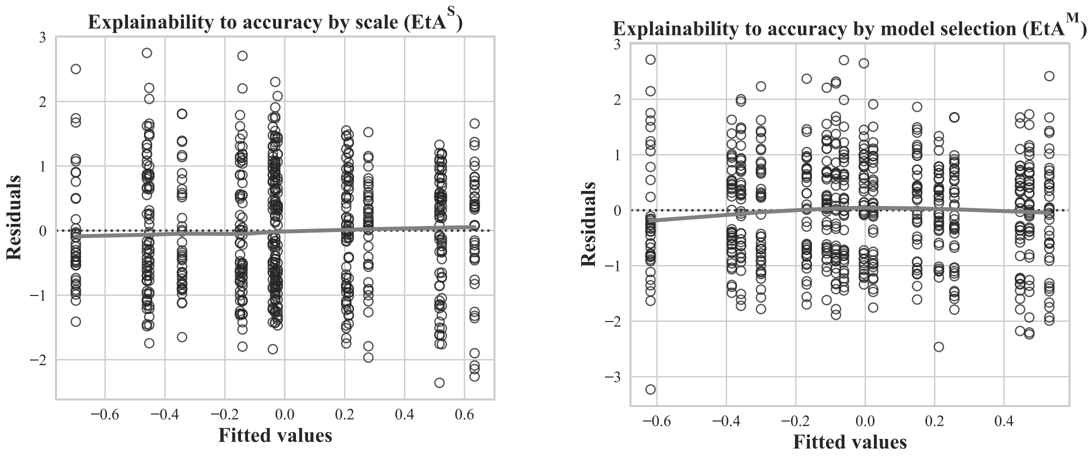


Figure 34: Residuals vs. fits plots for explainability to accuracy by scale (EtA^S) and by model selection (EtA^M). The residuals vs. fits plots of both variable measures suggest homoscedasticity of residuals.

5 Cost Reduction Performance with Target Costing: The Role of Product Design Interdependence and Uncertainty about Target Cost Difficulty

Abstract

Target costing involves the challenge to meet a specific and usually difficult product cost goal during product development. Cost goals that are seen as difficult but doable can enhance motivation and performance. However, setting such goals and improving performance will not always be possible in complex product development activities, which are the focus of this study. The empirical study is based on proprietary target costing data of development projects in a car company, uniquely combining data on cost reduction during the development stage and the production stage of new products. This study shows that product design interdependence moderates the impact of target cost difficulty on cost reduction performance. Product development projects become interdependent when these products share technology, such as common parts or product platforms. Interdependence makes product development more complex, and teams may have less complete task knowledge. As a result, specific, difficult target cost goals less positively improve cost reduction performance. Furthermore, this study shows that uncertainty about target cost difficulty reduces cost reduction performance during development. Because of complexity and resulting uncertainty, it may not be feasible to always set cost goals on such a level, that these are difficult but doable. When target cost difficulty is uncertain, teams may at some point conclude that a cost target is easy, and they would be able to reduce costs more than required. This may lead teams to reduce their cost reduction efforts and aim for reaching the cost target instead of achieving maximum cost reduction performance. We show that some cost reduction performance is shifted from the development to the production phase and that this is associated with less cost reduction performance in total.

Keywords: Target costing; New product development; Goal setting theory; Uncertainty; Product design interdependence; Modularity; Product platforms; Parts commonality

5.1 Introduction

The life-cycle stage of product development is crucial for managing the costs and profitability of products (Booker et al., 2007; Cooper & Slagmulder, 2004b; T. Davila, 2000; A. Davila et al., 2009; A. Davila & Wouters, 2004). More degrees of freedom exist to influence the design, features, and performance of a product, thereby affecting future costs and revenues, compared to the later production stage. Furthermore, cost reduction solutions are implemented earlier—from the start of production, instead of gradually during the production stage—so total cost savings are larger (Afonso et al., 2008). Finally, the costs of redesign activities are avoided, compared to when a product is first completely developed, launched, and subsequently needs to be changed in order to implement cost saving measures.

Target costing represents the approach for managing product costs during new product development that has often been studied in management accounting research (Ansari et al., 2006). The essential idea is that, first, the allowable product cost is determined by the attainable sales price and the required profit margins and second, the product cost is evaluated during the product development stage. If the estimated product cost exceeds the allowable product cost, the product design needs to be adjusted. In its purest form (also called the *cardinal rule*), product development cannot be completed until solutions have been found to meet the target cost (Cooper & Slagmulder, 1999). Thus, target costing involves the challenge for employees, typically in teams, to meet a specific and typically quite challenging goal (Booker et al., 2007). As predicted by goal setting theory (Locke & Latham, 2002), the presence of specific and challenging cost targets can lead to achieving more cost reduction during product development than general *do-your-best* goals (Everaert & Bruggeman, 2002; Everaert & Swenson, 2014; Gopalakrishnan et al., 2015).

However, target costing may not so straightforwardly enhance motivation and performance in a more complex organizational product development context. The effects of target costing are more intricate in organizational settings that also involve the need to manage the cost of shared resources (A. Davila & Wouters, 2004), where complexities from concurrent engineering play a role (Gopalakrishnan et al., 2015), where other, implicit incentives for developers exist (Mihm, 2010), and information overload may occur from the combination of costing information and nonfinancial performance measures in product development (Henri & Wouters, 2020).

The current chapter also addresses target costing in a more complex product development context, which is characterized by *product design interdependence* and *uncertainty about target cost difficulty*. First, we investigate the effect of product design interdependence on cost reduction performance in a target costing context. Firms may use not only target costing for separate, independent product development projects, but they may also manage costs through coordinated product development decisions across product development projects. Examples are the deliberate use of common parts or common processes (A. Davila & Wouters, 2004; Labro, 2004). As product development projects are more interdependent, more coordination is required and there are fewer degrees of freedom for teams to reduce the costs of their own product. We hypothesize a negative relationship between product design interdependence and cost reduction performance. More coordination requirements due to greater product design interdependence also means that product development is more difficult, and teams may have less complete task knowledge. Incomplete task knowledge can make goal setting less effective for increasing motivation and performance (Hirst, 1987). Therefore, we expect that product design interdependence reduces the effectiveness of specific and challenging cost goals for improving cost reduction performance. Accordingly, we hypothesize that product design interdependence moderates the relationship between target cost difficulty and cost reduction performance.

Second, we investigate the effects of uncertainty about target cost difficulty. A firm aiming to set goals that are difficult but doable, in line with goal setting theory (Locke & Latham, 2002), may actually set goals that turn out to be very easy or practically impossible to achieve. In the complex organizational context of target costing, the difficulty of cost goals may be uncertain because of long lead times of product development, which create uncertainty about future sales prices, product attributes, and costs. Moreover, long supply chains create uncertainty about the relationship between market prices and targets for the manufacturing costs of individual parts (Stadtherr & Wouters, 2021). We expect that uncertainty about target cost difficulty reduces the effectiveness of having specific goals for improving cost reduction performance. If goals turn out to be difficult but doable, they would be expected to enhance motivation and improve cost reduction performance. But, if it becomes clear during product development that the goal is easily achievable and the product cost will likely land below the cost target, employees may react differently. Instead of going for the maximum cost reduction performance, employees may *slow down* and try to land with a product cost that is around the target cost. More cost reduction performance is not required. By doing so, teams would exploit fewer cost reduction opportunities during the product development stage and potentially leave more of these for later, during the production stage of the product. Therefore, we hypothesize a negative relationship between cost reduction performance during development and cost reduction performance during production, and this relationship would be moderated by target cost attainment.

Furthermore, by considering both, cost reduction performance during development as well as during production, we investigate a basic premise of target costing. The fundamental idea of target costing is that the product development stage offers more degrees of freedom for changing the product than the production stage and, therefore, more potential for reducing costs exist during product development than during production. Reversely, achieving less cost reduction performance during development and postponing some cost reduction performance to the production stage, would result in less cost reduction performance in total, so when looking at the development and production stages together. Therefore, we hypothesize that achieving relatively more cost reduction performance during development than production is associated with greater total cost reduction performance.

The empirical research is based on proprietary archival data of a car company. The case company represents a complex product development setting with interdependence and uncertainty about target cost difficulty. For several hundreds of different parts, the data include the part cost at several points in time (at start of development, start of production, and 12 months after start of production), the part cost target (at start of development and start of production), and several part characteristics. We used these data for measuring cost reduction performance, target cost difficulty, and target cost attainment. Product design interdependence is measured in two ways: A variable for each part that describes how many other parts that are included in a particular car affect that part’s technical design, and by the distinction of platform parts versus hat parts. Hat parts are used for one car project only and interdependence is low. Platform parts are used over multiple projects because these encompass the basic architecture of a product, providing the basis for a series of derivative products. Interdependence is high for platform parts.

We find that cost reduction performance is positively related to target cost difficulty and negatively related to product design interdependence. We also find support for the hypothesized moderating effect: The relationship between target cost difficulty and cost reduction performance is weaker for parts with greater interdependence.

Furthermore, several results are consistent with the *slowing down* behavior because of uncertainty about cost target difficulty in the case company. Descriptively, we find a positive, nonlinear relationship between cost reduction performance and target cost attainment, suggesting that cost reduction performance during development flattens as target cost attainment is greater. Specifically, regarding our hypotheses, we find a negative relationship between cost reduction performance during development and during production and, as hypothesized, this relationship is moderated by target cost attainment. As target cost attainment is better, the negative relationship between cost reduction performance during development and during production is stronger. Furthermore, we construct a measure that captures the extent to which cost reduction is taking place more during development than during production, and we find that this measure is positively related to total cost reduction.

The contribution of this paper is to provide a better understanding of the effectiveness of target costing in more complex organizational product development contexts. First, our finding that target cost difficulty is related to more cost reduction performance during product development is based on archival company data, and thereby complements results from earlier studies that were based on experimental tasks (Everaert & Bruggeman, 2002; Gopalakrishnan et al., 2015). Second, we demonstrate that in a complex product development context, product design interdependence and uncertainty about target cost difficulty may both limit the effectiveness of target costing. Product design interdependence is associated with less cost reduction performance, and it moderates the relationship between target cost difficulty and cost reduction performance. Uncertainty about target cost difficulty means target costs cannot always be determined as difficult, but doable goals. This study uniquely combines data on cost reduction during the development and the production stage, and it shows that uncertainty about target cost difficulty leads to shifting cost reduction performance to the production stage and limits total cost reduction. These findings on product design interdependence and uncertainty about target cost difficulty complement earlier studies that have investigated the role of target costs in more complex organizational contexts (A. Davila & Wouters, 2004; Gopalakrishnan et al., 2015; Henri & Wouters, 2020; Mihm, 2010).

5.2 Literature Review and Hypotheses Development

5.2.1 Target Costing

Target costing represents the approach for managing product costs during new product development that has often been studied in management accounting research. Target costing is a principal technique for managing life-cycle costs during the product design and development stage (Ewert & Ernst, 1999; Kato, 1993) based on a strategic, systematic profit planning process (Ansari et al., 2006; Cooper & Slagmulder, 1999). Target cost management improves the generation of new ideas for product development and the reduction of cost (Tani, 1995). Target costing is mostly adopted in assembly-oriented firms facing intense competitive pressure, and high environmental uncertainty (Dekker & Smidt, 2003; Scarbrough et al., 1991; Tani et al., 1994). Ax et al. (2008) confirm the positive association between target costing adoption and the intensity of competition, however, could not support the relationship between perceived environmental uncertainty and target costing. Target costing is more likely to be adopted when senior managers gain cash-based compensation and is avoided when mangers are offered stock-based compensation, as target costing facilitates or inhibits their personal bonus (Navissi & Sridharan, 2017). The adoption of target costing in organizations often requires a change of attitude; from depicting costs as an output of product design towards the perception of cost as a result of market-driven requirements and prices (Ibusuki & Kaminski, 2007).

Target costs are calculated as the maximum cost of a future product, ensuring the profitability goals, whilst meeting all relevant customer requirements (Everaert et al., 2006). Compared to many other cost management methods with an internal focus, target costing takes an external perspective by incorporating the target sales prices (Yoshikawa et al., 1994). Further, target costing is an important approach for interorganizational cost management, as it fosters the collaboration of design teams across organizations to jointly manage costs (Cooper & Slagmulder, 2004b). The decision on the maximum cost depends on the estimated actual cost of a future product (also called ***drifting cost***) and the ***allowable cost*** (Everaert et al., 2006). The allowable cost is calculated by deducting the target profit margin from the planned sales price. Cost goals need to be clearly specified for each development team to ensure that the overall profit goals of a company are achieved (Tanaka, 1993). The target costing process can be expanded toward the suppliers, where the allowable costs of a part constitute the ultimate purchasing price that the firm pays the supplier (Ellram, 2006). However, the estimated costs often exceed the allowable cost and there is no guarantee that the allowable costs can be achieved by designers and suppliers (Cokins, 2002; Cooper & Slagmulder, 1999). Therefore, it is necessary to find a balance between the attainability of cost goals and the profitability of products, and target costs often lie in-between the allowable costs and the forecasted actual costs (Monden & Hamada, 1991; Tani et al., 1994). To reduce the gap between target costs and current costs, several engineering techniques can be applied such as quality function deployment, value engineering, and design for manufacture and assembly (Cooper & Slagmulder, 1999). Target costing can be combined with other managerial accounting techniques such as life cycle costing, activity-based costing, and Kaizen costing (Woods et al., 2012). Cost reduction performance is higher for derivative products with low complexity than for radical innovations with high complexity (Everaert et al., 2000). Further, participating in the process of target cost setting and evaluating performance on controllable information positively impact the reduction of cost (Monden et al., 1997).

5.2.2 The Importance of Goal Setting Theory for Target Costing

One important reason for the potential beneficial effects of target costing is the motivational effect that the specific and difficult cost target provides. We define ***target cost difficulty*** as how demanding teams perceive their task for developing a product that will meet the target cost as well as all other requirements at the end of the product development process. Cost estimates for initial designs often exceed the target cost (Cooper & Slagmulder, 2004a) so teams must find solutions to reduce costs. We define ***cost reduction performance during product development*** as how much a team is able to reduce the estimated cost of their product during the product development process.

Several studies in target costing found that specific, difficult cost targets lead to more cost reduction during product development than general *do-your-best* goals (Everaert & Bruggeman, 2002; Everaert & Swenson, 2014; Gopalakrishnan et al., 2015). Gopalakrishnan et al. (2015) investigate the cost reduction performance of specific and general cost goals under sequential and concurrent new product development processes. In their experiment, design groups of three are required to redesign a small toy truck according to certain product requirements. Groups with specific cost goals achieved significantly higher cost reduction than groups with general (*do-your-best*) goals in the sequential process, while no superior cost reduction performance was observed in the case of concurrent product development. Everaert and Bruggeman (2002) analyze the impact of cost goal specificity under high and low time pressure during product development. Participants in an experiment needed to design a carpet either with the general goal to minimize the product cost, or with a specific, quantitative cost goal. The specific cost target led to higher cost reduction performance under low time pressure compared to a general goal, whereby the provision of cost targets did not deteriorate the quality of designs. Under high time pressure, however, no significant improvement for cost goal specificity was observed. Finally, Everaert and Swenson (2014) describe their experiences with using a teaching case for target costing. Small teams redesign toy trucks having either the general goal to minimize cost, or the specific goal to achieve a particular target cost. Teams with specific goals engage in higher levels of cooperation, while teams with *do-your-best* goals have lower motivation and lack in criterions to assess their progress. Studies in accounting investigating budget goals outside product development also found that specific, difficult budget goals were associated with higher performance (Webb et al., 2010).

The effect of target costs on cost reduction performance can be explained with ***goal setting theory*** (Locke, 1968). Goal setting improves task performance when goals are specific, adequately challenging, individuals have sufficient ability, feedback is provided, rewards are granted, the supervisor is supportive, and goals are accepted (Locke et al., 1981). Goals affect performance through the direction of attention toward relevant activities, mobilization of additional effort, increasing persistence, and usage of task-relevant strategies (Locke & Latham, 2002). The effects of goal setting theory have been applied and validated for individuals, groups, and organizations (Locke & Latham, 2002).

Figure 35 provides the main components of the goal setting theory. Two major properties of goals are specificity and difficulty. ***Specificity*** describes the “degree of quantitative precision” (Locke et al., 1981, p. 4) to which a goal is declared in a clear and unambiguous manner, so that individuals understand what needs to be accomplished and allows measuring the progress of a task (Aunurrafiq et al., 2015). Goal specificity primarily reduces the variability of task performance by decreasing the vagueness about what needs to be achieved (Locke et al., 1989). The second main component is ***difficulty***, which can be distinguished between goal difficulty and task difficulty. The term goal difficulty describes the specific level of required performance for a given task, whereas task difficulty refers to the nature of the activity to be carried out (Locke et al., 1981). Goal difficulty is usually defined as an increase in required production of a given task during a given time period (Campbell & Ilgen, 1976). There is strong evidence for a positive linear association between the difficulty of goals and task performance, but only given sufficient commitment of individuals (Erez & Zidon, 1984; Locke & Latham, 2002).

Although the overall relationship between goal difficulty and task performance is positive, effects are subtler for very high levels of goal difficulty. ***Goal commitment*** plays an important role in the relationship between goal difficulty and task performance. Task performance is positively related to goal difficulty when goals are accepted and negatively related when goals are rejected. Goal commitment is problematic for very difficult goals that seem unattainable, no matter what. Because of this negative association between goal commitment and difficulty, goal setting theory predicts an inverse U-shaped relationship between goal difficulty and task performance (Erez & Zidon, 1984). Furthermore, ***goal importance*** plays a moderating role in the relationship between goal difficulty and task performance. The importance of goal attainment to individuals, including the importance of results of a task, also facilitate goal commitment (Locke & Latham, 2002). Further, important goals elicit persistent striving towards goal attainment (Miner, 2005).

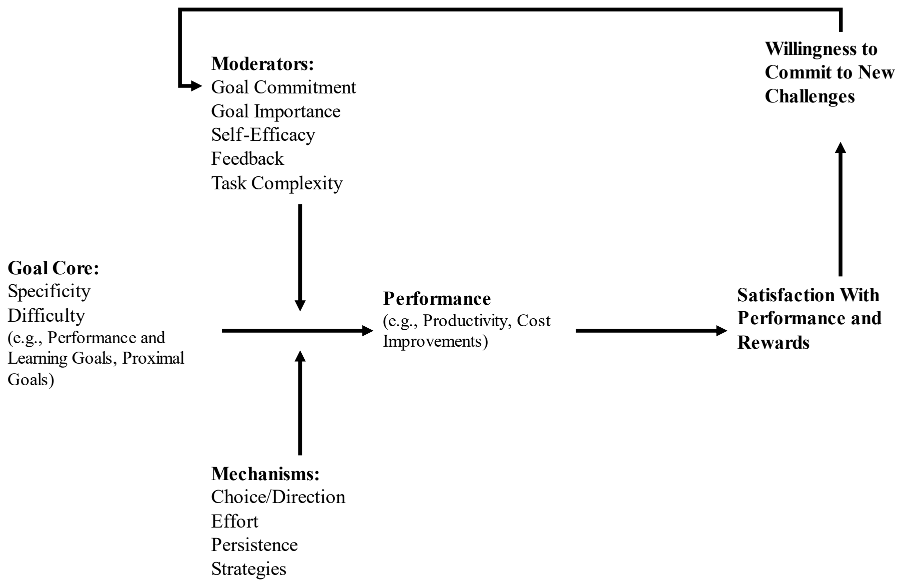


Figure 35: Important components of goal setting theory. Source: Locke and Latham (2002)

Since target costing is often done in (cross-functional) teams (Ansari et al., 2006; Dekker & Smidt, 2003; Wijewardena & Zoysa, 1999), it is also relevant to consider goal setting theory for teams. A meta-analysis of the effect of goal setting on group performance (Kleingeld et al., 2011) found a robust positive effect of specific group goals versus nonspecific goals on the performance of groups. “Goal setting thus appears to be at least as effective at the group level as at the individual level” (Kleingeld et al., 2011, p. 1294).

Thus, prior research often found a relationship between difficult individual and team goals. Yet, the literature does not provide evidence supporting this relationship in the context of target costing for the performance of development teams in actual organizations. The arguments above imply that the effectiveness of target costing arises from specific, difficult goals that motivate employees and teams in product development to achieve more cost reduction performance, leading to the following hypothesis:

H1: *Target cost difficulty and cost reduction performance during product development are positively related.*

However, studies in accounting show that the relationship between target difficulty and performance is often more complex (Bonner & Sprinkle, 2002; Feichter et al., 2018; Matějka, 2018). For example, Cheng et al. (2007) investigated goal conflict when people are assigned multiple goals, and they did not find a direct relationship between perceived overall goal difficulty and task performance, but this was mediated by goal conflict. Arnold and Artz (2015) found only a weak direct association between target difficulty and performance at the level of firm performance. But, the relationship between target difficulty and performance was negative for very difficult targets after controlling for target flexibility. Webb et al. (2013) found that higher targets can increase performance through higher effort, in the sense of working harder on known tasks. But, higher targets may also reduce performance by hindering *outside-the-box* thinking for the discovery of efficiency improvements for these tasks.

In the context of product development, too, the effectiveness of target costing is more intricate. Product development occurs in a highly complex organizational context, making it much more difficult to understand how target costing could lead to cost reduction and greater profitability of new products. For example, cost information may be incomplete and needs to be complemented with other kinds of information for managing product development activities, but this can cause information overload (Henri & Wouters, 2020). Moreover, engineers with specialized and private knowledge might overengineer their components and increase costs, because they overestimate the importance of their parts, they want to showcase their technical expertise, or avoid risks

(Mihm, 2010). Furthermore, some activities in product development sometimes occur simultaneously and iteratively, also called concurrent engineering. Some later development activities (i.e., technical product design) already start before earlier activities (i.e., concept development) have been fully completed (Gopalakrishnan et al., 2015). This means that some parts are already starting to be designed before all product specifications and features have been finalized, and consequently, later changes of these specifications or features necessitate making technical design adjustments. Compared to sequential product development, design teams receive new information more frequently and their goals change more often. Concurrent engineering greatly complicates product development activities and causes additional cognitive demands. Teams are more likely to have only incomplete task knowledge. As a result, having specific and challenging cost goals may be less effective for enhancing motivation and cost reduction performance (Hirst, 1987). Gopalakrishnan et al. (2015) found that groups with a specific cost goal have higher cost reduction performance than groups assigned a general cost goal only in the case of a sequential product development process, but not in a concurrent product development process. In this study, we investigate two complexities regarding the effectiveness of target costing, which arise from uncertainty and interdependence.

5.2.3 Effectiveness of Target Costing: Product Design Interdependence

Product design interdependence is the first complexity investigated in this study that may impact the effectiveness of target costing. We define ***product design interdependence*** in this study as the extent to which teams, when making product development decisions for their product, need to consider product development choices that other teams make for other products.[8] Firms also manage costs through coordinated product development decisions across product development projects. Examples are the deliberate use of common parts, common processes, or modular designs (A. Davila & Wouters, 2004; Degraeve et al., 2004). The focus shifts from managing the profitability of an individual product through target costing to managing the profitability of a product group (Granlund & Taipaleenmäki, 2005). The motivation is that some costs can also be affected by the combined effect of design choices from different product development projects, and these costs can hardly be addressed within the scope of the separate projects.

[8] Product design interdependence differs from interdependence in the study of target costing of Kee and Matherly (2021). That considered independence because of new products sharing production capacity.

Warehousing costs, for example, could be driven by the number of different stock-keeping units, R&D costs by the total number of different elements designed (i.e., parts, versions), purchasing transaction costs by the number of suppliers of the company, and manufacturing costs may be affected by the number of production runs. Nonetheless, the reduction of the number of different parts, production runs, or suppliers not necessarily lead to a reduction in cost, as the cost per unit may increase due to lager difficulties (Labro, 2004). Therefore, structural cost management is needed, which involves organizational design, process design, and product design, in order to generate an adequate cost structure (Anderson & Dekker, 2009). As an example, a car manufacturer may not only focus on the target costing approach of single car models, but also conducts cost management of parts and components, which are used across different car models (i.e., engine, axles, head-up displays).

Using common components or product platforms are important approaches in order to *coordinate the technical design decisions* during new product development over various products, which aim to decrease the overall cost for a group of products or increase total profit for the company. These approaches are especially important when launching many different products to the market. However, these approaches increase the interdependencies between product development projects. Development teams have fewer degrees of freedom for reducing the costs of their own product. Product design interdependence reduces the possibilities for achieving cost reduction performance in target costing.

H2: *Product design interdependence is associated with less cost reduction performance during product development.*

Product design interdependence also makes product development more difficult. Development teams need to exchange more information with other teams, they need to understand more about the development activities and product designs of other teams, and they need to find solutions for their own product design under more restrictions. Their task outcomes are more strongly influenced by the actions of others, and fully understanding such influences is difficult. As a result, employees may have less complete task knowledge (Hirst, 1987). Task knowledge refers to the knowledge about performing activities that are required for achieving particular goals, including knowing which activities to select or adjust under particular circumstances. Task knowledge is required so employees, stimulated by specific and difficult goals, can develop effective and efficient actions plans. Having less complete task knowledge hampers such cognitive activities, and some efforts might be misdirected and spent on irrelevant activities (Hirst, 1987). Incomplete task knowledge due to product design interdependence could therefore reduce the impact of goals on performance.

Empirical studies provide mixed support for the moderating effect of interdependence on the relationship between goal setting and performance. Hirst (1988) investigated how the effect of goal setting on intrinsic motivation was moderated by interdependence (this study did not investigate performance). For tasks with pooled interdependence, the assignment of specific and difficult goals raises intrinsic motivation, but for tasks with reciprocal interdependence, the assignment of specific and difficult goals reduced intrinsic motivation. Hirst and Yetton (1999), however, found that task interdependence did not moderate the effects of specific, difficult budget goals on the level of performance. Those goals increased performance, compared to *do-your-best* goals, on both low (pooled) and high (reciprocal) interdependent tasks.

In context of target costing and coordinated design decisions, product design interdependence refers to product design interdependencies between teams (or between product development employees) that each have their own cost goal for their own product. Studies in the goal setting literature mostly looked at interdependence between individual team members who have a common team goal (Kleingeld et al., 2011). The common goal may stimulate the individuals to collaborate more, but it may also stimulate free-riding behavior. For example, Aubé and Rousseau (2005) found that task interdependence moderated the relationship between the goal commitment of teams and their performance. Very few studies, however, considered task interdependence between individuals (or between teams) that do not have a common goal, which suggest that the interdependence negatively moderates the relationship between goal setting and performance. The experimental study of Saavedra et al. (1993) on the performance appraisal task showed that individual goals with reciprocal task interdependence lead to lower quantity and quality, compared to individual goals with pooled interdependent tasks or group goals with reciprocal interdependent tasks. In the tower-building task of T. R. Mitchell and Silver (1990) with high task interdependence, individual goals lead to worse results than group goals, a mix of individual goals and group goals, and no specific goals.

The interdependencies with other teams essentially increase the complexity of a team's (or individual's) task. Studies that investigated the role of task complexity are therefore indirectly related to our focus of task interdependence and incomplete task knowledge. The meta-analysis of Wood et al. (1987) supported the moderating role of task complexity. The positive performance effects of specific and difficult goals (vs. *do-your-best* goals) was greater on simple tasks than on complex tasks. Also, the positive performance effects of difficult goals (vs. moderate or easy goals) was greater on simple tasks than on complex tasks. This meta-analysis included studies at the individual level. However, the meta-analysis for teams of Kleingeld et al. (2011) did not find that task complexity would moderate the effect of specific, difficult group goals (versus *do-your-best* goals) on group performance.

These arguments imply that product design interdependence may not only constrain the possibilities for achieving cost reduction performance in target costing, but also limit the effect of specific, difficult goals on cost reduction performance, leading to the following hypothesis:

H3: *Product design interdependence moderates the relationship between target cost difficulty and cost reduction performance during product development.*

5.2.4 Effectiveness of Target Costing: Uncertainty About Target Cost Difficulty

This study also investigates another source of complexity that may impact the effectiveness of target costing, namely uncertainty about target cost difficulty. To enhance motivation and improve cost reduction performance, target costs should be difficult but doable. However, target setting is generally difficult (Feichter et al., 2018) and setting cost targets in product development at the appropriate level of difficulty will not always be possible. Uncertainty about target cost difficulty refers to lacking information at the start of the product development process for setting a cost target at such a level, that it will be difficult (but often possible, with much effort) to meet the target cost at the end of the product development process. As a result of this uncertainty, achieving the target cost sometimes turns out to be unexpectedly easy, other times, it appears to be simply impossible.

Long development lead times are one reason for this uncertainty about target difficulty. Developing a product can take years and at the start, limited information is available about products that will be competing in the market, future sales prices customers are willing to pay, and technological developments that affect product performance and costs. Target cost difficulty may also be uncertain in target costing because of supply chain structures. Reasoning from the sales prices to the allowable unit manufacturing cost is not simply a matter of subtracting the required profit margin.

Instead of one sales price, many different sales prices must be considered, if different product variations are sold at different sales prices in different countries. Instead of only the required profit margin, many margins and costs need to be considered, if products are distributed to the multiple points of sale via several links in long supply chains (Stadtherr & Wouters, 2017). Furthermore, from the sales price that *arrives* at the manufacturer, the nonmanufacturing costs are subtracted and then, the allowable unit manufacturing costs is disaggregated to major components and parts. So, there is a long and uncertain path from the future sales price of a product to today's manufacturing target cost for an individual part. For all such reasons, when cost goals are being determined, it may not be clear how difficult it will be to achieve those goals.

Studies in the literature on goal setting theory and studies in accounting have considered the uncertainty for setting goals. “Among the biggest impediments to goal setting is environmental uncertainty” (Latham, 2009, p. 170), meaning that the required information for setting goals may not be available, or become outdated due to a fast-changing environment. When the uncertainty raises, it becomes more and more difficult to be motivated by a goal and perform better (Latham, 2009). One reason for this effect is that it is difficult to adjust goals that turn out to be set at a too difficult or too easy challenging level. Adjusting goal difficulty on the basis of new information that becomes available is more difficult when goals are more challenging and specific (Polzer & Neale, 1995). Research looking at the effects of very difficult or very easy goals on motivation and performance basically found an inverted U-shaped relationship between goal difficulty and performance (Fang et al., 2005). Difficult but doable goals are associated with the highest performance, but very easy and very difficult goal are associated with lower performance (Latham & Locke, 1991). This inverted-U relationship between goal difficulty and performance occurs when individuals reduce or redirect their efforts to meet what is required of them and thereby adjust their performance to the goals. “Once a reference point (goal) is set, people view the outcome as binary, you either fail or succeed in achieving your goal. If one has no chance at succeeding, then effort decreases as it will be wasted on a task certain to fail. If one has a task that is very easy, then effort decreases once the goal is met, even though one may have been capable of greater performance” (Burdina et al., 2017, p. 78).[9]

[9] For example, it is reported that cab drivers are more likely to quit working once they reach their income goals on rainy days where they considerably earn more (Camerer et al., 1997).

We expect similar effects in the context of target costing. Specifically, teams work on projects for developing a part and they review the progressing design. The manufacturing cost is estimated, based on what is known about the current, semi-finished design. Besides costs, the design will also be evaluated in terms of the other requirements, such as product characteristics. As an example, the design of a passenger car also needs to satisfy the requirements for engine power, fuel consumption, emissions, noise, weight, safety, and aerodynamics. Target costing can also be seen from the perspective of stage-gate reviews (Hertenstein & Platt, 2000). Suppose the estimated product cost exceeds the cost target and teams believe achieving the cost target is challenging but could be possible. Stimulated by that level of goal difficulty, teams are stimulated to come up with solutions for improving the design, trying to meet all targets at a lower cost. These products may land at the target cost or above it (so still not fully meet the cost target), but *everything* has been done trying to achieve the target cost. However, suppose the product cost is already below the cost target (better than needed), or the product cost is still above the target, but teams believe it will be very easy to achieve the target. With such a low level of goal difficulty, teams are likely to slow down their cost reduction efforts. These products likely end very close to their cost target.

Thus, we expect that teams who, in the course of the progressing product development project, perceive a considerable likelihood of landing above the cost target (i.e., not achieving the target), are motivated to perform better. However, teams, who perceive a considerable likelihood of landing below the cost target (i.e., better than needed), may slow down and aim to achieve the target cost. Assuming that slowing down in front of an easy target is a stronger effect than speeding up when facing a difficult target, most products would end up with a cost that does not meet the target.

These performance enhancing and inhibiting effects of target cost attainment are modeled in **Figure 36**. At the start of product development, it is uncertain which cost levels are realistically attainable and, therefore, current costs at the start of product development (CC_0) are assumed to be normally distributed around the target cost (TC), so $CC_0 := N(100, 30)$ and $TC := 100$. Teams start their development activities, which can increase or decrease the current cost after product development (CC_s). This impact of their development activities on current costs (ΔCC) is assumed to be normally distributed $\Delta CC := N(0, 5)$ and $CC_S := CC_0 + \Delta CC$. Target cost attainment is defined as $(TC - CC_S)/TC$, which is the black line in **Figure 36**. Positive values indicate the favorable result that the target cost has been attained and costs are below the target; negative values indicate that costs are above the target and so the target has not been attained. However, the impact of SE teams on costs may not be symmetrical, due to performance enhancing and inhibiting effects. In the case of a cost overrun at the start of product development ($CC_0 > TC$), teams are stimulated to perform better and reduce the gap by a factor E. The current cost after product development is calculated as $CC_S := CC_0 - E(CC_0 - TC) + \Delta CC$, whereby $E := 0.5$ in our simulations. The higher the cost gap, the higher the absolute cost reduction performance during development becomes. In the case of underrunning the target cost at the start of product development ($CC_0 < TC$), teams slow down their performance to aim for the target cost. The current cost after product development is calculated as $CC_S := TC + \Delta CC$. As a result of both effects, the distribution of target cost attainment distribution is negatively skewed. The simulation of the performance enhancing and inhibiting effect produces a mean target cost attainment of -0.060 with a mode of -0.016. A negative skewed distribution is present when the mean is less than the mode, as is the case here.

H4: *The distribution of target cost attainment distribution after product development is negatively skewed.*

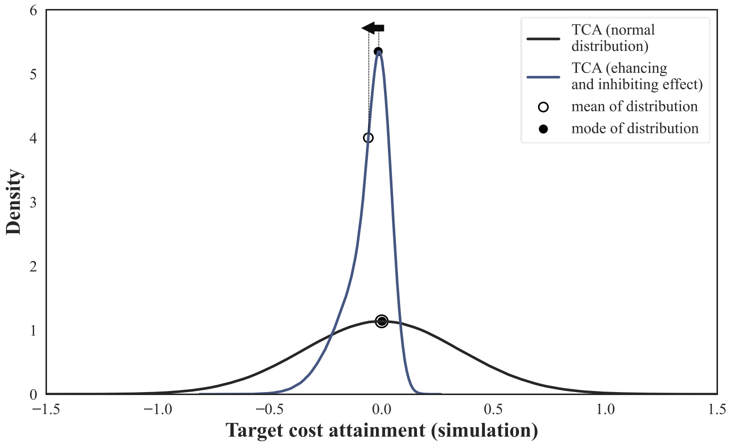


Figure 36: Simulation of the performance enhancing and inhibiting effects of target costs on target cost attainment. When current costs are higher than the target at the start of product development ($CC_0 > TC$), teams are stimulated to perform better and reduce the cost gap. When costs are below the target at the start of product development ($CC_0 < TC$), teams slow down their performance and aim for the target cost. Together, the performance enhancing and inhibiting effects produce a left-skewed distribution, which is characterized by a mean that is lower than the mode (left from the mode on the x-axis).

So, teams can do more or less during product development to find and implement solutions for improving the design and reducing the product cost. If a team needed to do less in terms of optimizing the product design, more cost reduction potential is left in the product design. The remaining cost reduction potential may, to some extent, be shifted to the production stage of the product. We define ***cost reduction performance during production*** (*CRPP*) as how much a team is able to reduce the cost of their product after product development, so during the production stage of the product. Less cost reduction is possible during production when more cost reduction has already been achieved during development. These arguments imply a tradeoff between achieving cost reduction performance during product development or during production, leading to the following hypothesis:

H5: *Cost reduction performance during the product production stage is negatively related to cost reduction performance during the development stage.*

The extent to which this effect occurs, may depend on how easy or difficult it was to achieve the cost target during product development. We define ***target cost attainment*** as the extent to which a team was able to achieve a product design with a cost that meets the target cost. A team fearing to not achieve the target cost experiences more pressure, compared to a team that feels secure about achieving the target. The greater pressure could lead a team to take less sustainable cost reduction measures that negatively affect cost reduction possibilities during production. For example, the team may reduce the quality of the product, leading to product design changes during production, or it could select cheaper suppliers that later cause more problems during production. As a result, the negative impact of cost reduction during development on cost reduction during production could be stronger in the case of low target cost attainment. If there is less pressure, a team not only needs to reduce costs less during development, but it may also do this in ways that less negatively affect subsequent cost reduction during production. They can develop a product that can later continue to be further improved, also in terms of cost, without having to take actions during production that partly undo and correct earlier cost reduction measures that were implemented during product development. The less the team is able to meet the cost target at the end of product development, the greater it probably experienced pressure and stress during product development. Thus, target cost attainment would moderate the relationship between cost reduction during development and during production in the following way: The more difficult it is to achieve target costs during product development, i.e., less target cost attainment, the stronger the subsequent cost reduction performance during the production stage will be affected, leading to the following hypothesis:

H6: *The negative relationship between cost reduction performance during the production and product development stages is moderated by target cost attainment.*

5.2.5 The Premise of Target Costing

The premise of target costing, as mentioned above, is that cost reduction is more feasible during product development than during production (Dowlatshahi, 1992; Keys, 1990). During product development, many decisions are being made that largely determine the product cost, such as the characteristics of the product, the underlying architecture of the product, the detailed technical design, selection of technologies, technical standards, components, materials, and suppliers (Krishnan & Ulrich, 2001). Obviously, these decisions influence each other. The choice of a particular technology, for example, has implications for the kinds of components that will be used, which focusses the choice of suppliers. Changing one element has consequences for many other. The solution space is small, however, because many restrictions apply, for example, concerning technical feasibility, legislation, technical standards, and customer preferences. Still, more degrees of freedom for making decisions exist during product development than during the production stage.

As an illustrative example: To save weight, space, and costs in a battery-electric vehicle, a new alloy steel for some axles and other parts of the drive train needs to be developed; moreover, manufacturing these parts with this new alloy steel requires particular process conditions (i.e., temperatures), for which the production equipment needs to be adjusted. Developing new parts and changing the production equipment might be feasible during the development stage of a car, but changing the design of the car, the parts, and the production process is less feasible during the production stage. These arguments imply that possibilities for reducing costs that are not actually implemented during product development may not always be transferrable to the production stage, leading to the following hypothesis:

H7: *Achieving relatively more cost reduction performance during development than during production is associated with greater total cost reduction performance.*

5.3 Research Method

5.3.1 Research Site

The study has been done on the basis of proprietary archival data on product development projects at a car company. This was a suitable research site for several reasons. First, because the company had been using target costing to manage costs during new product development since many years. Second, because this company kept an extensive database on product costs and other relevant variables throughout the development and production stages. The company used a pre-calculation system for tracking the estimated current cost and target cost of parts during product development, and a post-calculation system for tracking the actual costs during production. Third, because we had access to new product development teams to gain insight into the target costing process, the meaning of the archival data, and thereby the possibility to identify valid variable measures. New product development was performed in SE teams, including members form the technical development, procurement, sales, production/logistics, controlling, and quality departments. The SE teams were accountable for meeting the target costs for their specific parts during development and manage costs reduction during the production stage. SE teams were provided with feedback about the current cost gap by bi-monthly cost reports. Due to an increasing competitive pressure and disruptive changes in the automotive industry, the company experienced difficulties meeting their cost goals at the time the research was conducted. The researcher participated in SE teams and team meetings of the product controlling department.

New product development takes about five years in this company. The early phase of new product development involves the ideation and conceptualization of new car models. About three years before start of production, the so-called *serial development* starts, and product development is primarily performed by SE teams. At the beginning of the serial development phase, the target cost and current cost are calculated at the part level and are recorded in the pre-calculation system. After start of production, the post-calculation system is deployed until the end of production of the car model.

5.3.2 Data

We obtained data for several thousands of parts that were related to development projects for twelve different vehicle models. For each vehicle development project, the parts in the bill of materials were detailed in the pre-calculation system. These parts range from small parts (i.e., screws) to entire components (i.e., headlights) and to pre-assembled systems (i.e., front axle system). For each part, the target costs and current cost are updated continuously and reported bi-monthly. Target costs (*TC*) specify the cost goals of parts based on the sales price and the target profit margin. Current costs (*CC*) are the estimated cost for the current technical solution. For many parts, we could collect cost data at three points in time: target cost and current cost at start of serial development (t_0), target cost and current cost at the start of production (t_S) and current cost 12 months after SOP (t_{S+12}). These data points are illustrated in **Figure 37**. The cost data during production came from the post-calculation system, which contained data about the direct material costs of parts after the start of production. The so-called complex bill of materials of a car model contains all configurable parts, i.e., 20 different steering wheels or 6 different versions of headlights. In addition, we could collect several descriptive variables for each part, such as the responsible SE team and carry-over information (hat or platform part).

For some parts, not all these data points were available, either because at time of this study, the part had not reached 12 months after SOP, or because historical data were not comparable. Due to a modification of the pre-calculation system, the historical bills of materials can only be accessed until a certain point in time. If the start of serial development of a car model was prior this date, we used the costing data that was firstly available. The maximum offset between t_0 and the availability of historical bills of materials over the 12 car models amounts to 5 months.

Overall, the bills of materials of the twelve car models included 8,254 individual parts at t_0 and 9,527 parts at t_S, of which 3,708 parts were specified at both t_0 and t_S. Thus, many parts were added, replaced, or discarded during serial development. We discarded parts with incomplete information of relevant variables and dropped outliers according to the z-score and Cook's distance, resulting in 3,611 parts. In the following, we focus on parts where the current costs exceed the target costs at t_0, resulting in a sample of 1,827 parts. For this sample, the mean target cost at the start of serial development (*TC* at t_0) was €30.93 with a large standard deviation of €193.47. The mean current cost the same moment (*CC* at t_0) was €35.60 with a standard deviation of €218.94. Finally, the 1,827 parts from the pre-calculation system were merged with the respective one-year-after-SOP complex bills of materials, resulting in 597 parts with data on the current costs at t=0, t=SOP and t=SOP+12.[10]

[10] For two car models, the complex bills of materials were not available.

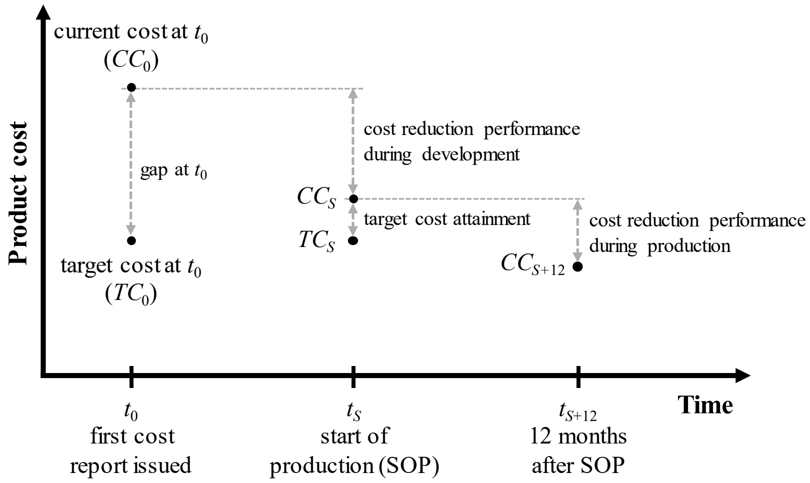


Figure 37: Current product cost, target product cost, cost reduction performance during the development stage (from t_0 to t_s) and cost reduction performance during the production stage (from t_s to t_{s+12})

5.3.3 Variable Measurement

Our dependent variables concern the cost reduction performance of parts achieved during development and during production. *Cost reduction performance during development (CRPD)* is measured as the relative reduction of current cost between t_0 and t_S. A positive *CRPD* value indicates a *positive performance*, so a cost reduction; a negative value indicates a cost increase. Descriptive statistics for all variables are provided in **Table 34**.

$$CRPD = \frac{CC_0 - CC_S}{CC_0} \tag{27}$$

where

CC_0: Current cost, $t = 0$
CC_S: Current cost, $t =$ SOP

Cost reduction performance during production (CRPP) is measured as the relative current cost reduction during the first year after start of production (t_S to t_{S+12}). A positive *CRPP* value indicates a reduction of cost during the first year in production.

$$CRPP = \frac{CC_S - CC_{S+12}}{CC_S} \tag{28}$$

where

CC_{S+12}: Current cost, t = SOP+12

Total cost reduction performance (*CRPT*) refers to both the development and production stage, and this is measured as the relative reduction of current cost between t_0 and t_{S+12}.

$$CRPT = \frac{CC_0 - CC_{S+12}}{CC_0} \tag{29}$$

The independent variables and control variables in this study are target cost difficulty, product design interdependence, target cost attainment, goal importance, and task uncertainty. We measure *target cost difficulty* by the relative cost gap (GAP_0) between the current cost and target cost at t_0. A higher cost gap indicates a more difficult cost goal for SE teams at the start of the serial development phase.

$$GAP_0 = \frac{CC_0 - TC_0}{TC_0} \tag{30}$$

Table 34: Descriptive statistics of the dependent and independent variables

Statistic	**Mean**	**Standard deviation**	**Median**	**Observations**
CRPD (cost reduction performance during development)	0.003	0.300	0.000	1,827
CRPP (cost reduction performance during production)	0.040	0.357	0.030	597
CRPT (total cost reduction performance)	0.031	0.447	0.069	597
GAP_0 (target cost difficulty)	0.219	0.328	0.120	1,827
IF (number of influencing product features)	1.885	1.052	2	597
TCA (target cost attainment)	-0.174	0.348	-0.099	1,827
CRPDoP (cost reduction performance during development over production)	-0.065	1.230	0.061	584
$\ln(relTC_0)$ (goal importance)	0.002	0.011	0.000	1,827
CTC (task uncertainty)	0.034	0.296	0.000	1,827

To measure *product design interdependence*, the distinction between hat parts and platform parts is relevant. The so-called hat concerns the upper part of the vehicle and determines its visual appearance, such as height, number of doors, and silhouette. It consists of an inside structure and visible parts such as side frames, roof, hood, fender, doors, trunk lid, and bumpers. The platform is the low part of the car where the powertrain, chassis subassemblies, and seats are connected. Not much of this is visible from the outside. Hat parts are used for one vehicle model only, while platform parts are shared between multiple vehicle models. Therefore, product design interdependence is measured at two levels: high for platform parts and low for hat parts. Of the 1,827 individual parts, 667 (36.5%) are platform parts and 1,160 (63.5%) are hat parts.

Additionally, we measure product design interdependencies by the number of influencing product features (*IF*). Some parts have multiple versions and which version to use depends on other parts in the configuration of the car. The *IF* number of a part is the number of features of other parts that determine the feasibility of using a particular part. The *IF* number is lower if the part is less dependent on the configuration of a vehicle, so if the part is not affected much by other features of a vehicle. For example, suppose the version of the head-up display depends on whether the vehicle is a left- or right-hand drive, the dashboard material, and the type of windshield. In that case, the *IF* number of a head-up display would be three, because three features need to be checked to select the appropriate version of this part. It also means that the SE team responsible for the head-up displays needs to coordinate with three other SE teams. The number of different head-up displays that are needed could be much more than three, because the three *IF* could create many constrains that necessitate a large number of different versions of the head-up display. As another example, the tow eye is an extreme case: Its design is independent from the rest of the vehicle and so the *IF* number is 0. Therefore, there is only one version of it. Data on *IF* were available for 597 parts in the post-calculation system, and of these, 442 (74.0%) parts are dependent on other features ($IF > 0$) and 155 (26.0%) are independent ($IF = 0$).[11]

P_{platf} := 1 if Platform part, 0 Hat part.

IF := number of influencing product features

We measure *target cost attainment* (*TCA*) as the relative deviation between target cost and current costs after product development, so at t_S. A positive *TCA* value indicates a *positive* (i.e., a favorable) result, meaning that the current costs are below target costs, a negative value indicates a target cost overrun.

$$TCA = \frac{TC_S - CC_S}{TC_S} \tag{31}$$

where

TC_S: Target cost, t = SOP

Cost reduction performance during development versus production is measured by comparing both cost reduction performances, after z-standardization $\left(\frac{X-\mu}{\sigma}\right)$, *Cost reduction performance during development over production* (*CRPDoP*) is calculated by the difference between both standardized cost reduction variables. A positive value indicates that relatively more cost reduction is realized during development compared to

[11] The number of influencing features is not available in the pre-calculation system and can only be retrieved from the corresponding complex bills of materials.

cost reduction during production. 13 outliers have been dropped according to the 3-sigma rule.

$$CRPDoP = \quad Z(CRPD) - Z(CRPP) \tag{32}$$

where

Z: z-standardization

We control for the effects of goal importance and task uncertainty. The measurement of *goal importance* is based on the target cost of a part relative to the total target cost of a car in t = 0 ($relTC_0$). We expect that parts with a higher impact on total cost will have higher perceived goal importance. The mean relative target cost amounts to 0.002 (SD 0.011). However, the distribution of part value is highly skewed to the right; the target cost in t_0 ranges from €0.003 to €3,802, while 95% of the parts are below a TC of €88.98. We expect that the difference of TC between €1 and €100 has much higher impact on perceived importance, than the difference between €3,001 and €3,100. To account for the skewness of part values, we apply logistic transformation.

$$\ln(relTC_0) = \quad \ln\left(\frac{TC_0}{\sum_{car} TC_0}\right) \tag{33}$$

Finally, *task uncertainty* during new product development arises in the company because the target cost of a part can change during development when substantial product changes are implemented. Such changes increasingly take place in case of high uncertainty about the technology, costumer, and competitor environment. Accordingly, we measure task uncertainty by the *change of target cost during development* (*CTC*). The change of target cost is calculated by the relative change of target cost between t_0 and t_S. From the 1,827 individual parts, 739 (40.45 %) parts changed their target cost during development.

$$CTC = \frac{TC_S - TC_0}{TC_0} \tag{34}$$

5.4 Results

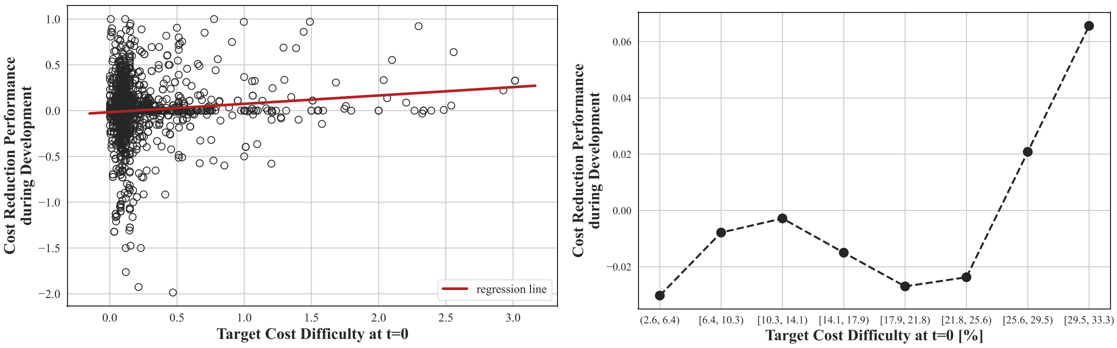
5.4.1 Cost Reduction Performance and Target Cost Difficulty (H1)

We first analyze the relationship between target cost difficulty and cost reduction performance during development (H1). **Figure 38a** shows the scatter plot and linear regression line between both variables, indicating a positive relationship. We conduct several multivariate OLS regression models to statistically test this effect, see **Table 35**.[12] Model 1 includes target cost difficulty (GAP_0), task uncertainty (CTC) and goal importance ($\ln(relTC_0)$) and is based on all observations (3,611 parts). Model 2 is based on more relevant observations, namely parts that require cost reduction during development (cost gaps at t_0: current costs above target costs) (1,827 parts). The positive and statistically significant coefficient for target cost difficulty (GAP_0) supports H1.

As we use goal setting theory as our theoretical framework, we additionally investigate the premise of the inverted-U relationship between goal difficulty and task performance to provide further robustness on the applicability of goal setting theory in the context of target costing. As we expect that the inverted U-shape relationship between target cost difficulty and *CPRD* is primarily in place for intermediate levels of target cost difficulty, we discard observations with very easy goals less than 2.6% GAP_0 (5th percentile) and very hard goals higher than 33.3% GAP_0 (85th percentile), resulting in a subset of 1,459 parts (80% of observations). **Figure 38b** depicts the mean *CRPD* over eight equally sized bins of GAP_0 to further bring out the relationship between both variables. The binned scatter plot shows an inverted-U-shape association for the cost gap from 2.6% to 21.8%. Model 3 statistically analyzes the inverted-U relationship for the 1,326 parts (72.58%) in the range of 2.6% and 21.8% GAP_0. It shows a significantly negative second-degree effect of cost gap at t_0, confirming the inverted-U association for intermediate goals. An explanation for the positive relationship beyond 21.8% GAP_0 could be mounting cost pressure. Such large target cost gaps create a *high-pressure* area

[12] For each OLS regression model (Model 1 – Model 11), we further perform residual analysis, where we do not find any abnormalities.

for SE teams, with additional task forces to find cost reduction opportunities and additional top management reporting. The higher the target cost gap, the larger the pressure to achieve cost reduction performance. However, our results below suggest that these measures may have unfavorable consequences later, during the production stage of the product.



a) *Scatter plot and regression line (1,827 observations). The regression line suggests a positive relationship between target cost difficulty and CRPD.*

b) *Binned scatter plot for intermediate levels of target cost difficulty (1,459 obverstions). The binned scatter plot indicates an inverse relationship between target goal diffcuity and CRPD for observations with a GAP_0 between 2.6% and 21.8%.*

Figure 38: Relationship between target cost difficulty (GAP_0) at the start of product development (t_0) and cost reduction performance during development (*CRPD*).

Table 35: The effect of target cost difficulty on cost reduction performance during development (*CRPD*)

	Cost reduction performance during development (CRPD)		
	Model 1	*Model 2*	*Model 3*
Constant	0.0033 (0.224)	**0.0509*** (2.115)	-0.0594 (-1.505)
GAP_0 (target cost difficulty)	**0.2133***** (16.942)	**0.1993***** (12.010)	**2.0100**** (3.014)
$(GAP_0)^2$			**-7.5760**** (-2.706)
CTC (task uncertainty)	**-0.7649***** (-59.017)	**-0.6513***** (-35.340)	**-0.8745***** (-37.030)
$\ln(relTC_0)$ (goal importance)	0.0016 (0.981)	**0.0079**** (2.946)	**0.0060*** (1.976)
Observations	3611	1827	1326
R^2	49.8%	42.0%	51.7%

Regression coefficients (and t-statistic)
+p<0.1; ∗p<0.05; ∗∗p<0.01; ∗∗∗p<0.001

5.4.2 Cost Reduction Performance and Product Design Interdependence (H2 and H3)

Next, we examine the relationship between product design interdependence and cost reduction performance during development. **Figure 39** depicts the scatter plot between the number of influencing features (*IF*) and *CRPD* which indicates a slightly negative relationship between *IF* and *CRPD*: More product design interdependence seems to be associated with less cost reduction during development. **Figure 40** compares the mean *CRPD* values for hat and platform parts which shows that the mean *CRPD* is lower (so less cost reduction) for platform parts than for hat parts. Both figures suggest a negative association between product design interdependence and *CRPD*.

Figure 41 depicts the scatter plots between target cost difficulty and *CRPD* for high and low interdependencies for both variable measurements. **Figure 41a** indicates a strong moderating effect of product design interdependence measured by the difference between hat parts and platform parts. **Figure 41b** shows no interaction effect of product design interdependence measured by the number of influencing product features.

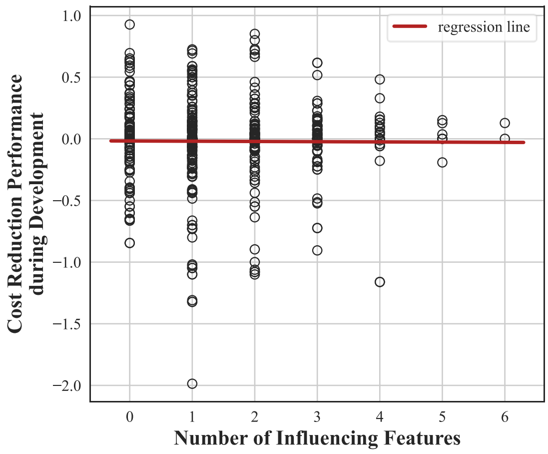


Figure 39: Scatter plot between the number of influencing features (*IF*) and *CRPD*. The regression line suggests a slightly negative relationship (597 observations).

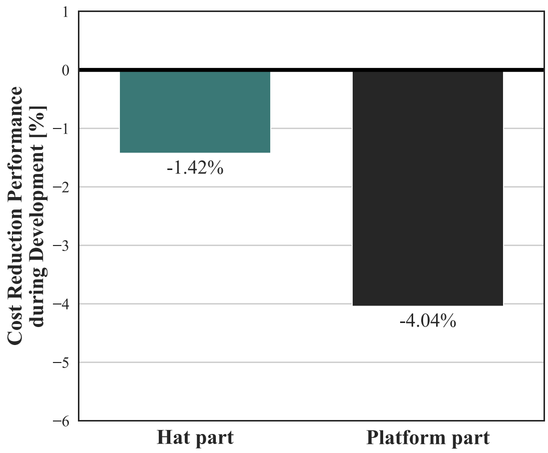


Figure 40: Comparison of the mean *CRPD* between hat and platform parts. The negative number indicates that, on average, costs increased during development. This effect was less for hat parts. Thus, cost reduction performance was better for hat parts, which have a lower product design interdependence.

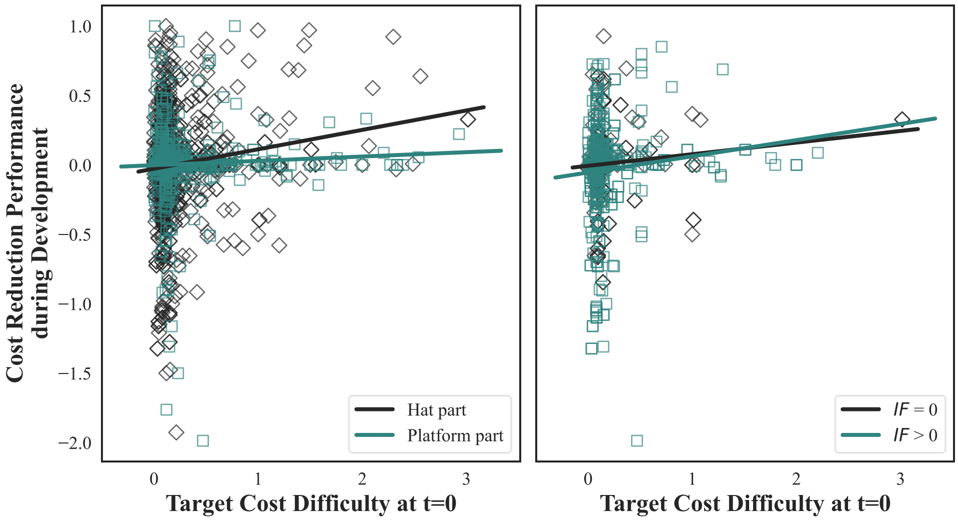


Figure 41: Scatter plot between target cost difficulty at t_0 (GAP_0) and *CRPD*, differentiated between products with high and low product design interdependencies. Product design interdependence is measured on the basis of hat parts versus platform parts and on the basis of the *IF* number. The relationship between GAP_0 and *CRPD* seems to be different for hat parts versus platform parts (moderation effect of product design interdependence). This effect is not visible on the basis of the *IF* number.

Table 36 shows the results of several OLS regression models. Model 4 investigates the main effect of product design interdependence measured on the basis of the distinction between hat parts and platform parts (P_{platf}), and Model 5 investigates the main effect of product design interdependence measured on the basis of the number of influencing features (*IF*). The negative and statistically significant coefficients for both measures of product design interdependence support H2 (although the result in Model 5 is only marginally significant). Model 6 and Model 7 additionally include the interaction effect of product design interdependence and target cost difficulty (GAP_0). Model 6 shows a statistically significant interaction effects of product design interdependence and target cost difficulty, supporting H3 that target cost difficulty moderates the relationship of product design interdependence and cost reduction performance during development. The results for Model 7 do not support H3.

Table 36: The effect of product design interdependence on cost reduction performance during development (*CRPD*)

	Cost reduction performance during development (CRPD)			
	Model 4	*Model 5*	*Model 6*	*Model 7*
Constant	**0.0587*** (2.409)	0.0980 (1.514)	0.0118 (1.767)	**-0.0208*** (-2.088)
GAP_0 (target cost difficulty)	**0.2015***** (12.123)	**0.3234***** (9.893)	**0.2611***** (11.804)	**0.3169***** (9.083)
CTC (task uncertainty)	**-0.6533***** (-35.424)	**-0.6893***** (-22.795)	**-0.6554***** (-35.672)	**-0.6906***** (-22.750)
$\ln(relTC_0)$ (goal importance)	**0.0079**** (2.955)	**0.0123**$^+$ (1.888)	**0.0075**** (2.794)	**0.0121**$^+$ (1.852)
P_{platf} (platform part)	**-0.0220*** (-1.979)		**-0.0211**$^+$ (-1.906)	
$GAP_0 \times P_{platf}$			**-0.1328***** (-4.061)	
IF (number of influencing product features)		**-0.0157**$^+$ (-1.850)		**-0.0162**$^+$ (-1.895)
$GAP_0 \times IF$				-0.0171 (-0.534)
Observations	1827	597	1827	597
R^2	42.1%	48.5%	42.6%	48.5%

Regression coefficients (and t-statistic)
In Model 6 and Model 7 the independent and moderating variables are mean-centered, except for the binary variable P_{platf}.
+p<0.1; ∗p<0.05; ∗∗p<0.01; ∗∗∗p<0.001

5.4.3 Cost Reduction Performance and Uncertainty About Target Cost Difficulty (H4-H6)

Because of uncertainty about target cost difficulty, achieving the target cost sometimes appears to be difficult, other times it turns out be very easy. According to H4, teams react to such outcomes and their cost reduction performance is enhanced or restricted, leading to a left skewed distribution of *TCA* at start of production (**Figure 36**). **Figure 42** shows the distribution of target cost attainment and demonstrates the expected negative (left) skewness. The mean target cost attainment amounts to -0.174 and the mode amounts to -0.030, supporting H4. This suggests that cost reduction performance is, indeed, reduced, when target costs are attained during development.

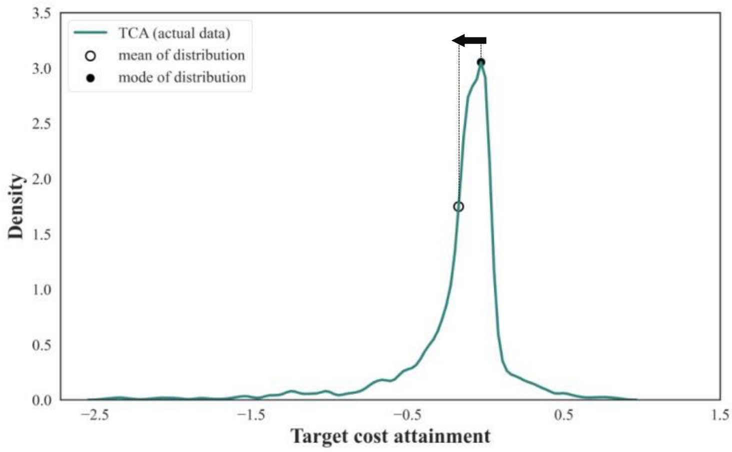


Figure 42: Target cost attainment (*TCA*) after product development (1,827 observations). The mean *TCA* is lower (left on the x-axis) than the mode, which indicates a left-skewed distribution. The shape of the actual distribution of *TCA* is similar to the simulated distribution of *TCA* based on the performance enhancing and inhibiting effects in **Figure 36**.

Next, we analyze the trade-off between cost reduction performance during the development and production stages, and the moderating effect of target cost attainment on this relationship. The association between *CRPD* and *CRPP* is depicted in **Figure 43**. The scatter plot indicates a negative relationship between both variables. Moreover, the mean *CRPP* is 0.123 in case no cost reduction is realized during development (*CRPD* = 0), which is greater than the mean value of *CRPP* of 0.024 when cost reduction during development has taken place (*CRPD* ≠ 0) (*t*-Test, *p*-value = 0.011). In order to specifically analyze the linear relationship between *CRPD* and *CRPP*, we drop the observations with no cost reduction during development (*CRPD* = 0), since observations with *CRPD* = 0 would obscure the linear effect.

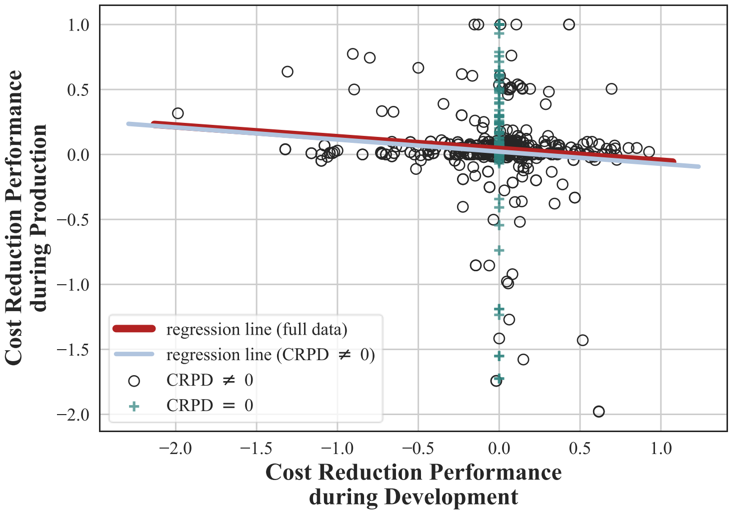


Figure 43: Scatter plot between cost reduction performance during *development* and cost reduction performance during *production*. The red regression line, which is based on all 597 observations, suggests that *CRPD* has a negative impact on *CRPD*. The more costs are reduced during development, the poorer the cost reduction performance during production. Parts with very high *CRPD* are even connected to negative *CRPP* values, which correspond to an increase in cost. The blue line is based on observations where *CRPD* is unequal to zero (497 observations), so where parts increase or decrease the current cost during development, and is comparable to the red line.

To analyze the inverse relationship between *CRPD* and *CRPP* we again conducted OLS regression analysis, shown in **Table 37**. We applied z-standardization to ensure comparability of both cost reduction periods. The negative, statistically significant coefficient of *CRPD* in Model 8 supports H5. Cost reduction during production is less, as more cost reduction has taken place during development. Model 9 includes the interaction effect of *CRPD* and target cost attainment. The statistically significant coefficient supports H6: The negative relationship between *CRPD* and *CRPP* depends on *TCA* during product development. The interaction effect is visualized in **Figure 44**. The regression lines are plotted for high and low values of target cost attainment based on Model 9 where a high (low) *TCA* value is indicated by 1 standard deviation above (below) the mean score. In the case of high *TCA*, the current cost after product development (at t_s) is at the target cost or better than the target cost, and the negative association between *CRPD* and *CRPP* is less strong. High cost reduction during development still allows cost reduction during production. However, if *TCA* is low, the current cost after product development is above the target cost (the cost goal is not achieved), and the negative association between *CRPD* and *CRPP* is much stronger. High cost reduction during development even leads to negative *CRPP* (unfavorable, so cost increases).

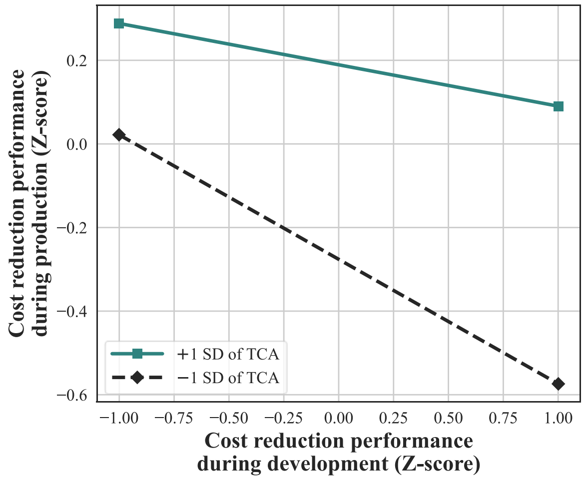


Figure 44: Moderating effect of target cost attainment on the relationship between cost reduction performance during *development* and cost reduction performance during *production*. The figure suggests that the negative relationship between *CRPD* and *CRPP* is moderated by target cost attainment (*TCA*). In the case of high *TCA* (+1 standard deviation (SD) of *TCA*, green line) the negative impact of *CRPD* on *CRPP* is less strong (flatter line). High cost reduction performance during development still allows for high cost reduction performance during production (positive *CRPP*). In the case of low *TCA* (-1 SD of *TCA*, black line) the negative effect of *CRPD* on *CRPP* is much stronger (steeper line). The high cost reduction performance during development in combination with low target cost attainment even leads to negative *CRPP* values, which correspond to an increase in cost during production. The figure is based on Model 9.

Table 37: The effect of cost reduction performance during development (*CRPD*) on cost reduction performance during production (*CRPP*).

	Cost reduction performance during production (CRPP)	
	Model 8	*Model 9*
Constant	0 (0)	-0.0434 (-0.950)
CRPD (cost reduction performance during development)	**-0.1143*** (-2.560)	**-0.1983***** (-4.094)
TCA (target cost attainment)		**0.2324***** (4.678)
CRPD×TCA		**0.0993**** (3.186)
Observations	497	497
R^2	1.3%	6.4%

Regression coefficients (and t-statistic)
+p<0.1; ∗p<0.05; ∗∗p<0.01; ∗∗∗p<0.001
All variables are standardized
Model 8 and Model 9 are without observations where *CRPD* = 0

5.4.4 The Effectiveness of Cost Reduction During Development and Production Stage (H7)

Finally, we examine the effectiveness of cost reduction during development relative to cost reduction during the production stage. The achievement of *CRPD* relative to *CRPP* is measured by cost reduction performance during development over production (*CRPDoP*). **Figure 45** exhibits the scatter plot between *CRPDoP* and total cost reduction performance (*CRPT*). The plot indicates a positive association between both variables. We separately illustrate parts where no cost reduction is realized during development (*CRPD* = 0). For these parts, *CRPT* corresponds to *CRPP* and *CRPDoP* corresponds to negative *CRPP*, which leads to the strong negative association in the scatter plot. The data set does not contain any observations where *CRPP* equals to zero.

In order to analyze the effect of *CRPDoP* on *CRPT*, we again conduct OLS regression analysis (see **Table 38**). Model 10 exhibits a significantly positive effect of *CRPDoP* on total cost reduction performance. The more cost reduction is realized during development relative to cost reduction during production, the higher the total cost reduction. To further bring out the linear association for parts where the costs changed in both periods, we discard observations where *CRPD* = 0 (Model 11). Model 11 exhibits a much larger regression coefficient and coefficient of determination. All in all, these results provide support for hypothesis 7.

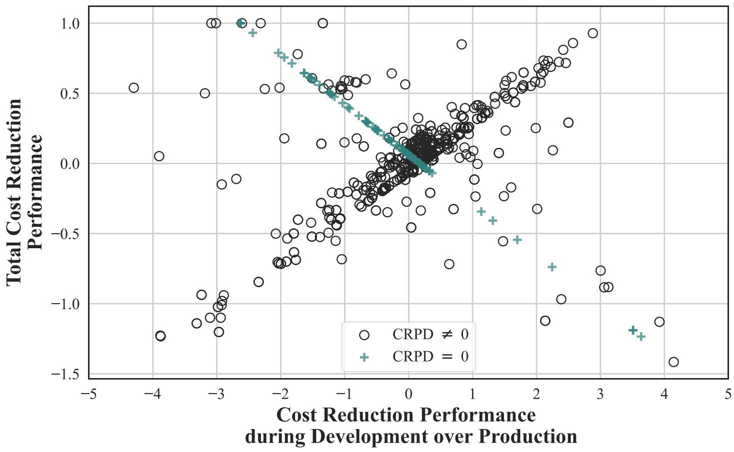


Figure 45: Scatter plot between cost reduction performance during development over production (*CRPDoP*) and total cost reduction performance (*CRPT*) for 584 observations. 96 parts, where the *CRPD* equals zero, are illustrated separately. There are no observations where *CRPP* equals zero. The scatter plot suggests a positive relationship between *CRPDoP* and *CRPT*. The more cost reduction is placed on the development phase relative to the production phase (higher positive *CRPDoP* values), the better the total cost reduction performance.

Table 38: The effect of cost reduction performance during development over production (*CRPDoP*) on the total cost reduction performance (*CRPT*).

	Total cost reduction performance (CRPT)	
	Model 10	*Model 11*
Const.	**0.0560***** (3.337)	0.0247 (1.497)
CRPDoP (cost reduction performance during development over production)	**0.0528***** (3.875)	**0.1362***** (10.140)
Observations	584	488
R^2	2.5%	17.5%

Regression coefficients (and t-statistic)
+p<0.1; ∗p<0.05; ∗∗p<0.01; ∗∗∗p<0.001
Model 11 is without observations where CRPD = 0

5.5 Discussion and Conclusion

This study adds to our understanding of the effectiveness of target costing in a complex organizational setting with uncertainty about target difficulty and complex supply chains and product structures. Several studies showed that specific, difficult goals were associated with more cost reduction performance than *do-your-best* goals in target costing (Everaert & Bruggeman, 2002; Gopalakrishnan et al., 2015). These studies used experiments with individual tasks, and as a first contribution, this study complement prior research by providing evidence on the relationship between goal difficulty and cost reduction performance on the basis of proprietary, archival company data. These data include cost targets and cost estimates at several points in time for several hundreds of parts. We find that more difficult cost goals are associated with more cost reduction performance during product development, which also supports the overall idea of target costing.

Second, we contribute to research on the effectiveness of target costing in a more complex product development context. Prior studies considered complexities due to the need to manage the costs of shared resources during product development (A. Davila & Wouters, 2004). Implicit incentives for product development employees (Mihm, 2010), information overload from combining cost information and nonfinancial performance measures (Henri & Wouters, 2020), or the complexities of concurrent engineering (Gopalakrishnan et al., 2015). We show that two further complex circumstances in product development activities may limit the effectiveness of target costing: product design interdependence and uncertainty about target cost difficulty.

Product design interdependence makes product development activities more complex by limiting a team's degrees of freedom for finding and implementing cost reduction possibilities, and by requiring more coordination with other teams. Accordingly, we find a negative relationship between product design interdependence and cost reduction performance during development. We also find that product design interdependence moderates the relationship between target cost difficulty and cost reduction performance during development. This result suggests it is worthwhile to look for approaches in which target costing could integrated with cost management strategies such as parts commonality that create interdependencies (Labro, 2004). Such approaches may have implications for how the costs of common elements are allocated to separate target-costing projects (Israelsen & Jørgensen, 2011).

Uncertainty about target cost difficulty implies that employees, in the course of a progressing product development project, may perceive a considerable likelihood of landing below the cost target (i.e., do better than needed). They may slow down and aim for reaching the cost target instead of achieving maximum cost reduction performance. Consistent with such responses, we find a negatively skewed distribution of target cost attainment. When some cost reduction performance is not needed during development, it can be shifted to the production stage. Accordingly, we find that cost reduction during product development and during production are negatively related, and target cost attainment moderates that relationship. These findings imply a tradeoff between achieving cost reduction performance during product development or during production, and the easier it is to achieve target costs during product development, the more cost reduction performance will be shifted to the production stage. Notably, this study uniquely combined data on cost reduction during the development and the production stage. Moreover, we find that by shifting cost reduction from the development stage to the production stage, total cost reduction performance is less. Thus, uncertainty about target cost difficult can have significant consequences for cost reduction performance, and it may limit the effectiveness of target costing.

These results provide several ideas for future research that is connected to broader topics in accounting. The phenomenon of constraining cost reduction performance if the target cost can easily be met, has some similarities with the ***target ratchet effect***. Target ratcheting refers to managers adjusting targets in the next period based on the actual results in a prior period. Anticipating target ratcheting, employees may limit their effort and performance, aiming to meet but not exceed performance targets, for example, in order to avoid that performance targets in the next period are adjusted upwardly, which is the target ratchet effect (Aranda et al., 2014). For example, Bouwens and Kroos (2011) find support for such behavior of managers reducing their effort, but the results of Webb et al. (2010) are not consistent with employees constraining their performance because of ratcheting concerns. Future research in target costing could investigate variables that have also been investigated in ratcheting studies, and which may also influence the performance constraining behavior in target costing, such as the presence of implicit agreements between managers and superiors that target adjustments based on the manager's past performance are restricted (Bol & Lill, 2015) or the availability of relative performance information for target setting (Casas-Arce et al., 2018).

Future research could also focus on adjustments of cost targets, in response to uncertainty of target cost difficulty. Research in accounting has investigated ***ex post target adjustments*** that are made during the period, or at the end of the period, when new information about goal difficulty becomes available (Kelly et al., 2015). Such adjustments are typically done through subjective adjustments (Bol & Smith, 2011; Gibbs et al., 2004; Höppe & Moers, 2011). Target adjustments can have beneficial performance effects, such as through procedural justice perceptions (Kelly et al., 2015). However, anticipation of adjustments may also weaken the incentive effect of targets and reduce performance (Arnold & Artz, 2015). Our study demonstrated the problem of too easy targets in target costing, which would imply *upward* target adjustments. These are unusual in prior studies (Arnold & Artz, 2015; Kelly et al., 2015). Subjective adjustments for uncontrollable events are made to compensate for bad luck, but not to punish for good luck (Bol & Smith, 2011). Future research in target costing could focus on upward target adjustments when cost targets turn out to be too easy. How could such adjustments be implemented? How would people react to that in terms of fairness perceptions? Would it be possible to stimulate teams to focus less on target cost attainment and more on cost reduction performance, encouraging teams to potentially achieve costs below the target costs, if that turns out to be possible. This would be related to research that has looked at rewards that not flatten completely beyond target achievement (Merchant et al., 2018). At the same time, other goals are important in product development, such as time-to-market and customer needs (T. Davila, 2000). A greater focus on cost reduction would make it even more important to closely monitor that other key goals are still achieved (Booker et al., 2007).

On the other hand, how problematic is our finding that too easy targets are likely to make teams constrain their performance and just achieve their cost target (but not do better)? Research in accounting has also investigated more nuanced ideas about target attainment and slack. Target achievement makes results more predictable, which is important for meeting outside expectations and for coordination within the firm (T. Davila & Wouters, 2005; Merchant & Manzoni, 1989). Constraining effort and performance in order to maximize target cost attainment instead of absolute performance may also have favorable effects for the organization. Future research could also try to better understand when and why target cost attainment might be more important than maximum cost reduction performance. Future research could specifically investigate the complex coordination problems, which could result from not meeting or beating target costs for many parts.

6 Conclusion

This dissertation thesis aimed to provide a better understanding on the management of cost during new product development in a complex and uncertain environment. Specifically, we investigate the usage of new technologies, such as big data and machine learning, and the application of target costing. Two research foci are examined: 1) *the applicability of big data and machine learning technology for cost management during new product development* and 2) *the impact of complexity and uncertainty on the effectiveness of the target costing approach*. In a three-year research project in cooperation with a car manufacturer, we developed four studies: a literature review, a case study, an experimental study, and a proprietary archival study. In the following, we briefly summarize our main contributions, discuss limitations, and suggest future research opportunities.

In the literature review study, we examined use cases, benefits, and issues of big data and machine learning technology in management accounting. First, we contribute to the literature by providing an overview of various use cases of both technologies in managerial accounting. We find that the body of literature is rather small, however, all main big data aspects (volume, velocity, variety, and veracity) and various machine learning methods can be applied with every main task of management accounting. Further, we find that big data and machine learning are mostly used for different tasks. Big data is primarily used for descriptive analyses and machine learning is mainly used for predictive tasks. Hence, we find that the simultaneous usage of big data and machine learning is uncommon in the field of management accounting. Second, we categorize benefits and issues of both technologies in the context of management accounting. Our systematic review finds that poor data quality and the lack in skills are the most critical issues of big data applications. For machine learning the interpretability problem and the complex training process are considered problematic. Big data offers opportunities by providing new insights, better decision making, and increasing the influential power of accountants. In regard with the application of machine learning, we find that higher accuracy, time-specific advantages, and larger independence from expert knowledge are the main benefits. Moreover, we find that the benefits and issues are highly dependent on the specific management accounting task.

In the third chapter, we investigate the applicability of machine learning and big data technology for cost estimation from a multi-generational perspective. We provided empirical evidence for the practicality of these technologies in cost management at a car manufacturer on the basis of six research aspects. First, we find that the complex machine learning models, such as GBR and ANN, only yield superior cost estimation accuracy for the estimation of subsequent product generations. Within the same generation, much simpler models, such as LAR, yield similar predictive accuracy levels. Second, we add to cost estimation literature by demonstrating the positive impact of big data application on predictive accuracy. Third, we propose a novel method incorporating the target price of a product to improve cost estimates for multi-generational cost estimation. Fourth, machine learning and big data technology was found to be more accurate than manual calculations from cost experts for total product costs. Yet, machine learning and big data performed less well on more granular cost levels. Finally, the case study shows that machine learning techniques are able capture the most important cost drivers (fifth aspect) and quantify the average costs of features (sixth aspect) reliably in most cases. We also find that machine learning models are most adequate in the product design and early development phase of new product development when there are rapid changes of product designs, low integration of expert knowledge, and cost estimates are not used for cost goal setting. Otherwise, more detailed and comprehensible cost estimates are required.

The experimental study investigates several factors that influence the relative importance of explainability to accuracy in the context of product cost estimation during new product development. First we empirically confirm that people actually perceive the trade-off between accuracy and explainability. Second, we provide evidence that the relative importance of explainability to accuracy matters for the selection of machine learning methods for cost estimation. Third, we contribute to the literature by identifying four factors that significantly influence the relative importance of explainability to accuracy for product cost estimation during new product development. Specifically, we introduce the *phase of product development*, *information uncertainty*, the *level of cost granularity*, and *target cost gap* as important antecedents of relative importance of explainability to accuracy. We show that despite having much higher accuracy, complex machine learning methods are considered adequate in only few situations during new product development. Complex machine learning models are more appropriate during the early development stage, when dealing with low information uncertainty, high level of cost granularity, and a low target cost gap.

In the proprietary archival study, we uniquely combine data on cost reduction during the development and production stage in order to analyze the effectivity of target costing in complex situations with high product interdependencies and uncertainty. First, we complement the literature by providing empirical evidence toward the positive impact of target cost difficulty on cost reduction performance during product development. Second, we show that interdependencies between parts and products can limit the effectiveness of cost reduction performance during new product development. Third, we find that uncertainty about cost goal difficulty moderates the allocation of cost reduction performance during development and the production stage. If teams perceive a considerable likelihood of landing below the cost target, they may slow down and don't aim to achieve the maximum cost reduction performance. Thereby, uncertainty about target cost difficulty leads to a shift of cost reduction performance from the development phase to the production stage, which limits the total cost reduction performance.

This dissertation thesis is limited by certain characteristics of the longitudinal case study research approach and factors related to the specific research methods in the four studies. The study-specific limitations were already discussed in each of the four main chapters. In the following, we point out the main limitations related to case study research approach. In a single-case design the findings potentially underlie biased qualitative data collection, subjective interpretations, and are neither statistically generalizable nor reproducible (Cooper & Slagmulder, 2004b). Therefore, the conclusions from the cases are not automatically transferable to other settings. Yet, this is not the target of this work. The objective was to provide a detailed and multifaceted analysis toward the usage of big data and machine learning technology and target costing when dealing with high complexity and uncertainty in product development. Notwithstanding its limitations, this dissertation makes several contributions to the literature by providing a detailed and thorough analysis of big data and machine learning and target costing in cost management.

The studies trigger several opportunities for future research. First, we find that the body of literature lacks field research, which reports on experienced implementations from the perspective of management accountants. Conducting more field work and case studies on the application of big data and machine learning at real companies, could generate further insights into this nascent topic. Second, it would be fruitful to investigate the operationalization of big data and machine learning technology in cost management during product development, such as the establishment of human-machine interfaces for cost estimation models. Third, future studies could conduct additional research on the model selection problem with regard to the accuracy-explainability trade-off. It would be interesting to analyze antecedents of the relative importance of explainability to accuracy for other activities in cost management. Fourth, future research could analyze the adjustments of cost targets, in response to the uncertainty of target cost difficulty. Future research in target costing could also investigate variables that have been investigated in ratcheting studies, which may also influence the performance constraining behavior in target costing.

7 References

Abadi, M., Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, . . . Xiaoqiang Zheng. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*. http://download.tensorflow.org/paper/whitepaper2015.pdf

ABDC. (2019, December 6). *ABDC Journal Quality List.* Australian Business Deans Council. https://abdc.edu.au/research/abdc-journal-quality-list/

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Afonso, P., Nunes, M., Paisana, A., & Braga, A. (2008). The influence of time-to-market and target costing in the new product development success. *International Journal of Production Economics*, *115*(2), 559–568. https://doi.org/10.1016/j.ijpe.2008.07.003

Agarwal, R., & Prasad, J. (1998). The antecedents and consequents of user perceptions in information technology adoption. *Decision Support Systems*, *22*(1), 15–29. https://doi.org/10.1016/S0167-9236(97)00006-7

Ajzen, I. (2002). Residual Effects of Past on Later Behavior: Habituation and Reasoned Action Perspectives. *Personality and Social Psychology Review*, *6*(2), 107–122. https://doi.org/10.1207/S15327957PSPR0602_02

Allaire, Y., & Firsirotu, M. E. (1989). Coping with strategic uncertainty. *MIT Sloan Management Review*, *30*(3). https://www.proquest.com/scholarly-journals/coping-with-strategic-uncertainty/docview/224962763/se-2?accountid=44866

Alonso, J. M., Castiello, C., & Mencar, C. (2015). Interpretability of Fuzzy Systems: Current Research Trends and Prospects. In J. Kacprzyk & W. Pedrycz (Eds.), *Springer Handbook of Computational Intelligence* (pp. 219–237). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-43505-2_14

Alonso, J. M., & Magdalena, L. (2011). HILK++: an interpretability-guided fuzzy modeling methodology for learning readable and comprehensible fuzzy rule-based classifiers. *Soft Computing*, *15*(10), 1959–1980. https://doi.org/10.1007/s00500-010-0628-5

Al-Qady, M., & El-Helbawy, S. (2016). Integrating Target Costing and Resource Consumption Accounting. *Journal of Applied Management Accounting Research*, *14*(1).

Amani, F. A., & Fadlalla, A. M. (2017). Data mining applications in accounting: A review of the literature and organizing framework. *International Journal of Accounting Information Systems*, *24*, 32–58. https://doi.org/10.1016/j.accinf.2016.12.004

Anderson, S. W., & Dekker, H. C. (2009). Strategic Cost Management in Supply Chains, Part 1: Structural Cost Management. *Accounting Horizons*, *23*(2), 201–220. https://doi.org/10.2308/acch.2009.23.2.201

Ansari, S., Bell, J., & Okano, H. (2006). Target Costing: Uncharted Research Territory. In *Handbooks of Management Accounting Research* (Vol. 2, pp. 507–530). Elsevier. https://doi.org/10.1016/S1751-3243(06)02002-5

Appelbaum, D., Kogan, A., Vasarhelyi, M., & Yan, Z. (2017). Impact of business analytics and enterprise systems on managerial accounting. *International Journal of Accounting Information Systems*, *25*(International Journal of Accounting Information Systems), 29–44. https://doi.org/10.1016/j.accinf.2017.03.003

Aranda, C., Arellano, J., & Davila, A. (2014). Ratcheting and the Role of Relative Target Setting. *The Accounting Review*, *89*(4), 1197–1226. https://doi.org/10.2308/accr-50733

Arnaboldi, M., Busco, C., & Cuganesan, S. (2017). Accounting, accountability, social media and big data: revolution or hype? *Accounting, Auditing & Accountability Journal*, *30*(4), 762–776. https://doi.org/10.1108/AAAJ-03-2017-2880

Arnold, M. C., & Artz, M. (2015). Target difficulty, target flexibility, and firm performance: Evidence from business units' targets. *Accounting, Organizations and Society*, *40*, 61–77. https://doi.org/10.1016/j.aos.2014.12.002

Aubé, C., & Rousseau, V. (2005). Team Goal Commitment and Team Effectiveness: The Role of Task Interdependence and Supportive Behaviors. *Group Dynamics: Theory, Research, and Practice*, *9*(3), 189–204. https://doi.org/10.1037/1089-2699.9.3.189

Aunurrafiq, Sari, R. N., & Basri, Y. M. (2015). The Moderating Effect of Goal Setting on Performance Measurement System-managerial Performance Relationship. *Procedia Economics and Finance*, *31*, 876–884. https://doi.org/10.1016/S2212-5671(15)01185-5

Ax, C., Greve, J., & Nilsson, U. (2008). The impact of competition and uncertainty on the adoption of target costing. *International Journal of Production Economics*, *115*(1), 92–103. https://doi.org/10.1016/j.ijpe.2008.04.010

Bansal, K., Vadhavkar, S., & Gupta, A. (1998). Neural networks based data mining applications for medical inventory problems. *International Journal of Agile Manufacturing*, *1*(2).

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Baryannis, G., Dani, S., & Antoniou, G. (2019). Predicting supply chain risks using machine learning: The trade-off between performance and interpretability. *Future Generation Computer Systems*, *101*, 993–1004. https://doi.org/10.1016/j.future.2019.07.059

Bendul, J., & Apostu, V. (2017). An Accuracy Investigation of Product Cost Estimation in Automotive Die Manufacturing. *International Journal of Business Administration*, *8*(7), 1. https://doi.org/10.5430/ijba.v8n7p1

Berchicci, L., & Bodewes, W. (2005). Bridging environmental issues with new product development. *Business Strategy and the Environment*, *14*(5), 272–285. https://doi.org/10.1002/bse.488

Berkovsky, S., Taib, R., & Conway, D. (2017). How to Recommend? In G. A. Papadopoulos, T. Kuflik, F. Chen, C. Duarte, & W.-T. Fu (Eds.), *Proceedings of the 22nd International Conference on Intelligent User Interfaces - IUI '17* (pp. 287–300). ACM Press. https://doi.org/10.1145/3025171.3025209

Bhimani, A., & Willcocks, L. (2014). Digitisation, ‘Big Data’ and the transformation of accounting information. *Accounting and Business Research*, *44*(4), 469–490. https://doi.org/10.1080/00014788.2014.910051

Bode, J. (2000). Neural networks for cost estimation: Simulations and pilot application. *International Journal of Production Research*, *38*(6), 1231–1254. https://doi.org/10.1080/002075400188825

Boehm, M., Kumar, A., & Yang, J. (2019). Data Management in Machine Learning Systems. *Synthesis Lectures on Data Management*, *14*(1), 1–173. https://doi.org/10.2200/S00895ED1V01Y201901DTM057

Bol, J. C., & Lill, J. B. (2015). Performance Target Revisions in Incentive Contracts: Do Information and Trust Reduce Ratcheting and the Ratchet Effect? *The Accounting Review*, *90*(5), 1755–1778. https://doi.org/10.2308/accr-51050

Bol, J. C., & Smith, S. D. (2011). Spillover Effects in Subjective Performance Evaluation: Bias and the Asymmetric Influence of Controllability. *The Accounting Review*, *86*(4), 1213–1230. https://doi.org/10.2308/accr-10038

Bonner, S. E., & Sprinkle, G. B. (2002). The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. *Accounting, Organizations and Society*, *27*(4-5), 303–345. https://doi.org/10.1016/S0361-3682(01)00052-6

Booker, D. M., Drake, A. R., & Heitger, D. L. (2007). New Product Development: How Cost Information Precision Affects Designer Focus and Behavior in a Multiple Objective Setting. *Behavioral Research in Accounting*, *19*(1), 19–41. https://doi.org/10.2308/bria.2007.19.1.19

Bouwens, J., & Kroos, P. (2011). Target ratcheting and effort reduction. *Journal of Accounting and Economics*, *51*(1-2), 171–185. https://doi.org/10.1016/j.jacceco.2010.07.002

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Chapman & Hall/CRC.

Brown, S. L., & Eisenhardt, K. M. (1995). Product Development: Past Research, Present Findings, and Future Directions. *Academy of Management Review*, *20*(2), 343. https://doi.org/10.2307/258850

Brownlee, J. (2019, August 12). *A Tour of Machine Learning Algorithms*. https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

Burdina, M., Hiller, R. S., & Metz, N. E. (2017). Goal attainability and performance: Evidence from Boston marathon qualifying standards. *Journal of Economic Psychology*, *58*, 77–88. https://doi.org/10.1016/j.joep.2017.01.001

Burton-Jones, A., & Hubona, G. S. (2006). The mediation of external variables in the technology acceptance model. *Information & Management*, *43*(6), 706–717. https://doi.org/10.1016/j.im.2006.03.007

Cai, X., & Tyagi, S. (2014). Development of a Product Life-Cycle Cost Estimation Model to Support Engineering Decision-Making in a Multi-Generational Product Development Environment. *Journal of Cost Analysis and Parametrics*, *7*(3), 219–235. https://doi.org/10.1080/1941658X.2014.982403

Calude, C. S., & Longo, G. (2017). The Deluge of Spurious Correlations in Big Data. *Foundations of Science*, *22*(3), 595–612. https://doi.org/10.1007/s10699-016-9489-4

Campbell, D. J., & Ilgen, D. R. (1976). Additive effects of task difficulty and goal setting on subsequent task performance. *Journal of Applied Psychology*, *61*(3), 319–324. https://doi.org/10.1037/0021-9010.61.3.319

Campolongo, F., & Braddock, R. (1999). The use of graph theory in the sensitivity analysis of the model output: a second order screening method. *Reliability Engineering & System Safety*, *64*(1), 1–12. https://doi.org/10.1016/S0951-8320(98)00008-8

Caputo, A. C., & Pelagagge, P. M. (2008). Parametric and neural methods for cost estimation of process vessels. *International Journal of Production Economics*, *112*(2), 934–954. https://doi.org/10.1016/j.ijpe.2007.08.002

Carlton Sapp, Daren Brabham, Joseph Antelmi, Henry Cook, Thornton Craig, Soyeb Barot, Doreen Galli, Sumit Pal, Sanjeev Mohan, & George Gilbert. (2019). *2019 Planning Guide for Data and Analytics.* Gartner. https://www.gartner.com/en/doc/361501-2019-planning-guide-for-data-and-analytics

Casas-Arce, P., Holzhacker, M., Mahlendorf, M. D., & Matějka, M. (2018). Relative Performance Evaluation and the Ratchet Effect. *Contemporary Accounting Research*, *35*(4), 1702–1731. https://doi.org/10.1111/1911-3846.12385

Cavalieri, S., Maccarrone, P., & Pinto, R. (2004). Parametric vs. neural network models for the estimation of production costs: A case study in the automotive industry. *International Journal of Production Economics*, *91*(2), 165–177. https://doi.org/10.1016/j.ijpe.2003.08.005

CGMA. (2013, October 16). *From Insight to Impact Unlocking Opportunities in Big Data.* Chartered Global Management Accountants. https://www.cgma.org/resources/reports/insight-to-impact-big-data.html

CGMA. (2016, February 2). *Joining the dots: Decision making for a new era.* Chartered Global Management Accountants. https://www.cgma.org/resources/joining-the-dots-decision-making-for-a-new-era.html

Chan, S. L., Lu, Y., & Wang, Y [Yan] (2018). Data-driven cost estimation for additive manufacturing in cybermanufacturing. *Journal of Manufacturing Systems, 46*, 115–126. https://doi.org/10.1016/j.jmsy.2017.12.001

Chen, D., & Burrell, P. (2001). Case-Based Reasoning System and Artificial Neural Networks: A Review. *Neural Computing & Applications*, *10*(3), 264–276. https://doi.org/10.1007/PL00009897

Cheng, M. M., Luckett, P. F., & Mahama, H. (2007). Effect of perceived conflict among multiple performance goals and goal difficulty on task performance. *Accounting & Finance*, *47*(2), 221–242. https://doi.org/10.1111/j.1467-629X.2007.00215.x

Chollet, F., & others. (2015). *Keras*. https://keras.io

Chou, J.-S., Tai, Y., & Chang, L.-J. (2010). Predicting the development cost of TFT-LCD manufacturing equipment with artificial intelligence models. *International Journal of Production Economics*, *128*(1), 339–350. https://doi.org/10.1016/j.ijpe.2010.07.031

Chou, J.-S., & Tsai, C.-F. (2012). Preliminary cost estimates for thin-film transistor liquid–crystal display inspection and repair equipment: A hybrid hierarchical approach. *Computers & Industrial Engineering*, *62*(2), 661–669. https://doi.org/10.1016/j.cie.2011.11.037

Chua, F. (2013). *Big Data: Its Power and Perils.* Association of Chartered Certified Accountants, ACCA. https://www.accaglobal.com/content/dam/acca/global/PDF-technical/futures/pol-afa-bdpap.pdf

Cokins, G. (2002). Integrating target costing and ABC. *Journal of Cost Management*, *16*(4), 13-22.

Cooper, R., & Slagmulder, R. (1999). Develop profitable new products with target costing. *MIT Sloan Management Review*, *40*(4), 23–33.

Cooper, R., & Slagmulder, R. (2004a). Achieving full-cycle cost management. *MIT Sloan Management Review*, *46*(1), 45-52.

Cooper, R., & Slagmulder, R. (2004b). Interorganizational cost management and relational context. *Accounting, Organizations and Society*, *29*(1), 1–26. https://doi.org/10.1016/S0361-3682(03)00020-5

Coussement, K., Benoit, D. F., & Antioco, M. (2015). A Bayesian approach for incorporating expert opinions into decision support systems: A case study of online consumer-satisfaction detection. *Decision Support Systems*, *79*, 24–32. https://doi.org/10.1016/j.dss.2015.07.006

Coussement, K., Lessmann, S., & Verstraeten, G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decision Support Systems*, *95*, 27–36. https://doi.org/10.1016/j.dss.2016.11.007

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Curran, R., Gilmour, M., McAlleenan, C., Kelly, P., & Wang, J [J.] (2007). Development of an Evolutionary Cost Management System for Dynamic Product Definition. In *7th AIAA ATIO Conf, 2nd CEIAT Int'l Conf on Innov and Integr in Aero Sciences,17th LTA Systems Tech Conf; followed by 2nd TEOS Forum.* American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2007-7707

Davila, A., Foster, G., & Oyon, D. (2009). Accounting and Control, Entrepreneurship and Innovation: Venturing into New Research Opportunities. *European Accounting Review*, *18*(2), 281–311. https://doi.org/10.1080/09638180902731455

Davila, A., & Wouters, M. (2004). Designing Cost-Competitive Technology Products through Cost Management. *Accounting Horizons*, *18*(1), 13–26. https://doi.org/10.2308/acch.2004.18.1.13

Davila, T. (2000). An empirical study on the drivers of management control systems' design in new product development. *Accounting, Organizations and Society, 25*(4-5), 383–409. https://doi.org/10.1016/S0361-3682(99)00034-3

Davila, T., & Wouters, M. (2005). Managing budget emphasis through the explicit design of conditional budgetary slack. *Accounting, Organizations and Society, 30*(7-8), 587–608. https://doi.org/10.1016/j.aos.2004.07.002

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, *35*(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982

Degraeve, Z., Labro, E., & Roodhooft, F. (2004). Total cost of ownership purchasing of a service: The case of airline selection at Alcatel Bell. *European Journal of Operational Research*, *156*(1), 23–40. https://doi.org/10.1016/j.ejor.2003.08.002

Dekker, H., & Smidt, P. (2003). A survey of the adoption and use of target costing in Dutch firms. *International Journal of Production Economics, 84*(3), 293–305. https://doi.org/10.1016/S0925-5273(02)00450-4

Demoulin, N. T., & Coussement, K. (2020). Acceptance of text-mining systems: The signaling role of information quality. *Information & Management*, *57*(1), 103120. https://doi.org/10.1016/j.im.2018.10.006

Deng, S., & Yeh, T.-H. (2010). Applying least squares support vector machines to the airframe wing-box structural design cost estimation. *Expert Systems with Applications*, *37*(12), 8417–8423. https://doi.org/10.1016/j.eswa.2010.05.038

Deng, S., & Yeh, T.-H. (2011). Using least squares support vector machines for the airframe structures manufacturing cost estimation. *International Journal of Production Economics*, *131*(2), 701–708. https://doi.org/10.1016/j.ijpe.2011.02.019

Desai, V. (1997). Credit-scoring models in the credit-union environment using neural networks and genetic algorithms. *IMA Journal of Management Mathematics*, *8*(4), 323–346. https://doi.org/10.1093/imaman/8.4.323

Diaz, G. I., Fokoue-Nkoutche, A., Nannicini, G., & Samulowitz, H. (2017). An effective algorithm for hyperparameter optimization of neural networks. *IBM Journal of Research and Development*, *61*(4/5), 9:1-9:11. https://doi.org/10.1147/JRD.2017.2709578

Dishaw, M. T., & Strong, D. M. (1998). Supporting software maintenance with software engineering tools: A Computed task–technology fit analysis. *Journal of Systems and Software*, *44*(2), 107–120. https://doi.org/10.1016/S0164-1212(98)10048-1

Dishaw, M. T., & Strong, D. M. (1999). Extending the technology acceptance model with task–technology fit constructs. *Information & Management*, *36*(1), 9–21. https://doi.org/10.1016/S0378-7206(98)00101-3

Doğan, S. Z., Arditi, D., & Murat Günaydin, H. (2008). Using Decision Trees for Determining Attribute Weights in a Case-Based Model of Early Cost Prediction. *Journal of Construction Engineering and Management*, *134*(2), 146–152. https://doi.org/10.1061/(ASCE)0733-9364(2008)134:2(146)

Doshi-Velez, F., & Kim, B. (2017). *Towards A Rigorous Science of Interpretable Machine Learning*. http://arxiv.org/pdf/1702.08608v2

Dowlatshahi, S. (1992). Product design in a concurrent engineering environment: an optimization approach. *International Journal of Production Research*, *30*(8), 1803–1818. https://doi.org/10.1080/00207549208948123

Duverlie, P., & Castelain, J. M. (1999). Cost Estimation During Design Step: Parametric Method versus Case Based Reasoning Method. *The International Journal of Advanced Manufacturing Technology*, *15*(12), 895–906. https://doi.org/10.1007/s001700050147

Dysert, L. (2007). Is estimate accuracy an oxymoron? A publication of the american association of cost engineers. *Cost Engineering*, *49*(1), 32–36.

Earley, C. E. (2015). Data analytics in auditing: Opportunities and challenges. *Business Horizons*, *58*(5), 493–500. https://doi.org/10.1016/j.bushor.2015.05.002

Economist Intelligence Unit, E. I. (2013, April 29). *The Data Directive: Focus on the CFO*. https://eiuperspectives.economist.com/technology-innovation/data-directive/article/data-directive-focus-cfo-summary-paper

Eisenhardt, K. M., & Tabrizi, B. N. (1995). Accelerating Adaptive Processes: Product Innovation in the Global Computer Industry. *Administrative Science Quarterly*, *40*(1), 84. https://doi.org/10.2307/2393701

Elish, M. C., & Boyd, D. (2018). Situating methods in the magic of Big Data and AI. *Communication Monographs*, *85*(1), 57–80. https://doi.org/10.1080/03637751.2017.1375130

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *The Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Ellram, L. M. (2006). The Implementation of Target Costing in the United States: Theory Versus Practice. *The Journal of Supply Chain Management*, *42*(1), 13–26. https://doi.org/10.1111/j.1745-493X.2006.04201003.x

Elsevier B.V. (2017). *Scopus Content Coverage Guide*. https://www.elsevier.com/?a=69451

Elsevier B.V. (2019). *Scopus Advanced search*. https://www.scopus.com

Emmert-Streib, F., Yli-Harja, O., & Dehmer, M. (2020). Explainable artificial intelligence and machine learning: A reality rooted perspective. *WIREs Data Mining and Knowledge Discovery*, *10*(6). https://doi.org/10.1002/widm.1368

Erez, M., & Zidon, I. (1984). Effect of goal acceptance on the relationship of goal difficulty to performance. *Journal of Applied Psychology*, *69*(1), 69–78. https://doi.org/10.1037/0021-9010.69.1.69

Everaert, P., Boër, G., & Bruggeman, W. (2000). The impact of target costing on cost, quality and development time of new products: conflicting evidence from lab experiments. *Tijdschrift Voor Economie En Management*, *45*(3), 283–302.

Everaert, P., & Bruggeman, W. (2002). Cost targets and time pressure during new product development. *International Journal of Operations & Production Management*, *22*(12), 1339–1353. https://doi.org/10.1108/01443570210452039

Everaert, P., Loosveld, S., van Acker, T., Schollier, M., & Sarens, G. (2006). Characteristics of target costing: theoretical and field study perspectives. *Qualitative Research in Accounting & Management*, *3*(3), 236–263. https://doi.org/10.1108/11766090610705425

Everaert, P., & Swenson, D. W. (2014). Truck Redesign Case: Simulating the Target Costing Process in a Product Design Environment. *Issues in Accounting Education*, *29*(1), 61–85. https://doi.org/10.2308/iace-50623

Ewert, R., & Ernst, C. (1999). Target costing, co-ordination and strategic cost management. *European Accounting Review*, *8*(1), 23–49. https://doi.org/10.1080/096381899336131

Fang, E., Evans, K. R., & Zou, S. (2005). The moderating effect of goal-setting characteristics on the sales control systems–job performance relationship. *Journal of Business Research*, *58*(9), 1214–1222. https://doi.org/10.1016/j.jbusres.2004.03.006

Farineau, T., Rabenasolo, B., Castelain, J. M., Meyer, Y., & Duverlie, P. (2001). Use of Parametric Models in an Economic Evaluation Step During the Design Phase. *The International Journal of Advanced Manufacturing Technology*, *17*(2), 79–86. https://doi.org/10.1007/s001700170195

Feichter, C., Grabner, I., & Moers, F. (2018). Target Setting in Multi-Divisional Firms: State of the Art and Avenues for Future Research. *Journal of Management Accounting Research*, *30*(3), 29–54. https://doi.org/10.2308/jmar-52158

Filomena, T. P., Neto, F. J. K., & Duffey, M. R. (2009). Target costing operationalization during product development: Model and application. *International Journal of Production Economics*, *118*(2), 398–409. https://doi.org/10.1016/j.ijpe.2008.12.007

Fischer, R. (2017, June 2). *Logistikkennzahlen: Effizienz in Beschaffung, Lager und Versand: Fertigungstiefe und Anteil der Materialkosten an den Gesamtkosten.* Haufe Finance Office Premium Online. https://www.haufe.de/finance/haufe-finance-office-premium/logistikkennzahlen-effizienz-in-beschaffung-lager-und-31-fertigungstiefe-und-anteil-der-materialkosten-an-den-gesamtkosten_idesk_PI20354_HI7397802.html

Florén, H., Frishammar, J., Parida, V., & Wincent, J. (2018). Critical success factors in early new product development: a review and a conceptual model. *International Entrepreneurship and Management Journal*, *14*(2), 411–427. https://doi.org/10.1007/s11365-017-0458-3

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How ‘big data’ can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, *165*, 234–246. https://doi.org/10.1016/j.ijpe.2014.12.031

Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *European Conference on Computational Learning Theory*, *55*(1), 119–139. https://doi.org/10.1006/jcss.1997.1504

Frias-Martinez, E., Magoulas, G., Chen, S., & Macredie, R. (2005). Modeling human behavior in user-adaptive systems: Recent advances using soft computing techniques. *Expert Systems with Applications*, *29*(2), 320–329. https://doi.org/10.1016/j.eswa.2005.04.005

Friedman, J. H. (1997). Data mining and statistics: What's the connection. In *Proceedings of the 29th Symposium on the Interface Between Computer Science and Statistics*.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Friedman, J. H., & Meulman, J. J. (2003). Multiple additive regression trees with application in epidemiology. *Statistics in Medicine*, *22*(9), 1365–1381. https://doi.org/10.1002/sim.1501

Friedman, J. H., & Popescu, B. E. (2008). Predictive learning via rule ensembles. *The Annals of Applied Statistics*, *2*(3), 916–954. https://doi.org/10.1214/07-AOAS148

Gärtner, B., & Hiebl, M. R. W. (2017). Issues with Big Data. In M. Quinn & E. Strauss (Eds.), *The Routledge Companion to Accounting Information Systems* (pp. 161–172). Routledge. https://doi.org/10.4324/9781315647210-13

Gedikli, F., Jannach, D., & Ge, M. (2014). How should I explain? A comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies*, *72*(4), 367–382. https://doi.org/10.1016/j.ijhcs.2013.12.007

Germani, M., Mandolini, M., & Cicconi, P. (2011). Manufacturing cost estimation during early phases of machine design. In *Proceedings of the 18th International Conference on Engineering Design (ICED 11): Impacting Society through Engineering Design*, Lyngby/Copenhagen, Denmark.

Gibbs, M., Merchant, K. A., van der Stede, W. A., & Vargus, M. E. (2004). Determinants and Effects of Subjectivity in Incentives. *The Accounting Review*, *79*(2), 409–436. https://doi.org/10.2308/accr.2004.79.2.409

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). *Explaining Explanations: An Overview of Interpretability of Machine Learning*. http://arxiv.org/pdf/1806.00069v3

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, *106*(494), 746–762.

Golmohammadi, D. (2011). Neural network application for fuzzy multi-criteria decision making problems. *International Journal of Production Economics*, *131*(2), 490–504. https://doi.org/10.1016/j.ijpe.2011.01.015

González, J., Rojas, I., Pomares, H., Herrera, L. J., Guillén, A., Palomares, J. M., & Rojas, F. (2007). Improving the accuracy while preserving the interpretability of fuzzy function approximators by means of multi-objective evolutionary algorithms. *International Journal of Approximate Reasoning*, *44*(1), 32–44. https://doi.org/10.1016/j.ijar.2006.02.006

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. http://www.deeplearningbook.org/

Gopalakrishnan, M., Libby, T., Samuels, J. A., & Swenson, D. (2015). The effect of cost goal specificity and new product development process on cost reduction performance. *Accounting, Organizations and Society*, *42*, 1–11. https://doi.org/10.1016/j.aos.2015.01.003

Gorry, G. A., & Scott Morton, M. S. (1989). A framework for management information systems. *MIT Sloan Management Review*, *30*(3), 49–61.

Gotthardt, M., Koivulaakso, D., Paksoy, O., Saramo, C., Martikainen, M., & Lehner, O. (2020). Current state and challenges in the implementation of smart robotic process automation in accounting and auditing. *ACRN Journal of Finance and Risk Perspectives*, *9*(1). https://doi.org/10.35944/jofrp.2020.9.1.

Granlund, M., & Taipaleenmäki, J. (2005). Management control and controllership in new economy firms—a life cycle perspective. *Management Accounting Research*, *16*(1), 21–57. https://doi.org/10.1016/j.mar.2004.09.003

Green, S., McKinney Jr, E., Heppard, K., & Garcia, L. (2018). Big Data, digital demand and decision-making. *International Journal of Accounting & Information Management*, *26*(4), 541–555. https://doi.org/10.1108/IJAIM-02-2017-0019

Grolinger, K., Hayes, M., Higashino, W. A., L'Heureux, A., Allison, D. S., & Capretz, M. A. (2014, June 27 – July 2). Challenges for MapReduce in Big Data. In *2014 IEEE World Congress on Services* (pp. 182–189). IEEE. https://doi.org/10.1109/SERVICES.2014.41

Gruss, R., Abrahams, A. S., Fan, W., & Wang, G. A. (2018). By the numbers: The magic of numerical intelligence in text analytic systems. *Decision Support Systems*, *113*, 86–98. https://doi.org/10.1016/j.dss.2018.07.004

Gudem, M., Steinert, M., & Welo, T. (2014). From lean product development to lean innovation: Searching for a more valid approach for promoting utilitarian and emotional value. *International Journal of Innovation and Technology Management*, *11*(02), 1450008. https://doi.org/10.1142/S0219877014500084

Guia, J. D. de, Concepcion II, R. S., Calinao, H. A., Tobias, R. R., Dadios, E. P., & Bandala, A. A. (2020, December 1–3). Solar Irradiance Prediction Based on Weather Patterns Using Bagging-Based Ensemble Learners with Principal Component Analysis. In *2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC)* (pp. 1–6). IEEE. https://doi.org/10.1109/R10-HTC49770.2020.9356988

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A Survey of Methods for Explaining Black Box Models. *ACM Computing Surveys*, *51*(5), 1–42. https://doi.org/10.1145/3236009

Gunning, D., & Aha, D. (2019). DARPA’s Explainable Artificial Intelligence (XAI) Program. *AI Magazine*, *40*(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

Hall, P., & Gill, N. (2018). *Introduction to Machine Learning Interpretability* (Second edition). O'Reilly.

Hatwell, J., Gaber, M. M., & Azad, R. M. A. (2021). gbt-HIPS: Explaining the Classifications of Gradient Boosted Tree Ensembles. *Applied Sciences*, *11*(6), 2511. https://doi.org/10.3390/app11062511

Henderson, R. M., & Clark, K. B. (1990). Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. *Administrative Science Quarterly*, *35*(1), 9. https://doi.org/10.2307/2393549

Henri, J.-F., & Wouters, M. (2020). Interdependence of management control practices for product innovation: The influence of environmental unpredictability. *Accounting, Organizations and Society*, *86*, 101073. https://doi.org/10.1016/j.aos.2019.101073

Herrmann, A., Gassmann, O., & Eisert, U. (2007). An empirical study of the antecedents for radical product innovations and capabilities for transformation. *Journal of Engineering and Technology Management*, *24*(1-2), 92–120. https://doi.org/10.1016/j.jengtecman.2007.01.006

Herstatt, C., Verworn, B., & Nagahira, A. (2004). Reducing project related uncertainty in the "fuzzy front end" of innovation: a comparison of German and Japanese product innovation projects. *International Journal of Product Development*, *1*(1), Article 4890, 43. https://doi.org/10.1504/IJPD.2004.004890

Hertenstein, J. H., & Platt, M. B. (2000). Performance Measures and Management Control in New Product Development. *Accounting Horizons*, *14*(3), 303–323. https://doi.org/10.2308/acch.2000.14.3.303

Hey, J. H., Joyce, C. K., & Beckman, S. L. (2007). Framing innovation: negotiating shared frames during early design phases. *Journal of Design Research*, *6*(1/2), Article 15564, 79. https://doi.org/10.1504/JDR.2007.015564

Hirst, M. K. (1987). The effects of setting budget goals and task uncertainty on performance: A theoretical analysis. *Accounting Review*, *62*(4), 774–784.

Hirst, M. K. (1988). Intrinsic motivation as influenced by task interdependence and goal setting. *The Journal of Applied Psychology*, *73*(1), 96–101. https://doi.org/10.1037/0021-9010.73.1.96

Hirst, M. K., & Yetton, P. W. (1999). The effects of budget goals and task interdependence on the level of and variance in performance: a research note. *Accounting, Organizations and Society*, *24*(3), 205–216. https://doi.org/10.1016/S0361-3682(98)00037-3

Holton, C. (2009). Identifying disgruntled employee systems fraud risk through text mining: A simple solution for a multi-billion dollar problem. *Decision Support Systems*, *46*(4), 853–864. https://doi.org/10.1016/j.dss.2008.11.013

Hooker, S., Erhan, D., Kindermans, P.-J., & Kim, B. (2018, June 28). *A Benchmark for Interpretability Methods in Deep Neural Networks*. http://arxiv.org/pdf/1806.10758v3

Hopgood, A. A. (2005). The State of Artificial Intelligence. In *Advances in Computers* (Vol. 65, pp. 1–75). Elsevier. https://doi.org/10.1016/S0065-2458(05)65001-2

Höppe, F., & Moers, F. (2011). The Choice of Different Types of Subjectivity in CEO Annual Bonus Contracts. *The Accounting Review*, *86*(6), 2023–2046. https://doi.org/10.2308/accr-10132

Horn, D., & Salvendy, G. (2009). Measuring consumer perception of product creativity: Impact on satisfaction and purchasability. *Human Factors and Ergonomics in Manufacturing*, *19*(3), 223–240. https://doi.org/10.1002/hfm.20150

Horvitz, E., & Mulligan, D. (2015). Policy forum. Data, privacy, and the greater good. *Science (New York, N.Y.)*, *349*(6245), 253–255. https://doi.org/10.1126/science.aac4520

Hu, X., Xia, B., Skitmore, M., & Chen, Q. (2016). The application of case-based reasoning in construction management research: An overview. *Automation in Construction*, *72*, 65–74. https://doi.org/10.1016/j.autcon.2016.08.023

Hua Tan, K., Peng Lim, C., Platts, K., & Shen Koay, H. (2006). An intelligent decision support system for manufacturing technology investments. *International Journal of Production Economics*, *104*(1), 179–190. https://doi.org/10.1016/j.ijpe.2005.02.010

Huysmans, J., Dejaeger, K., Mues, C., Vanthienen, J., & Baesens, B. (2011). An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. *Decision Support Systems*, *51*(1), 141–154. https://doi.org/10.1016/j.dss.2010.12.003

Hwang, J. T. G., & Ding, A. A. (1997). Prediction Intervals for Artificial Neural Networks. *Journal of the American Statistical Association*, *92*(438), 748–757. https://doi.org/10.1080/01621459.1997.10474027

IBM. (2013). *The Four V's of Big Data*. https://www.ibmbigdatahub.com/infographic/four-vs-big-data

Ibusuki, U., & Kaminski, P. C. (2007). Product development process with focus on value engineering and target-costing: A case study in an automotive company. *International Journal of Production Economics*, *105*(2), 459–474. https://doi.org/10.1016/j.ijpe.2005.08.009

Institute of Management Accountants. (2008). *Definition of Management Accounting: Practice of Management Accounting*. https://www.imanet.org/-/media/6c984e4d7c854c2fb40b96bfbe991884.ashx?as=1&mh

International Group of Control. (2012). *Controlling Process Model*. Haufe-Lexware.

Israelsen, P., & Jørgensen, B. (2011). Decentralizing decision making in modularization strategies: Overcoming barriers from dysfunctional accounting systems. *International Journal of Production Economics*, *131*(2), 453–462. https://doi.org/10.1016/j.ijpe.2010.12.020

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer.

Järvenpää, M. (2007). Making Business Partners: A Case Study on how Management Accounting Culture was Changed. *European Accounting Review*, *16*(1), 99–142. https://doi.org/10.1080/09638180701265903

Joseph, J., & Vetrivel, A. (2012). Impact of Target Costing and Activity Based Costing on Improving the Profitability of Spinning Mills in Coimbatore-Empirical study on Spinning Mills. *Journal of Contemporary Research in Management*, *7*(2).

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Kato, Y. (1993). Target costing support systems: lessons from leading Japanese companies. *Management Accounting Research*, *4*(1), 33–47. https://doi.org/10.1006/mare.1993.1002

Kee, R., & Matherly, M. (2021). Target Costing in the Presence of Product and Production Interdependencies. In C. Akroyd & L. L. Burney (Eds.), *Advances in Management Accounting. Advances in Management Accounting* (Vol. 22, pp. 135–158). Emerald Publishing Limited. https://doi.org/10.1108/S1474-7871(2013)0000022011

Kelly, K. O., Webb, R. A., & Vance, T. (2015). The Interactive Effects of Ex Post Goal Adjustment and Goal Difficulty on Performance. *Journal of Management Accounting Research*, *27*(1), 1–25. https://doi.org/10.2308/jmar-50729

Kendall, K., Mangin, C., & Ortiz, E. (1998). Discrete event simulation and cost analysis for manufacturing optimisation of an automotive LCM component. *Composites Part a: Applied Science and Manufacturing*, *29*(7), 711–720. https://doi.org/10.1016/S1359-835X(98)00003-7

Keys, L. K. (1990). System life cycle engineering and DF'X'. *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, *13*(1), 83–93. https://doi.org/10.1109/33.52854

Khurana, A. (1998). Towards holistic “front ends” in new product development. *An International Publication of the Product Development & Management Association*, *15*(1), 57–74. https://doi.org/10.1016/S0737-6782(97)00066-0

Kim, G.-H., An, S.-H., & Kang, K.-I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and Environment*, *39*(10), 1235–1242. https://doi.org/10.1016/j.buildenv.2004.02.013

Kim, J., & Wilemon, D. (2002). Focusing the fuzzy front–end in new product development. *R&D Management*, *32*(4), 269–279. https://doi.org/10.1111/1467-9310.00259

Kim, K. (2003). Application of a hybrid genetic algorithm and neural network approach in activity-based costing. *Expert Systems with Applications*, *24*(1), 73–77. https://doi.org/10.1016/S0957-4174(02)00084-2

Kingma, D. P., & Ba, J. (2014, December 22). *Adam: A Method for Stochastic Optimization*. http://arxiv.org/pdf/1412.6980v9

Kleingeld, A., van Mierlo, H., & Arends, L. (2011). The effect of goal setting on group performance: A meta-analysis. *The Journal of Applied Psychology*, *96*(6), 1289–1304. https://doi.org/10.1037/a0024315

Kostakis, H., Sarigiannidis, C., Boutsinas, B., Varvakis, K., & Tampakas, V. (2008). Integrating activity-based costing with simulation and data mining. *International Journal of Accounting & Information Management*, *16*(1), 25–35. https://doi.org/10.1108/18347640810887744

Krishnan, V., & Ulrich, K. T. (2001). Product Development Decisions: A Review of the Literature. *Management Science*, *47*(1), 1–21. https://doi.org/10.1287/mnsc.47.1.1.10668

Kuzey, C., Uyar, A., & Delen, D. (2019). An investigation of the factors influencing cost system functionality using decision trees, support vector machines and logistic regression. *International Journal of Accounting & Information Management*, *27*(1), 27–55. https://doi.org/10.1108/IJAIM-04-2017-0052

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, *34*(3), 387–394. https://doi.org/10.1016/j.ijinfomgt.2014.02.002

Labro, E. (2004). The Cost Effects of Component Commonality: A Literature Review Through a Management-Accounting Lens. *Manufacturing & Service Operations Management*, *6*(4), 358–367. https://doi.org/10.1287/msom.1040.0047

Landry, M., Erlinger, T. P., Patschke, D., & Varrichio, C. (2016). Probabilistic gradient boosting machines for GEFCom2014 wind forecasting. *International Journal of Forecasting*, *32*(3), 1061–1066. https://doi.org/10.1016/j.ijforecast.2016.02.002

Laney, D. (2001, February 6). *3D data management: Controlling data volume, velocity and variety*. META group. https://scribd.com/document/452060252/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety-pdf

Langmann, C. (2019). *Digitalisierung im Controlling*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-25017-1

Latham, G. P. (2009). Motivate Employee Performance through Goal Setting. In E. A. Locke (Ed.), *The Blackwell Handbook of Principles of Organizational Behaviour* (pp. 161–178). Blackwell Publishing Ltd.

Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes*, *50*(2), 212–247. https://doi.org/10.1016/0749-5978(91)90021-K

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, *52*(2).

Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons*, *63*(2), 157–170. https://doi.org/10.1016/j.bushor.2019.10.005

Legg, S., & Hutter, M. (2007). A Collection of Definitions of Intelligence. *Frontiers in Artificial Intelligence and Applications*. http://arxiv.org/pdf/0706.3639v1

Li Qian, & Ben-Arieh, D. (2008). Parametric cost estimation based on activity-based costing: A case study for design and development of rotational parts. *International Journal of Production Economics*, *113*(2), 805–818. https://doi.org/10.1016/j.ijpe.2007.08.010

Liang, T.-P., Chandler, J. S., Han, I., & Roan, J. (1992). An empirical investigation of some data effects on the classification accuracy of probit, ID3, and neural networks. *Contemporary Accounting Research*, *9*(1), 306–328. https://doi.org/10.1111/j.1911-3846.1992.tb00883.x

Linkov, I., Galaitsi, S., Trump, B. D., Keisler, J. M., & Kott, A. (2020). Cybertrust: From Explainable to Actionable and Interpretable Artificial Intelligence. *Computer*, *53*(9), 91–96. https://doi.org/10.1109/MC.2020.2993623

Lipton, Z. C. (2018). The mythos of model interpretability. *Queue*, *16*(3).

Liu, H., Cocea, M., & Gegov, A. (2016). Interpretability of Computational Models for Sentiment Analysis. In W. Pedrycz & S.-M. Chen (Eds.), *Studies in Computational Intelligence: Vol. 639. Sentiment Analysis and Ontology Engineering* (Vol. 639, pp. 199–220). Springer International Publishing. https://doi.org/10.1007/978-3-319-30319-2_9

Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, *3*(2), 157–189. https://doi.org/10.1016/0030-5073(68)90004-4

Locke, E. A., Chah, D.-O., Harrison, S., & Lustgarten, N. (1989). Separating the effects of goal specificity from goal level. *Organizational Behavior and Human Decision Processes*, *43*(2), 270–287. https://doi.org/10.1016/0749-5978(89)90053-8

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. *American Psychologist*, *57*(9), 705–717. https://doi.org/10.1037//0003-066x.57.9.705

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969-1980. *Psychological Bulletin*, *90*(1), 125–152. https://doi.org/10.1037/0033-2909.90.1.125

Losbichler, H., & Lehner, O. M. (2021). Limits of artificial intelligence in controlling and the ways forward: a call for future accounting research. *Journal of Applied Accounting Research*, *22*(2), 365–382. https://doi.org/10.1108/JAAR-10-2020-0207

Loyer, J.-L., Henriques, E., Fontul, M., & Wiseall, S. (2016). Comparison of Machine Learning methods applied to the estimation of manufacturing cost of jet engine components. *International Journal of Production Economics*, *178*, 109–119. https://doi.org/10.1016/j.ijpe.2016.05.006

Lucae, S., Rebentisch, E., & Oehmen, J. (2014). Understanding the Front-end of Large-scale Engineering Programs. *Procedia Computer Science*, *28*, 653–662. https://doi.org/10.1016/j.procs.2014.03.079

Luxburg, U. von, & Schölkopf, B. (2011). Statistical Learning Theory: Models, Concepts, and Results. *Handbook of the History of Logic*, *10*, 651–706. https://doi.org/10.1016/B978-0-444-52936-7.50016-1

Mardini, G. H., & Alkurdi, A. (2021). Artificial Intelligence Literature in Accounting: A Panel Systematic Approach. In A. Hamdan, A. E. Hassanien, A. Razzaque, & B. Alareeni (Eds.), *Studies in Computational Intelligence: Vol. 935. The Fourth Industrial Revolution: Implementation of Artificial Intelligence for Growing Business Success* (Vol. 935, pp. 311–323). Springer International Publishing. https://doi.org/10.1007/978-3-030-62796-6_18

Markus, L. M. (2001). Toward a Theory of Knowledge Reuse: Types of Knowledge Reuse Situations and Factors in Reuse Success. *Journal of Management Information Systems*, *18*(1), 57–93. https://doi.org/10.1080/07421222.2001.11045671

Matějka, M. (2018). Target Setting in Multi-Divisional Organizations. *Journal of Management Accounting Research*, *30*(3), 13–27. https://doi.org/10.2308/jmar-52159

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard Business Review*, *90*(10), 60–68.

McKinsey Global Institute. (2011, May 1). *Big data: The next frontier for innovation, competition, and productivity*. https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation

Merchant, K. A., & Manzoni, J.-F. (1989). The achievability of budget targets in profit centers: A field study. *The Accounting Review*, *64*(3), 539–558.

Merchant, K. A., Stringer, C., & Shantapriyan, P. (2018). Setting Financial Performance Thresholds, Targets, and Maximums in Bonus Plans. *Journal of Management Accounting Research*, *30*(3), 55–73. https://doi.org/10.2308/jmar-52116

Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2020). Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities. *Information Systems Management*, 1–11. https://doi.org/10.1080/10580530.2020.1849465

Mihm, J. (2010). Incentives in New Product Development Projects and the Role of Target Costing. *Management Science*, *56*(8), 1324–1344. https://doi.org/10.1287/mnsc.1100.1175

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Miner, J. B. (2005). *Organizational Behavior: Essential theories of motivation and leadership. one*. M.E. Sharpe.

Mitchell, T. R., & Silver, W. S. (1990). Individual and group goals when workers are interdependent: Effects on task strategies and performance. *Journal of Applied Psychology*, *75*(2), 185–193. https://doi.org/10.1037/0021-9010.75.2.185

Mitchell, T. M. (1999). Machine learning and data mining. *Communications of the ACM*, *42*(11), 30–36. https://doi.org/10.1145/319382.319388

Moenaert, R. K., Meyer, A. de, Souder, W. E., & Deschoolmeester, D. (1995). R&D/marketing communication during the fuzzy front-end. *IEEE Transactions on Engineering Management*, *42*(3), 243–258. https://doi.org/10.1109/17.403743

Mohseni, S., Zarei, N., & Ragan, E. D. (2018). *A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems* (3-4). https://doi.org/10.1145/3387166

Monden, Y., Akter, M., & Kubo, N. (1997). Target Costing Performance Based on Alternative Participation and Evaluation Methods: A Laboratory Experiment. *Managerial and Decision Economics*, *18*(2), 113–129. https://doi.org/10.1002/(SICI)1099-1468(199703)18:2<113::AID-MDE814>3.0.CO;2-Y

Monden, Y., & Hamada, K. (1991). Target costing and kaizen costing in Japanese automobile companies. *Journal of Management Accounting Research*, *3*(1), 16–

34.
http://proxy1.ncu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=9701273024&site=ehost-live

Moon, J.-W., & Kim, Y.-G. (2001). Extending the TAM for a World-Wide-Web context. *Information & Management*, *38*(4), 217–230. https://doi.org/10.1016/S0378-7206(00)00061-6

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, *33*(2), 161–174.

Navissi, F., & Sridharan, V. G. (2017). Determinants of Target Costing Adoption: A Research Note. *Journal of Management Accounting Research*, *29*(1), 67–77. https://doi.org/10.2308/jmar-51501

Neary, P. (2018). Automatic Hyperparameter Tuning in Deep Convolutional Neural Networks Using Asynchronous Reinforcement Learning. In P. Neary (Ed.), *Automatic hyperparameter tuning in deep convolutional neural networks using asynchronous reinforcement learning* (pp. 73–77). IEEE. https://doi.org/10.1109/ICCC.2018.00017

Niazi, A., Dai, J. S., Balabani, S., & Seneviratne, L. (2006). Product Cost Estimation: Technique Classification and Methodology Review. *Journal of Manufacturing Science and Engineering*, *128*(2), 563–575. https://doi.org/10.1115/1.2137750

Nielsen, S. (2018). Reflections on the applicability of business analytics for management accounting – and future perspectives for the accountant. *Journal of Accounting & Organizational Change*, *14*(2), 167–187. https://doi.org/10.1108/JAOC-11-2014-0056

Norman, D. A., & Verganti, R. (2014). Incremental and Radical Innovation: Design Research vs. Technology and Meaning Change. *Design Issues*, *30*(1), 78–96. https://doi.org/10.1162/DESI_a_00250

Odar, M., Kavčič, S., & Jerman, M. (2015). The Role of a Management Accounting System in the Decision-Making Process: Evidence from a Post-Transition Economy. *Engineering Economics*, *26*(1). https://doi.org/10.5755/j01.ee.26.1.4873

Oladipupo, T. (2010). Types of Machine Learning Algorithms. *New Advances in Machine Learning.* Advance online publication. https://doi.org/10.5772/9385

O'Leary, D. E. (2013). Artificial Intelligence and Big Data. *IEEE Intelligent Systems*, *28*(2), 96–99. https://doi.org/10.1109/MIS.2013.39

Otley, D. T., & Berry, A. J. (1994). Case study research in management accounting and control. *Management Accounting Research*, *5*(1), 45–65. https://doi.org/10.1006/mare.1994.1004

Partovi, F. Y., & Anandarajan, M. (2002). Classifying inventory using an artificial neural network approach. *Computers & Industrial Engineering*, *41*(4), 389–404. https://doi.org/10.1016/S0360-8352(01)00064-X

Patanakul, P., Chen, J., & Lynn, G. S. (2012). Autonomous Teams and New Product Development. *Journal of Product Innovation Management*, *29*(5), 734–750. https://doi.org/10.1111/j.1540-5885.2012.00934.x

Peake, G., & Wang, J [Jun] (2018). Explanation Mining. In Y. Guo & F. Farooq (Eds.), *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2060–2069). ACM. https://doi.org/10.1145/3219819.3220072

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A. and Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214–232. https://doi.org/10.1016/j.envsoft.2016.02.008

Pickard, M. D., & Cokins, G. (2015). From Bean Counters to Bean Growers: Accountants as Data Analysts—A Customer Profitability Example. *Journal of Information Systems*, *29*(3), 151–164. https://doi.org/10.2308/isys-51180

Piltaver, R., Luštrek, M., Gams, M., & Martinčić-Ipšić, S. (2016). What makes classification trees comprehensible? *Expert Systems with Applications*, *62*, 333–346. https://doi.org/10.1016/j.eswa.2016.06.009

Platts, K. W. (1993). A Process Approach to Researching Manufacturing Strategy. *International Journal of Operations & Production Management*, *13*(8), 4–17. https://doi.org/10.1108/01443579310039533

Polzer, J. T., & Neale, M. A. (1995). Constraints or Catalysts? Reexamining goal Setting Within the Context of Negotiation. *Human Performance, 8*(1), 3–26. https://doi.org/10.1207/s15327043hup0801_2

Preece, A. (2018). Asking ‘Why’ in AI: Explainability of intelligent systems – perspectives and challenges. *Intelligent Systems in Accounting, Finance and Management*, *25*(2), 63–72. https://doi.org/10.1002/isaf.1422

Pullen, A., Weerd-Nederhof, P. de, Groen, A., Song, M., & Fisscher, O. (2009). Successful Patterns of Internal SME Characteristics Leading to High Overall Innovation Performance. *Creativity and Innovation Management*, *18*(3), 209–223. https://doi.org/10.1111/j.1467-8691.2009.00530.x

Qi, J., Zhang, Z., Jeon, S., & Zhou, Y. (2016). Mining customer requirements from online reviews: A product improvement perspective. *Information & Management*, *53*(8), 951–963. https://doi.org/10.1016/j.im.2016.06.002

Reid, S. E., & Brentani, U. de (2004). The Fuzzy Front End of New Product Development for Discontinuous Innovations: A Theoretical Model. *Journal of Product Innovation Management*, *21*(3), 170–184. https://doi.org/10.1111/j.0737-6782.2004.00068.x

Rey, M. I., Galende, M., Fuente, M. J., & Sainz-Palmero, G. I. (2017). Multi-objective based Fuzzy Rule Based Systems (FRBSs) for trade-off improvement in accuracy and interpretability: A rule relevance point of view. *Knowledge-Based Systems*, *127*, 67–84. https://doi.org/10.1016/j.knosys.2016.12.028

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, June 17). *Model-Agnostic Interpretability of Machine Learning*. http://arxiv.org/pdf/1606.05386v1

Richins, G., Stapleton, A., Stratopoulos, T. C., & Wong, C. (2017). Big Data Analytics: Opportunity or Threat for the Accounting Profession? *Journal of Information Systems*, *31*(3), 63–79. https://doi.org/10.2308/isys-51805

Rikhardsson, P., & Yigitbasioglu, O. (2018). Business intelligence & analytics in management accounting research: Status and future focus. *International Journal of Accounting Information Systems*, *29*, 37–58. https://doi.org/10.1016/j.accinf.2018.03.001

Rom, A., & Rohde, C. (2006). Enterprise resource planning systems, strategic enterprise management systems and management accounting. *Journal of Enterprise Information Management*, *19*(1), 50–66. https://doi.org/10.1108/17410390610636878

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408. https://doi.org/10.1037/h0042519

Roy, R. (2003). *Cost Engineering: Why, What and How? Decision Engineering Report Series*. Cranfield University.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Runkler, T. A. (2016). *Data Analytics*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-14075-5

Rush, C., & Roy, R. (2000). Analysis of cost estimating processes used within a concurrent engineering environment throughout a product life cycle. In *7th ISPE International Conference on Concurrent Engineering: Research and Applications,* Lyon, France.

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (Third edition, Global edition). *Always learning*. Pearson.

Saavedra, R., Earley, P. C., & van Dyne, L. (1993). Complex interdependence in task-performing groups. *Journal of Applied Psychology*, *78*(1), 61–72. https://doi.org/10.1037/0021-9010.78.1.61

Saggi, M. K., & Jain, S. (2018). A survey towards an integration of big data analytics to big insights for value-creation. *Information Processing & Management*, *54*(5), 758–790. https://doi.org/10.1016/j.ipm.2018.01.010

Samek, W., Wiegand, T., & Müller, K.-R. (2017, August 28). *Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models*. http://arxiv.org/pdf/1708.08296v1

Scapens, R. W. (1990). Researching management accounting practice: The role of case study methods. *The British Accounting Review*, *22*(3), 259–281. https://doi.org/10.1016/0890-8389(90)90008-6

Scarbrough, P., Nanni, A. J., & Sakurai, M. (1991). Japanese management accounting practices and the effects of assembly and process automation. *Management Accounting Research*, *2*(1), 27–46. https://doi.org/10.1016/S1044-5005(91)70025-5

Schmidt, J. B., Sarangee, K. R., & Montoya, M. M. (2009). Exploring New Product Development Project Review Practices. *Journal of Product Innovation Management*, *26*(5), 520–535. https://doi.org/10.1111/j.1540-5885.2009.00678.x

Schniederjans, M. J., & Garvin, T. (1997). Using the analytic hierarchy process and multi-objective programming for the selection of cost drivers in activity-based costing. *European Journal of Operational Research*, *100*(1), 72–80. https://doi.org/10.1016/S0377-2217(96)00302-5

Scholkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., & Bengio, Y. (2021). Toward Causal Representation Learning. *Proceedings of the IEEE*, *109*(5), 612–634. https://doi.org/10.1109/JPROC.2021.3058954

Shawver, T. J. (2005). Merger Premium Predictions Using a Neural Network Approach. *Journal of Emerging Technologies in Accounting*, *2*(1), 61–72. https://doi.org/10.2308/jeta.2005.2.1.61

Shin, Y. (2015). Application of Boosting Regression Trees to Preliminary Cost Estimation in Building Construction Projects. *Computational Intelligence and Neuroscience*, *2015*, 149702. https://doi.org/10.1155/2015/149702

Shtub, A., & Zimerman, Y. (1993). A neural-network-based approach for estimating the cost of assembly systems. *International Journal of Production Economics*, *32*(2), 189–207. https://doi.org/10.1016/0925-5273(93)90068-V

Smith, A. E., & Mason, A. K. (1997). Cost Estimation Predictive Modeling: Regression versus Neural Network. *The Engineering Economist*, *42*(2), 137–161. https://doi.org/10.1080/00137919708903174

Song, M., & Montoya-Weiss, M. M. (2001). The effect of perceived technological uncertainty on japanese new product development. *Academy of Management Journal*, *44*(1), 61–80. https://doi.org/10.2307/3069337

Stadtherr, F., & Wouters, M. (2017). Cost management and modular product design strategies. In E. Harris (Ed.), *Routledge companions in business, management and accounting. The Routledge companion to performance management and control* (pp. 54–86). Routledge. https://doi.org/10.4324/9781315691374

Stadtherr, F., & Wouters, M. (2021). Extending target costing to include targets for R&D costs and production investments for a modular product portfolio—A case study. *International Journal of Production Economics*, *231*, 107871. https://doi.org/10.1016/j.ijpe.2020.107871

Stockton, D. J., Khalil, R. A., & Mukhongo, L. M. (2013). Cost model development using virtual manufacturing and data mining: part II—comparison of data mining algorithms. *The International Journal of Advanced Manufacturing Technology*, *66*(9-12), 1389–1396. https://doi.org/10.1007/s00170-012-4416-5

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tan, K. H., Zhan, Y., Ji, G., Ye, F., & Chang, C. (2015). Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph. *International Journal of Production Economics*, *165*, 223–233. https://doi.org/10.1016/j.ijpe.2014.12.034

Tanaka, T. (1993). Target costing at Toyota. *Journal of Cost Management*, *7*(Spring), 4–11.

Tang, Y.-C. (2009). An approach to budget allocation for an aerospace company—Fuzzy analytic hierarchy process and artificial neural network. *Neurocomputing*, *72*(16-18), 3477–3489. https://doi.org/10.1016/j.neucom.2009.03.020

Tani, T. (1995). Interactive control in target cost management. *Management Accounting Research*, *6*(4), 399–414. https://doi.org/10.1006/mare.1995.1028

Tani, T., Okano, H., Shimizu, N., Iwabuchi, Y., Fukuda, J., & Cooray, S. (1994). Target cost management in Japanese companies: current state of the art. *Management Accounting Research*, *5*(1), 67–81. https://doi.org/10.1006/mare.1994.1005

Tatikonda, M. V., & Rosenthal, S. R. (2000). Technology novelty, project complexity, and product development project execution success: a deeper look at task uncertainty in product innovation. *IEEE Transactions on Engineering Management*, *47*(1), 74–87. https://doi.org/10.1109/17.820727

Thore Olsson, A.-C., Johannesson, U., & Schweizer, R. (2018). Decision-making and cost deviation in new product development projects. *International Journal of Managing Projects in Business*, *11*(4), 1066–1085. https://doi.org/10.1108/IJMPB-02-2018-0029

Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, *66*(8), 1352–1362. https://doi.org/10.1057/jors.2014.103

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, *14*(3), 207–222. https://doi.org/10.1111/1467-8551.00375

Tripathi, S., Muhr, D., Brunner, M., Jodlbauer, H., Dehmer, M., & Emmert-Streib, F. (2021). Ensuring the Robustness and Reliability of Data-Driven Knowledge Discovery Models in Production and Manufacturing. *Frontiers in Artificial Intelligence*, *4*, 576892. https://doi.org/10.3389/frai.2021.576892

Tseng, K. C., El-Ganzoury, W., & Abdalla, H. (2005). Integration of Non-networked Software Agents for Collaborative Product Development. *Computer-Aided*

Design and Applications, *2*(1-4), 377–386. https://doi.org/10.1080/16864360.2005.10738386

Tushman, M. L., & Nadler, D. A. (1978). Information Processing as an Integrating Concept in Organizational Design. *Academy of Management Review*, *3*(3), 613–624. https://doi.org/10.5465/amr.1978.4305791

Tyagi, S., Cai, X., & Yang, K. (2015). Product life-cycle cost estimation: a focus on the multi-generation manufacturing-based product. *Research in Engineering Design*, *26*(3), 277–288. https://doi.org/10.1007/s00163-015-0196-x

Tyebjee, T. T. (1987). Behavioral biases in new product forecasting. *International Journal of Forecasting*, *3*(3-4), 393–404. https://doi.org/10.1016/0169-2070(87)90032-X

van der Aalst, W. (2016). *Process Mining*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-49851-4

van der Linden, J., & van de Leemput, C. (2015). Observatory of students' uses of computer-based tools. *Psychologie Française*, *60*(2), 145–157. https://doi.org/10.1016/j.psfr.2015.02.002

VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data* (First edition). O'Reilly.

Vasarhelyi, M. A., Kogan, A., & Tuttle, B. M. (2015). Big Data in Accounting: An Overview. *Accounting Horizons*, *29*(2), 381–396. https://doi.org/10.2308/acch-51071

Venkatesh, Morris, & Davis (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425. https://doi.org/10.2307/30036540

Verlinden, B., Duflou, J. R., Collin, P., & Cattrysse, D. (2008). Cost estimation for sheet metal parts using multiple regression and artificial neural networks: A case study. *International Journal of Production Economics*, *111*(2), 484–492. https://doi.org/10.1016/j.ijpe.2007.02.004

Wang, Q. (2007). Artificial neural networks as cost engineering methods in a collaborative manufacturing environment. *International Journal of Production Economics*, *109*(1-2), 53–64. https://doi.org/10.1016/j.ijpe.2006.11.006

Wang, S.-G., Tse, S.-K., & Chow, S.-C. (1990). On the measures of multicollinearity in least squares regression. *Statistics & Probability Letters*, *9*(4), 347–355. https://doi.org/10.1016/0167-7152(90)90145-W

Warren, J. D., Moffitt, K. C., & Byrnes, P. (2015). How Big Data Will Change Accounting. *Accounting Horizons*, *29*(2), 397–407. https://doi.org/10.2308/acch-51069

Webb, R. A., Jeffrey, S. A., & Schulz, A. (2010). Factors Affecting Goal Difficulty and Performance When Employees Select Their Own Performance Goals: Evidence from the Field. *Journal of Management Accounting Research*, *22*(1), 209–232. https://doi.org/10.2308/jmar.2010.22.1.209

Webb, R. A., Williamson, M. G., & Zhang, Y. M. (2013). Productivity-Target Difficulty, Target-Based Pay, and Outside-the-Box Thinking. *The Accounting Review*, *88*(4), 1433–1457. https://doi.org/10.2308/accr-50436

Wijewardena, H., & Zoysa, A. de (1999). A Comparative Analysis of Management Accounting Practices in Australia and Japan: An Empirical Investigation. *The International Journal of Accounting*, *34*(1), 49–70. https://doi.org/10.1016/S0020-7063(99)80003-X

Wixom, B. H., & Todd, P. A. (2005). A Theoretical Integration of User Satisfaction and Technology Acceptance. *Information Systems Research*, *16*(1), 85–102. https://doi.org/10.1287/isre.1050.0042

Wlodarczyk, T. W., & Hacker, T. J. (2014). Current trends in predictive analytics of big data. *International Journal of Big Data Intelligence*, *1*(3), Article 66326, 172. https://doi.org/10.1504/IJBDI.2014.066326

Wood, R. E., Mento, A. J., & Locke, E. A. (1987). Task complexity as a moderator of goal effects: A meta-analysis. *The Journal of Applied Psychology*, *72*(3), 416–425. https://doi.org/10.1037/0021-9010.72.3.416

Woods, M., Taylor, L., & Fang, G. C. G. (2012). Electronics: A case study of economic value added in target costing. *Management Accounting Research*, *23*(4), 261–277. https://doi.org/10.1016/j.mar.2012.09.002

Wouters, M., & Morales, S. (2014). The Contemporary Art of Cost Management Methods during Product Development. *Advances in Management Accounting,*, *24*, 259–346. https://doi.org/10.1108/S1474-787120140000024008

Wouters, M., Morales, S., Grollmuss, S., & Scheer, M. (2021). Methods for Cost Management during Product Development: A Review and Comparison of Different Literatures. In C. Akroyd & L. L. Burney (Eds.), *Advances in Management Accounting. Advances in Management Accounting* (Vol. 26, pp. 139–274). Emerald Publishing Limited. https://doi.org/10.1108/S1474-787120150000026005

Wu, S., & Akbarov, A. (2011). Support vector regression for warranty claim forecasting. *European Journal of Operational Research*, *213*(1), 196–204. https://doi.org/10.1016/j.ejor.2011.03.009

Wu, X., Zhu, X., Wu, G.-Q., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, *26*(1), 97–107. https://doi.org/10.1109/TKDE.2013.109

Wynn, D. C., Grebici, K., & Clarkson, P. J. (2011). Modelling the evolution of uncertainty levels during design. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *5*(3), 187–202. https://doi.org/10.1007/s12008-011-0131-y

Xia, H., Vu, H. Q., Law, R., & Li, G. (2020). Evaluation of hotel brand competitiveness based on hotel features ratings. *International Journal of Hospitality Management*, *86*, 102366. https://doi.org/10.1016/j.ijhm.2019.102366

Xia, Y., Liu, C., Li, Y., & Liu, N. (2017). A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications*, *78*, 225–241. https://doi.org/10.1016/j.eswa.2017.02.017

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th edition). Sage.

Yoshikawa, T., Innes, J., & Mitchell, F. (1994). Applying functional cost analysis in a manufacturing environment. *International Journal of Production Economics*, *36*(1), 53–64. https://doi.org/10.1016/0925-5273(94)90148-1

Yukl, G. A., & Latham, G. P. (1978). Interrelationships among employee participation, individual differences, goal difficulty, goal acceptance, goal instrumentality, and performance. *Personnel Psychology*, *31*(2), 305–323. https://doi.org/10.1111/j.1744-6570.1978.tb00449.x

Zhan, Y., & Tan, K. H. (2020). An analytic infrastructure for harvesting big data to enhance supply chain performance. *European Journal of Operational Research*, *281*(3), 559–574. https://doi.org/10.1016/j.ejor.2018.09.018

Zhang, X., Wang, Y [Yuan], & Li, Z. (2021). User acceptance of machine learning models – Integrating several important external variables with technology acceptance model. *The International Journal of Electrical Engineering & Education*, 002072092110052. https://doi.org/10.1177/00207209211005271

Zhang, Y. F., Fuh, J., & Chan, W. T. (1996). Feature-based cost estimation for packaging products using neural networks. *Computers in Industry*, *32*(1), 95–113. https://doi.org/10.1016/S0166-3615(96)00059-0

Zhang, Y., Zhang, M., Li, J., Liu, G., Yang, M. M., & Liu, S. (2021). A bibliometric review of a decade of research: Big data in business research – Setting a research agenda. *Journal of Business Research*, *131*, 374–390. https://doi.org/10.1016/j.jbusres.2020.11.004

Zhu, D.-S., Lin, T. C.-T., & Hsu, Y.-C. (2012). Using the technology acceptance model to evaluate user attitude and intention of use for online games. *Total Quality Management & Business Excellence*, *23*(7-8), 965–980. https://doi.org/10.1080/14783363.2012.704269