European Accounting Review

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rear20

Explainability Versus Accuracy of Machine Learning Models: The Role of Task Uncertainty and Need for Interaction with the Machine Learning Model

Dominik Hammann & Marc Wouters

To cite this article: Dominik Hammann & Marc Wouters (25 Feb 2025): Explainability Versus Accuracy of Machine Learning Models: The Role of Task Uncertainty and Need for Interaction with the Machine Learning Model, European Accounting Review, DOI: 10.1080/09638180.2025.2463961

To link to this article: https://doi.org/10.1080/09638180.2025.2463961

European Accounting Review, 2025
https://doi.org/10.1080/09638180.2025.2463961

Check for updates

Explainability Versus Accuracy of Machine Learning Models: The Role of Task Uncertainty and Need for Interaction with the Machine Learning Model

DOMINIK HAMMANN* and MARC WOUTERS *,**

**Department of Economics and Management, Karlsruhe Institute of Technology, Karlsruhe, Germany **Amsterdam Business School, University of Amsterdam, Amsterdam, Netherlands*

(*Received: April 2023; accepted: January 2025*)

ABSTRACT This paper investigates the importance of explainability versus accuracy of machine learning (ML) models. We propose that greater task uncertainty makes people want to interact more with the ML model, which increases the importance of explainability relative to accuracy. We focus on the use of ML models for product cost estimation during new product development. The paper provides mixed-methods evidence on the trade-off between explainability and accuracy of ML models. Specifically, we find support for an inverse relationship between explainability and accuracy from the perspective of cost experts. We also find that the accurate but complex and less explainable ML model of gradient boosted regression (GBR) was preferred in only a few situations; mostly, the more basic, better explainable models of multiple linear regression (MLR) and case-based reasoning (CBR) were preferred, although these were less accurate. This suggests that lack of explainability can indeed be a major limitation for the application of ML models. Furthermore, we investigate specific characteristics that could increase task uncertainty and the importance of explainability in our context: project unpredictability, product cost granularity, predecessor product availability, target cost gap, and product development phase.

Keywords: Machine learning; Explainability; Cost estimation; Task uncertainty;

1. Introduction

Machine learning receives increasing attention in management accounting (Abernethy et al., 2023; Amani & Fadlalla, 2017; Fehrenbacher et al., 2023; Kuzey et al., 2019; LaValle et al., 2011; Losbichler & Lehner, 2021; Mahlendorf et al., 2023; Nielsen, 2018; Repenning et al., 2022; Rikhardsson & Yigitbasioglu, 2018; Sutton et al., 2016). However, the applicability of complex machine learning (ML) models is often limited by the ability to provide explainable results (Arnaboldi et al., 2022; Bertomeu, 2020; Chou et al., 2010; Coussement et al., 2017; Lehner et al., 2022; Ranta et al., 2023). In the context of ML, *explainability* is the ability to

---

Correspondence Address: Marc Wouters, Karlsruhe Institute of Technology, Kaiserstraße 89, 76133 Karlsruhe, Germany. Email: marc.wouters@kit.edu

Paper accepted by Matthias Mahlendorf.

demonstrate the reasons for a behavior or the ability to produce insights about the cause of an outcome (Gilpin et al., 2018). Some ML models are more explainable than others, and there is generally a trade-off between accuracy and explainability. *Accuracy* 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 ML models are not very explainable, and vice versa (Adadi & Berrada, 2018). State-of-the-art ML techniques with high accuracy are often referred to as black boxes (Chye Koh & Kee Low, 2004; Kim et al., 2004; Zhang, 2019). This is problematic since explainability is often required to build trust and acceptance of ML models (Nielsen, 2022; Zhang et al., 2023) and is therefore crucial in managerial accounting (Appelbaum et al., 2017; Kietzmann & Pitt, 2020; Quattrone, 2016).

The inverse relationship between explainability and accuracy is known as the interpretability problem (Cavalieri et al., 2004; Loyer et al., 2016; Verlinden et al., 2008). Explainable AI aims to make artificial intelligence systems more understandable to users while maintaining accuracy (Adadi & Berrada, 2018; Arnaboldi et al., 2022; Preece, 2018). Explainable AI is a promising branch of artificial intelligence research that requires more attention in management accounting research (Ranta et al., 2023). Nevertheless, also after improving the explainability of complex ML models, the fundamental trade-off between explainability and accuracy remains (Barredo Arrieta et al., 2020; Emmert-Streib et al., 2020; Linkov et al., 2020). Given this trade-off, when do people prefer accuracy and are willing to sacrifice some explainability, but when do they prioritize explainability at the expense of some accuracy? Only very few empirical studies have investigated this question (Baryannis et al., 2019; Rey et al., 2017).

We focus on the trade-off between explainability and accuracy in the context of cost estimation during product development. Product development is overall an interesting and specific setting in terms of the role of management accounting (Gopalakrishnan et al., 2015; Grabner et al., 2018; Guo et al., 2019; Janka & Guenther, 2018; Moll, 2015; Taipaleenmäki, 2014). The use of ML models is particularly interesting to be investigated in product development because these models are promising for cost estimation tasks (Caputo & Pelagagge, 2008; Loyer et al., 2016; Wang, 2007) and cost estimation is particularly relevant during product development, when the product design can still be changed for reducing costs on the basis of cost estimation results (Ansari et al., 2006; Booker et al., 2007; Henri & Wouters, 2020; Wouters et al., 2021). Explainability is required, because cost management experts need to understand and be able to verify the results, so they are better able to explicate and justify these to other people they interact with in cross-functional development teams. Also, explainability is needed so cost management experts are able to modify and improve cost estimation models. Moreover, cost estimation results need to be discussed within the wider decision-making processes of product development. Cost estimation models do not capture all relevant considerations for product development decisions, so the possibility to connect the cost estimation results to other considerations makes explainability more relevant and potentially valuable for the firm. At the same time, accuracy, so the degree to which the estimated cost differs from the actual cost for a product (Dysert, 2007), is necessary to make informed decisions with regard to the design, costs, and profitability of future products. This paper empirically investigates management accountants' trade-offs between explainability and accuracy in this context. We address the research question: Which factors determine the relative importance of explainability versus accuracy of ML models for product cost estimation during new product development?

We reason that the importance of explainability relative to accuracy depends on the user's *need to interact with the ML model*, for example, to compare model results with the guestimates of human experts, to understand the impact of particular assumptions, to analyze differences between model results and a reference point, such as the cost of a previous product or a target cost. The interactions with the ML model require a greater level of explainability of the model.

We propose that greater *task uncertainty* makes people want to interact with the model more. Task uncertainty refers to the perceived difference between the information that is needed and that is available for a particular task (Chapman, 1997; Davila, 2000; Galbraith, 1973). Task uncertainty implies that more job-relevant information must be processed during task execution to achieve a particular level of performance (Galbraith, 1973). Essentially, when task uncertainty is larger, cost estimation results trigger more questions and probing about why the model produces particular results, making explainability more important.

This paper contributes to the literature by providing mixed-methods evidence on the trade-off between explainability and accuracy of ML models. Cost experts at a German car manufacturer, who are dealing with product cost estimation during new product development, conducted several variations of a cost estimation task that mirrored their actual work. First, we empirically show the inverse relationship between explainability and accuracy of ML models from the perspective of product controllers (Cavalieri et al., 2004; Loyer et al., 2016; Ranta et al., 2023; Verlinden et al., 2008). Second, we show that this trade-off influences the selection of ML models. We empirically support that lack of explainability can be a limitation for applying ML models, as suggested in accounting research (Arnaboldi et al., 2022; Bertomeu, 2020; Chye Koh & Kee Low, 2004; Lehner et al., 2022; Ranta et al., 2023) and in studies on cost estimation (Cavalieri et al., 2004; Loyer et al., 2016; Verlinden et al., 2008).

Third, we demonstrate the importance of task uncertainty in relation to the use of ML models. This broadens the understanding of the role of task uncertainty for the design and use of management accounting systems (Chapman, 1997; Chong & Johnson, 2007; Davila, 2000; Williams & Seaman, 2002; Ylinen & Gullkvist, 2012). We investigate the impact of five factors that are expected to create task uncertainty and the need to interact with the model in the context of cost estimation during product development, so these factors would influence the importance of explainability relative to accuracy. We find that explainability is relatively more important when project unpredictability is high, product cost estimation granularity concerns a complete product, the target cost gap is large, and the product development phase is early. This contributes to research on the trade-off between explainability and accuracy of ML models (Alonso et al., 2015; Baryannis et al., 2019; Frias-Martinez et al., 2005) by introducing and empirically testing context-specific factors that influence the relative importance of explainability and accuracy. Finally, this study provides a comparison of a direct and indirect measurement of the relative importance of explainability and accuracy. The results differ, suggesting the indirect measurement should be used only cautiously.

In Section 2, we review research on the interpretability problem and factors that are expected to influence the trade-off between accuracy and explainability. In Section 3, we formulate hypotheses. We describe the research method in Section 4. Section 5 reports on the results and Section 6 concludes the paper.

2. Literature Review

2.1. The Interpretability Problem of ML Models

The interpretability problem describes the inverse relationship between the accuracy and explainability of ML models (Barredo Arrieta et al., 2020). Explainability is needed for justification, improvement, and knowledge discovery (Adadi & Berrada, 2018; Bracci, 2023). However, the most accurate ML models usually are not very explainable (Adadi & Berrada, 2018; Huysmans et al., 2011). In particular, for large data sets and complex problems, the highest accuracy scores are often only achieved by models that even experts have problems to interpret. Gunning and

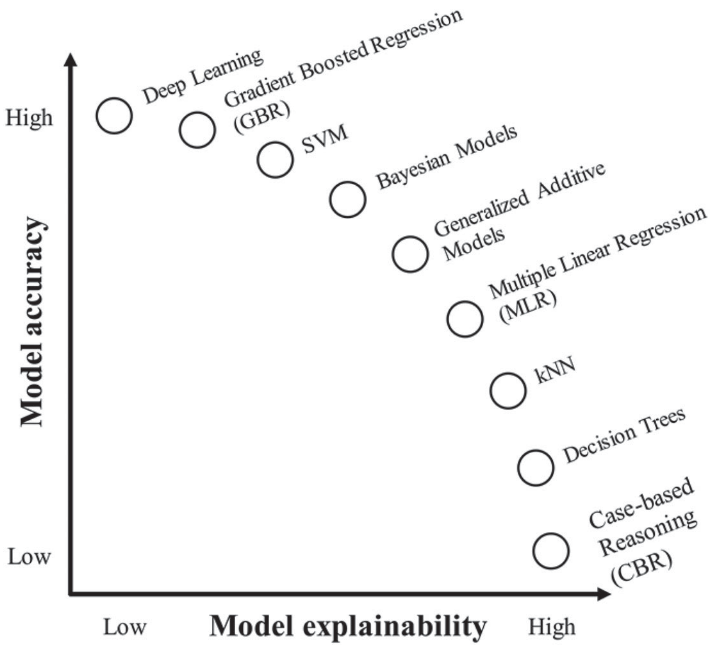

Figure 1. A conceptual model on the trade-off between model explainability and accuracy. Adapted from Barredo Arrieta et al. (2020)

Aha (2019) and Barredo Arrieta et al. (2020) conceptualized the trade-off between explainability and accuracy of ML models along these dimensions, see Figure 1.

Research aims to solve the interpretability problem with explainable AI (Adadi & Berrada, 2018; Guidotti et al., 2019; Hall & Gill, 2018; Preece, 2018; Zhang et al., 2022). This comprises methods to maintain the high accuracy levels of complex ML models while improving the comprehensibility of models through additional explanations. Nonetheless, these techniques usually cannot completely solve the interpretability problem and a trade-off remains (Alonso & Magdalena, 2011; Barredo Arrieta et al., 2020; Emmert-Streib et al., 2020; Linkov et al., 2020). Thus, data scientists, including management accountants, need to choose between ML models according to the accuracy and explainability needs in a given situation and task (Boehm et al., 2019; Tripathi et al., 2021).[1]

Empirical studies of how people prioritize explainability versus accuracy are lacking, however. Mainly conceptual papers suggest factors that would make explainability or accuracy relatively more important for the ML model selection process. Explainability could be more important when the objective is to integrate knowledge of human experts in the ML system (Alonso et al., 2015; Frias-Martinez et al., 2005; Ylinen & Ranta, 2024) or when an ML model needs to be validated against prior knowledge of human experts (Alonso et al., 2015; Commerford et al., 2022).

[1] A related but different stream of literature addresses algorithm aversion. Humans tend to discount advice that is provided by algorithms and rely more readily on human advice, even if they receive feedback that the algorithms are more accurate. Algorithm aversion may be stronger, for example, if the computer-generated input suggests bad news in light of business goals people have (Chen et al., 2022). In auditing, it may be stronger if the computer-generated input is contradicting management's estimates, especially when these are based on more objective data sources (Commerford et al., 2022). In the context of the accuracy-explainability trade-off, however, there is no comparison of humans versus algorithms, but only a comparison of different algorithms and the prioritization of two attributes of these.

Explainability is also likely to be more important when there is low trust toward artificial intelligence systems and there is the need to convince recipients of outcomes (Adadi & Berrada, 2018; Alonso et al., 2015). Furthermore, in the case of legally and ethically relevant tasks, explanations are required and black-box models may not be acceptable (Bonsón et al., 2021; Cho et al., 2020). Among the few empirical studies we could identify, Baryannis et al. (2019) found that significant costs of inaccurate predictions can place more importance on accuracy relative to explainability, and Rey et al. (2017) found that the relevance of rules in rule-based systems can have an important role in the accuracy-explainability trade-off. Thus, more empirical studies of the importance of explainability versus accuracy are needed.

2.2. Cost Estimation in New Product Development

Much of the research on cost estimation during product development is of a technical nature and addresses cost estimation methods (Niazi et al., 2006). Recent research also developed cost estimation methods based on ML, which is a very promising application of ML models (Bodendorf et al., 2021; Kadir et al., 2020; Relich & Pawlewski, 2018; Voltolini et al., 2019).

Empirical research of cost estimation in new product development is much rarer. Questions such as which cost estimation methods are used, on what does this depend, or which experiences do users have with various kinds of models are hardly being explored. Some teaching cases in accounting provide empirical insights, such as Dikolli and Sedatole (2004) and Román (2011), and several case studies explicitly address the role of cost estimation in product development, such as Nixon (1998) and Li Qian and Ben-Arieh (2008). Cost information can guide decisions made during product development (Booker et al., 2007; Hertenstein & Platt, 2000), not only within product development projects of a single organization but also when organizations collaborate to co-develop a product (Cooper & Slagmulder, 2004; van der Meer-Kooistra & Scapens, 2015).

However, cost estimation is difficult during product development. Much information about product design are still unknown or not stable (Florén et al., 2018; Kim & Wilemon, 2002). Also, costs may be affected by complex interactions between products (Davila & Wouters, 2004). For example, when designers choose to use the same part for several different products, manufacturing costs may decrease due to economics of scale, manufacturing costs may increase because of over-specification, development costs may decrease because fewer parts have to be developed, but greater complexity may increase the development costs of these parts (Labro, 2004). The cost impact of approaches such as parts commonality or product modularity can be difficult to estimate (Jørgensen & Messner, 2010).

We focus on the use of ML models for cost estimation in the context of new product development for several reasons. ML models could be very suitable for cost estimation (Cavalieri et al., 2004; Loyer et al., 2016). Cost estimation is particularly relevant to inform decisions taken during product development because there are still many degrees of freedom for managing costs through product design decisions. However, cost estimation is also especially challenging in product development, due to lacking data about the product design and other unpredictable and potentially changing circumstances (Newnes et al., 2008). Thus, product development could be an area where management accountants would want to improve their cost estimation methods and potentially use ML models. So far, however, empirical research on the use of ML models in this context is limited.

3. Hypothesis Development

3.1. Task Uncertainty and the Importance of Explainability Relative to Accuracy

We discuss task uncertainty as the basis for the importance of explainability relative to accuracy, which refers to the significance and weight a decision maker attaches to the attributes explainability and accuracy of an ML model relative to each other. Essentially, when task uncertainty is large, cost estimation triggers questions and there is probing about *why the model produces particular results*. Explainability is important to understand what is driving the results, to integrate knowledge of human experts in the model, and to improve predictions (Alonso et al., 2015; Frias-Martinez et al., 2005). Accuracy is also important for cost estimation, but improving accuracy would not always be a priority. Reasonably accurate results are sometimes good enough for decisions, or attempts to improve accuracy are futile and just leading to spurious accuracy (Niazi et al., 2006).

Task uncertainty has been defined as the perceived difference between the information that is needed and that is available for a particular task (Chapman, 1997; Davila, 2000; Galbraith, 1973). This may not just be a matter of the decision maker knowing exactly what information they require and being able to establish how much is missing. Beyond that, the decision maker may find it difficult to assess the quality of information that is available or to specify which information they require. That implies the decision maker would not be sure which relevant information is missing and how important that missing information would be.

Decision makers can undertake activities to gather more information that is known to be needed for a decision, but also activities to understand the relevance and quality of the available information, to identify additional potentially relevant information, to explore data sources, and to produce more information (Chong & Johnson, 2007; Williams & Seaman, 2002; Ylinen & Gullkvist, 2012). Task uncertainty generally implies that more job-relevant information must be processed during task execution to achieve a particular level of performance (Galbraith, 1973). Looking at this in the context of using ML models, we propose that greater task uncertainty makes people want to interact with the ML model more, to understand the results and the reasons behind those results, as well as to further develop and improve ML models. Need for interaction with the ML model would make explainability relatively more important.

Task uncertainty is often large in product development activities (Chenhall & Moers, 2015; Davila et al., 2009) so the need for interaction with the ML model used for cost estimation may be considerable. Cost estimation can often only be approximate because data are missing and the impact of factors that are influencing product costs can only be partially modeled (Bodendorf et al., 2021; H'mida et al., 2006). Domain experts need to understand and be able to verify the analyses and results the ML model provides (Barredo Arrieta et al., 2020). They want to become confident that the model is appropriate for their cost estimation task. That would be even more important, if the cost estimation data are heavily discussed in product development teams and/or with senior management, and the cost management experts should be able to explicate the analyses and results. To put themselves in that position, cost management experts would want to interact with the model to evaluate the impact of varying assumptions on the cost estimation results. They would also compare the results to reference points and analyze differences, such as the known costs of other comparable products, target cost values for the new product that is being developed, or experience-based cost estimates that they or other human experts could provide. They may also want to adjust the results that are based on ML models to reflect expert knowledge about cost behavior they and others find relevant, but that the model seems to not have considered. Moreover, the experience in working with the ML models for cost estimation could lead the cost management experts to modify, further develop, and improve these models (Bodendorf et al., 2021).

3.2. Hypotheses

People are confronted in various ways with task uncertainty in their work and the study was designed to account for various circumstances in which task uncertainty arises in product development tasks. This would create a greater need to interact with the ML model for cost estimation in that context, which we expect to influence the relative importance of explainability and accuracy. We draw on practice-defined variables for representing task uncertainty in product development tasks. Practice-defined variables have the advantage they capture management accounting phenomena practitioners want to understand, in practitioners' own language. However, it is important to clearly define their underlying theoretical properties (Luft & Shields, 2003; Herschung et al., 2018). The practice-defined variables in this study are based on intense discussions with the case organization about their product development processes and the issues they were facing. They used very similar expressions during these conversations and our terminology was clearly recognizable for them. At the same time, these variables are well-known in the innovation management literature (e.g., Ansari et al., 2006; Ax et al., 2008; Davila et al., 2009; Gopalakrishnan et al., 2015; Henri & Wouters, 2020; Janka & Guenther, 2018).

3.2.1. Project unpredictability

Project unpredictability here refers to limited availability and stability of the information concerning the product that needs to be developed and the conditions for the product development project. Unpredictability during new product development stems from various sources such as unforeseen changes regarding environment, technology, consumers, competition, and resources (Song & Montoya-Weiss, 2001). It makes intensive communication in teams more important and requires more information processing (Tushman & Nadler, 1978). Projects with high unpredictability 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). In unpredictable settings, product development is characterized by experimental and iterative problem-solving approaches (Brown & Eisenhardt, 1995; Eisenhardt & Tabrizi, 1995). Project unpredictability implies that there will be more changes happening in the product description (product specifications and product target cost, for example) and the conditions for the NPD project (such as lead time and development budget). Therefore, there will be a greater need to understand how and why such changes impact cost estimates, which implies more interaction with the ML model.

Furthermore, project unpredictability means that more assumptions will have to be made about unknown future events, as well as more variations of assumptions to engage in contingency planning (Allaire & Firsirotu, 1989), such as assumptions about the benefits of new products (Berchicci & Bodewes, 2005). Furthermore, cost estimates often require many assumptions to create leeway for satisfying technical requirements and organizational objectives (Tyebjee, 1987). Understanding the impact of assumptions and alternative assumptions on the cost estimation results will be important, which also creates more need for interaction with the ML model.

On the other hand, project unpredictability could make accuracy less of a priority. When many elements of the product and project development project are unknown or unstable, cost estimation is understood to be quite inaccurate. Trying to improve it may seem important, but when the basis is formed by arbitrary assumptions and changing information, such attempts are only creating spurious accuracy. We expect that receivers of cost estimates have lower expectations about accuracy when dealing with unpredictable situations, but understanding assumptions and how the model works are more important to them. In contrast, we expect that accuracy is more important if project information is known and stable, when cost experts can rely on facts rather than assumptions.

In sum, task uncertainty arises in this context through the unpredictability of new product development projects, making it more important for the cost management expert to interact with the ML model, which increases the importance of the explainability of the model. At the same time, unpredictability makes accuracy difficult to fundamentally improve.

H1: The importance of explainability relative to accuracy is greater when project unpredictability is high than when this is low.

3.2.2. Product cost granularity

Costs can be estimated at different levels of granularity, ranging from the costs of entire products to individual parts. At more detailed, granular levels, people have specific technical product development tasks for particular components. The development teams, which often comprise people from different departments, work with specific, homogenous sets of few components. A detailed decomposition of cost supports these development teams (Filomena et al., 2009). The impact of a team's design choices on the cost of the components they are working on may be quite straightforward to understand and not require much probing into the results the ML model is producing. Accuracy would be more important relative to explainability for these teams.

At more aggregated levels, people have management tasks for the overall product. Product development involves a heterogeneous set of many components, the optimization of the overall concept of a product, management of development teams, and lateral coordination between these teams. For example, the team that is responsible for the costs of the entire product must understand the reasons for too high costs, consider product changes, and discuss actions with several teams. This coordination task requires interaction with the ML model, such as probing into the cost estimation results, understanding what the model is doing and why, going through alternative design choices that affect several parts throughout the entire product that different teams are working on, understanding cost interdependencies between various kinds of parts, and assessing the validity of results. Such interaction with the ML model makes explainability more important.

Accuracy would likely be less of a concern for a more aggregate cost level because cost estimates are likely to be more accurate at a more aggregate level. Aggregation in statistical ML models often creates some canceling out of random estimation errors. This is consistent with Tseng et al. (2005) who find that in the conceptual design work that concerns the entire product, breadth of data is more important than precision, while during the detailed design of separate parts, data need to be precise.[2] Thus, we expect a higher importance of explainability relative to accuracy when estimating less granular costs (i.e., costs of a complete product) compared to more granular costs (i.e., costs of subcomponents).

H2: The importance of explainability relative to accuracy is greater for a complete product (less granular) cost estimation than for a subcomponent (more granular) cost estimation.

[2]When it comes to forecasts made by *humans*, aggregation may have very different effects. People find it difficult to consider many different factors when creating a forecast. Disaggregating a forecast enables them to consider distinct factors in separate forecasts, such as repeat customers, promotions, or various kinds of external events that are relevant for different sources of demand. This can increase forecast accuracy, see Brüggen et al. (2021).

3.2.3. Predecessor product availability

The availability of a predecessor product implies low product novelty (Horn & Salvendy, 2009; Tatikonda & Rosenthal, 2000) and incremental innovation (Henderson & Clark, 1990; Reid & de Brentani, 2004). This is characterized by changing existing products (Abdul Ali et al., 1993; A. Ali, 1994) and often relies on internal information (Herstatt et al., 2004). Individual components are refined, while the core designs of components and the overall architecture remain the same. Incremental innovations are targeted toward existing markets (Reid & de Brentani, 2004). 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). Project teams must be familiar with the work done in the past, to successfully establish products with incremental innovation (Patanakul et al., 2012).

For cost estimation, this means it is important to be knowledgeable about the technical background of products and to understand the cost estimates for the current product in relation to the predecessor product. Accordingly, we expect that in-depth comparisons with the predecessor product are being made. The existence of cost data for the predecessor product creates a need for interaction with the ML model to understand the current cost estimates for the new product in relation to the cost data for the predecessor product. Which costs differences are significant and why is that the case? What is the cost impact of different technical choices compared to the predecessor product? The greater need for interaction with the ML model in case a predecessor product is available makes explainability more important.

Accuracy may be less of a concern and priority if a predecessor product is available because achieving a high level of accuracy with all ML models is already more feasible. On the other hand, if no predecessor product is available, probing into cost differences is not relevant, but achieving a reasonable level of accuracy is challenging and a priority.

H3: The importance of explainability relative to accuracy is greater when a predecessor product is available than when this is not available.

3.2.4. Target cost gap

The target cost gap concerns the difference between the cost goal and the estimated cost of the current product design, which is still under development. We expect this to matter for the relative importance of explainability and accuracy. A larger target cost gap makes it more demanding to develop a product that meets the cost goals as well as all other product requirements. Finding measures for decreasing or even closing the cost gap can be a complex task, leading to discussions on the interactions between product characteristics, product design, and sales prices. Based on the technical requirements for particular product features, relationships between costs and functionalities are analyzed, and non-essential functionalities may be sacrificed in order to afford more important components (Thore Olsson et al., 2018). We expect that a larger cost gap implies higher cost pressure for the project that will demand more explanations and discussions with higher management levels. This creates a greater need for interaction with the ML model, making explainability more important.

Furthermore, a number of cost management methods, such as value engineering, functional analysis, quality function deployment, and design for manufacture and assembly, help to reduce costs during product development (Al-Qady & El-Helbawy, 2016; Ax et al., 2008). Using such methods requires detailed information (Everaert et al., 2006) and much interaction with the ML model, again making explainability more important.

Accuracy may be less of a concern when the cost gap is large. A large cost gap is problematic, regardless of whether this can be estimated very accurately or reasonably accurate. Information about a large cost gap is actionable, even if the exact size of the gap is less accurately estimated.

Accuracy is more important when the cost gap is small because on the basis of the cost estimate, it would be concluded that targets are (almost) achieved and (almost) nothing more needs to be undertaken to reduce costs. Information about a small cost gap is only actionable, if that cost gap can be accurately estimated.

H4: The importance of explainability relative to accuracy is greater when the target cost gap is large than when this is small.

3.2.5. Product development phase

Finally, we include the product development phase. The product development process can be divided into the early and the late phase (Kim & Wilemon, 2002). The early phase includes activities such as product strategy formulation, ideation, product definition, and project planning (Khurana, 1998), when companies develop product concepts and decide whether to invest in the development of these products (Moenaert et al., 1995). Since many technical details of the product have not yet been determined, development teams build on assumptions (Hey et al., 2007) and there is often just a rough cost estimate based on the experience of cost experts (Rush & Roy, 2000). In the late phase, the product concepts are taken to the next level and actual technical realizations are being devised. Detailed cost estimates are available, using computer-based tools to calculate manufacturing costs in greater detail (Bode, 2000).

Thus, the early product development phase creates task uncertainty, making explainability important in order to better understand the financial state of the entire project and the impact of specific assumptions. Although the early product development phase creates task uncertainty in a somewhat comparable way to the more holistic notion of project unpredictability in H1, notably different is the circumstance that the early product development phase offers time to still react and to change course later in the project. This situation could make accuracy even less of a concern and the prioritization of explainability of accuracy even more pronounced than in case of the more general project unpredictability.

H5: The importance of explainability relative to accuracy is greater in the early phase of new product development than in the late phase.

4. Research Method

A central motivation for the research design was involving expert research participants to consider the use of ML models for a realistic cost estimation task in product development. Several ML models that are applicable to this task were included and the accuracy of these models was measured using company archival data. In Section 4, we discuss the research participants, the task, the included ML models, the measured accuracy of these models with data of AutomotiveCompany, measurement of the dependent variable, the detailed steps, and the analysis of the qualitative data.[3]

[3]We collaborated with AutomotiveCompany on several research topics on the interface of product development and management accounting. Research topics were proposed by the researchers, considering the relevance for our research goals, and agreed by the case company. This included an in-depth, exploratory case study on the use of machine learning models for cost estimation during product development (Hammann, 2024). The specific question investigated in the current paper, as well as the research method, was entirely driven by our own research interests. Participation of individual management accountants was without any pressure or incentives from the case company and achieved by the researcher formulating criteria for participants and approaching people in order to achieve a stratified sample.

4.1. Research Participants

The research participants came from one organization (AutomotiveCompany, a disguised name) and the task was tuned to their company context. Forty cost experts from the product controlling department of AutomotiveCompany were included based on stratified convenience sampling. They had work experience in product controlling of at least three years, their average total work experience was 10.57 years (SD 6.80), and 35 (88%) of the participants were male. Their average ML experience was 2.00 (SD 1.13) on a 7-point Likert scale. Accordingly, most participants can be regarded as ML novices (Mohseni et al., 2021). The cost experts worked in various sub-departments as complete vehicle controllers (10), assembly group controllers (10), or parts controllers (20). The Appendix offers details on the measurement of work experience and ML experience. Section A of the Online Appendix provides details of the stratified sample and an overview of all participants. We include work experience, ML experience, gender, and sub-department as control variables.

4.2. Task and Conditions

At AutoCompany, the participating cost experts were involved in estimating the cost impact of product design changes (for example, changing a product feature such as a head-up display or panoramic roof). Accordingly, participants were 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.* We created treatments with five factors that produce task uncertainty in product development, summarized in Table 1. The clarity and recognizability of the descriptions of these treatment levels were verified with the company.

The five factors at two levels potentially create $5^2 = 32$ treatment combinations. To avoid participant fatigue, we reduced this to the $2^{5-1} = 16$ treatment combinations shown in Table 2. Specifically, the treatment combinations in Table 2 are of a so-called Resolution V fractional factorial design, which enables estimating direct effects and two-way interactions, but not three-factor interactions (or higher) (Margolin, 1969). This disadvantage is not relevant for our study.

Each participant conducted the task for all the 16 treatment combinations in a single session. The order was randomized for each participant. This design offered several advantages for this study. First, it enabled doing the study with expert participants. This is a strength of this study, but it also means the number of research participants was limited. Therefore, each participant conducted the task multiple times. This was also the reason for having a fractional factorial design: doing the task 32 times would be too much, but we considered 16 times feasible. Second, we used the awareness of the treatment combinations as a means to obtain additional qualitative information by discussing the variations of the task with each expert participant. More details on the qualitative part of the study are provided in Sections 4.5 and 4.7. The task was conducted by screen-shared audio conferences. The language was German because the participants were German native speakers. The mean duration of the session was 61 min (SD 8.51).

The sessions took place over a period of 84 days. Research participants did not have knowledge of who the other participants were, but if they happened to know, they could communicate. This was not a problem, because the ML models and treatment combinations were transparent to all participants anyway. However, we did explicitly request participants not to talk about the sessions and in particular not about their ideas regarding the importance of explainability and accuracy.

4.3. ML Models for the Cost Estimation Task

We selected three ML models for the cost estimation task, see also Section B of the Online Appendix. The three ML models vary in explainability and accuracy (see Table 3) such that the case-based reasoning (CBR) model is the most explainable and the least accurate, while the gradient boosted regression (GBR) model is the most accurate and least explainable, with multiple linear regression (MLR) falling in the middle between the two extremes. These assumptions were tested using firm data and the results are in Table 3, showing that the order of the measured accuracy of the three ML models is consistent with the expectations. More information for these tests is in Section C of the Online Appendix.

Case-based reasoning (CBR) is particularly strong regarding explainability. It essentially estimates costs based on analogies with similar products, whereby the ML model identifies the characteristics to find comparable products. Neither complex algorithms are deployed, nor complex associations between dependent and independent variables are assumed (Duverlie & Castelain, 1999). However, the effectivity of CBR highly depends on the similarity with past cases (Roy, 2003) and CBR is limited in terms of inter- or extrapolating costs and considering interdependencies between product features. We expect high explainability but low accuracy for the CBR model for product cost estimation.

Multiple linear regression (MLR) is easily understandable on the basis of the regression coefficients (James et al., 2013), making the prediction 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 ML models (Loyer et al., 2016). We expect medium explainability and medium accuracy for the MLR model.

Table 1. Contrast coding of the independent variables (treatments)

Variable	Code	Level	Description
Project Unpredictability (H1)	−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
Product Cost Granularity (H2)	−1	product	Your objective is the estimation of complete vehicle costs
	+1	component	Your objective is the estimation of costs on assembly level (engine, chassis, body, or electrics)
Predecessor Product Availability (H3)	−1	no	Your objective is to estimate the cost of new car without a predecessor
	+1	yes	Your objective is to estimate the cost of a car with a predecessor
Target Cost Gap (H4)	−1	small	The estimated costs of the current design surpass the target costs by 5%
	+1	large	The estimated costs of the current design surpass the target costs by 20%
Product Development Phase (H5)	−1	early	The task is conducted 66 months before SOP
	+1	late	The task is conducted 45 months before SOP[a]

[a]45 months before SOP is late in the product development process in automotive since many decisions on product design and agreements with suppliers are then fixed. 66 months before SOP is around two years earlier and this point in time, many more degrees of freedom – as well as uncertainties – exist, which is why this can be considered early.

Table 2. 2-level fractional factorial design with 2^{5-1} treatment combinations

Treatment combination	Project unpredictability	Product cost granularity	Predecessor product availability	Target cost gap	Product development phase
1	Low	Product	No	Large	Early
2	Low	Product	No	Small	Late
3	High	Product	No	Small	Early
4	High	Product	No	Large	Late
5	Low	Component	No	Small	Early
6	Low	Component	No	Large	Late
7	High	Component	No	Large	Early
8	High	Component	No	Small	Late
9	Low	Product	Yes	Small	Early
10	Low	Product	Yes	Large	Late
11	High	Product	Yes	Large	Early
12	High	Product	Yes	Small	Late
13	Low	Component	Yes	Large	Early
14	Low	Component	Yes	Small	Late
15	High	Component	Yes	Small	Early
16	High	Component	Yes	Large	Late

Table 3. Accuracy and explainability of the selected machine learning models

	Expected		Measured accuracy[b]	Perceived[c]	
ML model[a]	Accuracy	Explainability	NMAE	Accuracy	Explainability
CBR	Low	High	13.84 (5.12)	3.25 (1.26)	6.69 (0.38)
MLR	Medium	Medium	10.96 (4.51)	4.58 (1.01)	6.08 (0.79)
GBR	High	Low	8.77 (4.77)	5.85 (0.95)	2.67 (0.95)

[a]Machine learning (ML) models gradient boosted regression (GBR), multiple linear regression (MLR), and case-based reasoning (CBR).
[b]Estimation of the costs for the subsequent product generation. The predictive accuracy is measured by the Normalized Mean Absolute Error (NMAE), so a lower value indicates greater accuracy. Means and standard deviations (in parentheses) over car configurations.
[c]Perceptions are measured with 7-point Likert scales. Means and standard deviations (in parentheses) over all participants.

Gradient boosted regression (GBR) is essentially a combination of several simpler models, such as multiple regression models and decision trees. However, it is difficult to obtain explanations from such complex ensembles, which is why 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, explainability is expected to be low, but accuracy high.

4.4. *Measurement of the Relative Importance of Explainability to Accuracy*

The dependent variable of is the relative importance of explainability to accuracy (*EtA*). We asked participants after each task about the perceived relative importance of explainability to accuracy on a 10-point bipolar Likert scale (EtA^S) that ranged from 10 (explainability is most important) to 1 (accuracy is most important). See the Appendix with this paper. Results based on this direct measurement are used for testing the hypotheses.

We also measured the relative importance of explainability to accuracy indirectly. The indirect measurement was based on model selection (EtA^M), following Agarwal and Prasad (1998). At the start of a session, the participant rated three ML models according to their perceived explainability and accuracy, see the Appendix with this paper. For explainability, our measurement scale was based on the items of Gedikli et al. (2014), Berkovsky et al. (2017), Piltaver et al. (2016) and Demoulin and Coussement (2020). Next, each time they conducted the cost estimation task, the participant chose one of the models to solve the task. The relative importance of explainability to accuracy for that specific treatment combination was derived from the selected model and that particular participant's perception of the explainability and accuracy of the selected model. Because this indirect measurement is more complex and requires more assumptions, we did not rely on it for hypothesis testing, but we included it for comparing results, because a direct measurement may not always be suitable or feasible.

4.5. Procedure

The study consisted of 10 steps that are shown in the Appendix with this paper. These steps cover participant instruction, task completion, and discussion. *Participant instruction* comprised the first seven steps: (1) We gathered some participant data. (2) We provided the participant with explanations of the CBR, MLR and GBR models, which can be found in Section D of the Online Appendix. This was necessary to ensure everyone had the same level of understanding of the ML models. (3) We measured whether the participant had understood the information by asking them to rate the explainability of the models using the first three items. (4) The participant watched short video clips about the three ML models that we had created with a demonstration tool. (5) The participant rated each model using the next two explainability items. (6) We provided the participant with the measured accuracy results for each machine learning model based on the archival data of AutomotiveCompany. (7) The participant rated the accuracy of each model.

For *task completion*, (8) we provided the participant with an explanation of the task. (9) The participant was provided with 16 vignettes that reflected the combinations of the 5 uncertainty factors as laid out in Table 2. Section E of the Online Appendix provides an example of one of the vignettes that was used. Each time, the participant rated the relative importance of explainability to accuracy on a Likert scale (EtA^S) and by choosing one of the three models they would use for that combination of the five factors (EtA^M).

In the *discussion*, (10) we interviewed the participant about their choices to better understand why they made the decisions they did. For each vignette, we asked whether explainability or accuracy would be more important to them, and why that would be so. We probed to understand their reasons and asked for specific examples. Since the company did not allow the recording of these comments, detailed notes were taken by the researcher during and directly after the discussion.

4.6. Regression Analysis

We estimated regression models to test the hypotheses based on 640 observations: 40 participants $\times$ 16 treatment combinations. Each of the 640 rows in the data table included the 2 dependent variables EtA^S and EtA^M, 5 manipulated independent variables, a participant identification number, and a few measured participant variables (sub-department, work experience, ML experience and gender). These measured variables were the same for each participant across the 16 times they participated. Thus, the 640 rows were not completely independent, as would be the case if 640 different people had participated. We accounted for the fact that each participant generated

16 observations by estimating random effects regression models:

$$\begin{aligned} EtA_{it}^{S} = {} & \beta_0 + \beta_1 ProjectUnpredictability_{it} + \beta_2 ProductCostGranularity_{it} \\ & + \beta_3 PredecessorProductAvailability_{it} + \beta_4 TargetCostGap_{it} \\ & + \beta_5 ProductDevelopmentPhase_{it} + \beta_6 WorkExperience_i \\ & + \beta_7 MachineLearningExperience_i + \beta_8 Gender_i + \beta_9 CompleteVehicleDepartment_i \\ & + \beta_{10} PartsDepartment_i + \alpha_i + \epsilon_{it} \end{aligned}$$

where: i: participant, t: treatment combination, α_i: participant-specific random effect for participant i, ϵ_{it}: error term. Standard errors are clustered by participant by adjusting the standard errors using a clustered covariance matrix.

Without control variables, all individual effects are captured by the participant-specific effect that is estimated for each of the 40 participants (α_i), as a kind of individually adjusted intercept. The control variables in this study are also individual effects because they have the same value for all observations generated by one individual. The random effects model allows for including control variables in this case. The control variables refine the participant-specific effects on the basis of a participant's scores on the individual control variables (through β_6, β_7, β_8, β_9, and β_{10}). These individual effects of the control variables do not affect the slopes of the main effects (β_1, β_2, β_3, β_4, and β_5), although the level of statistical significance sometimes may vary between the models with and without the control variables, because the controls and variables of interest are orthogonal by design in this specific case.

4.7. Analysis of Qualitative Data

We conducted content analysis using the tool QDA Miner and combining a deductive and inductive approach. We prepared a table summarizing our arguments for each of the hypotheses and we deductively coded the participants' statements on the basis of these arguments and the keywords of the hypotheses. We compared the explanations provided by the participants with our arguments for the hypotheses. In addition, when a particular statement could not be deductively coded, we created an additional code for clustering statements. These might potentially provide new ideas. These codes were split into possible additional arguments for the hypotheses and possible explanations for unexpected results. We then pruned the long lists of clustered participants' statements to what we considered to be the most interesting arguments or explanations.

5. Results

5.1. Perceived Accuracy and Explainability

We measured whether the participants had understood the information provided during the participant instruction steps. We expected an increasing level of perceived accuracy (CBR < MLR < GBR) and the average scores of perceived accuracy shown in Table 3 are consistent with this. We expect a decreasing level of explainability over the three ML models (CBR > MLR > GBR) and the average scores of perceived explainability for the three models shown in Table 3 are also consistent with this. The Cronbach's alpha of the five items of explainability is 0.963. We also calculated the intraclass correlation coefficient (ICC) for interrater reliability (Aguinis et al., 2018). An ICC-value of higher than 0.75 is referred to as good reliability and a higher than 0.9 to excellent reliability. For the explainability rating over the

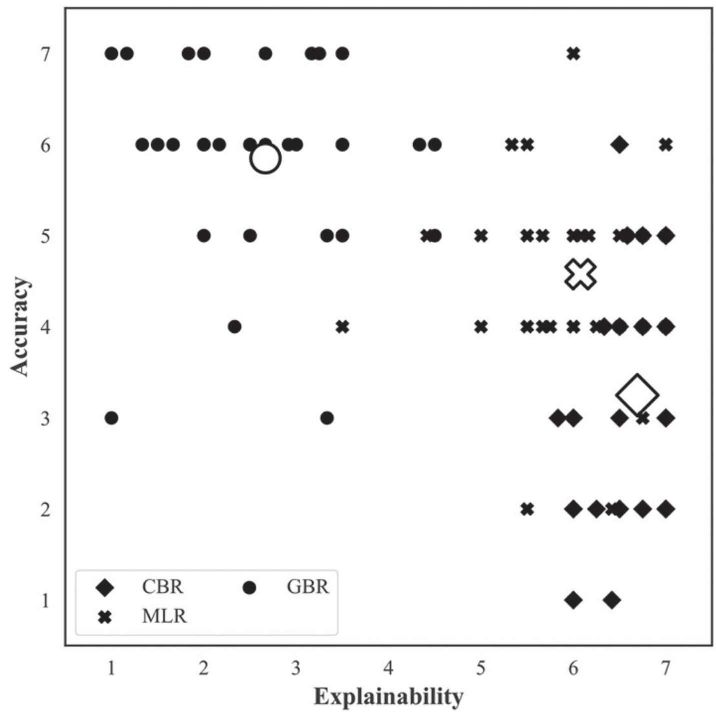

Figure 2. Perceived levels of explainability and accuracy of the three models: gradient-boosted regression (GBR), multiple linear regression (MLR), and case-based reasoning (CBR). As part of the participant instruction steps, the 40 participants rated each of the three models regarding explainability and accuracy, yielding 120 observations. The scatter plot suggests an inverse relationship between perceived explainability and accuracy

three models and 40 participants, the ICC is 0.915 (F-statistic 432.13, p-value $< .001$) and for accuracy the ICC is 0.863 (F-statistic 253.74, p-value $< .001$).

We analyzed the assumption about the inverse relationship between perceived accuracy and explainability of the three models. The 120 observations (40 participants $\times$ 3 models) are plotted in Figure 2. The association between perceived explainability and accuracy follows the expected concave function, supporting that some models perform primarily on accuracy, while others perform mainly on explainability. Furthermore, we tested for fatigue and find this is not a problem; see the figures and analyses in Section F of the Online Appendix.

5.2. Hypotheses Testing

Overall, participants tend to select simpler cost estimation methods in most situations. Despite being less accurate, cost experts mostly choose either the CBR (27%) or the MLR model (47%); in only 26% of the cases, the more accurate but also more complex GBR model is preferred. This is consistent with the finding that on average, they consider explainability slightly more important than accuracy: the mean EtA^S over the treatment combinations and participants amounts to 5.68 (SD 2.32) on a 10-point Likert scale.

We find that the perceived importance of explainability relative to accuracy is important when choosing between alternative ML models. The frequency distributions of the selected models for all participants and treatment conditions are depicted in Figure 3. As expected, in situations primarily requiring explainability, CBR is selected; whereas the GBR model is selected in situations which require mainly accuracy. The ANOVA shows significant differences

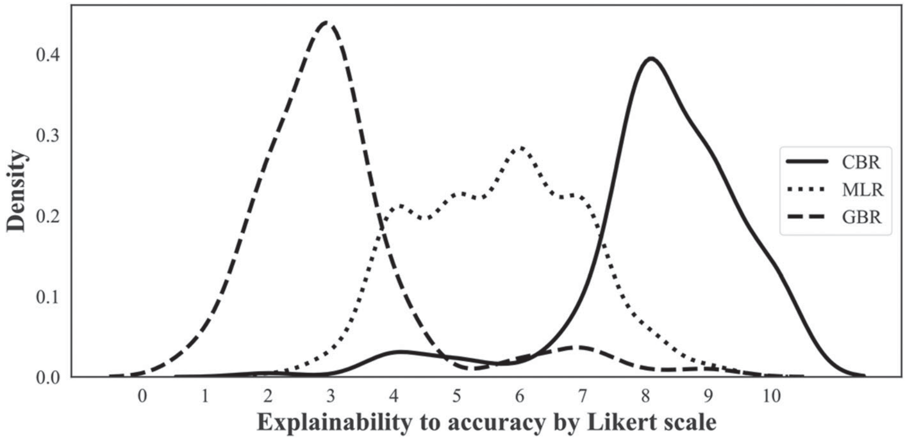

Figure 3. Model selection dependent on the perceived importance of explainability to accuracy EtA^S. The gradient–boosted regression (GBR) model is mostly selected in situations with low levels of perceived explainability to accuracy (accuracy is more important), whereas the case-based reasoning (CBR) model is mainly chosen in situations with high levels of perceived explainability to accuracy (explainability is more important)

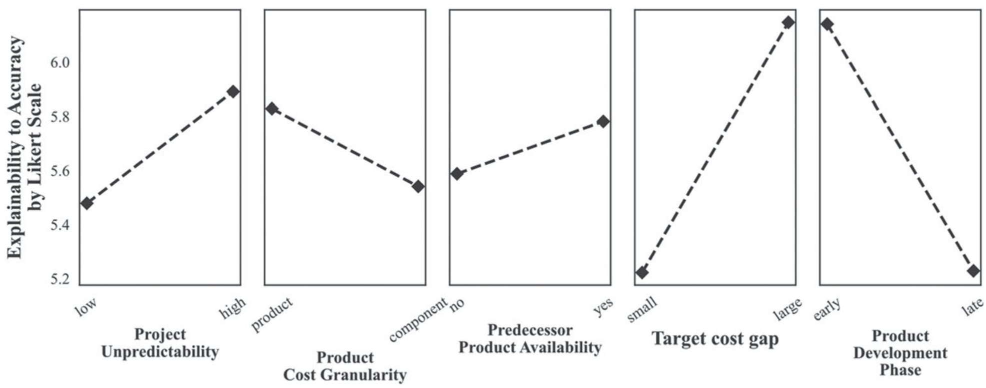

Figure 4. Visualization of differences for the five independent variables. The mean importance of explainability to accuracy, based on the direct measurement on a 10-point Likert scale (EtA^S) over all observations is 5.68. This is differentiated here for the level of each independent variable (e.g., 'low' and 'high'). The figure suggests that explainability is more important than accuracy for *high* project unpredictability, *product* cost granularity, predecessor product availability *yes*, *large* target cost gaps, and an *early* product development phase

(F-statistic $= 608.72$, p-value $< .001$) of the mean perceived importance of explainability to accuracy (EtA^S) between the three models.

The effects of the five factors are depicted in Figure 4. For each of these factors, we show the mean of the EtA^S score at the two different levels of the factor. The figure suggests that consistent with the hypotheses, explainability is relatively more important when project unpredictability is high, product cost granularity concerns an entire vehicle, a predecessor product is available, the target cost gap is large, and the product development phase is early.

Further descriptive statistics for the control variables (characteristics of the participants) are shown in Table 4. For each of these variables, we show the mean of the EtA^S score in relation to the different levels of these variables. ML experience or sub-department do not appear to have an effect. EtA^S does differ for the three groups with different levels of work experience, but without a clear relationship. Only for gender is the difference statistically significant, suggesting that male participants place greater importance on explainability relative to accuracy.

We used a random effects regression model to test the main effects of the five independent variables. We evaluated the assumptions of linear regression with binary independent variables. First, we checked the assumption of normality with the aid of quantile-quantile-plots and find no abnormalities (Section G of the Online Appendix). Second, residual plots of both variable measures suggest homoscedasticity of residuals (Section G of the Online Appendix). Third, Levene's test over the 16 treatment combinations confirms equal variances between the treatment combinations (p-value = .427 for EtA^S and .740 for EtA^M).

The results of the regression analysis are summarized in Table 5. We estimated models with and without control variables, which makes no difference for the substantive results. The signs of the coefficients are mostly in line with the hypothesized directions and the descriptive effects in Figure 4. As expected in H1, project unpredictability is positively and statistically significantly associated with the importance of explainability relative to accuracy. Greater unpredictability makes explainability more important. Supporting H2, product cost granularity is negatively related to the importance of explainability relative to accuracy, but the effect is only marginally statistically significant. For a complete vehicle (less granular cost estimation), more importance is placed on explainability. H3 is not supported. The negative coefficient is opposite to the hypothesis, but this is not statistically significant. The availability of a predecessor product does not affect the importance of explainability relative to accuracy. The findings support the positive relationship between target cost gap and the importance of explainability relative to accuracy (H4). A larger cost gap makes explainability more important. Finally, supporting H5, we observe a negative relationship between product development phase and the importance of explainability relative to accuracy. Explainability is more important in the early phase.

The magnitude of the coefficients indicates that target cost gap and product development phase have the strongest relationship with the importance of explainability relative to accuracy, which is also consistent with Figure 4. Furthermore, some interesting differences can be noted. The coefficient for the product development phase (H5) is larger than for project unpredictability (H1). Both factors create task uncertainty because much is unclear about the product and the project. A difference is that the early product development phase explicitly offers some response time when new information becomes available, whereas general project unpredictability may require a very sudden reaction. If there is 'time to think,' explainability could be more important relative to accuracy compared to when immediate action is required and decision-makers may just want accurate data and may be less concerned about the 'why' behind those numbers.

The coefficient for the target cost gap (H4) is very large and statistically significant, whereas the coefficient for predecessor product availability (H3) is not statistically significant. In both situations, the importance of explainability is driven by the existence of reference points, in the form of cost data for a predecessor model (H3) or as explicit targets for the current model (H4). We find that the latter is much more forceful. The need for explainability increases more strongly due to facing a large cost gap than because there is a predecessor model. A reason could be that in case of a large cost gap, it is not only about probing the results to get an understanding of cost estimates in relation to other numbers, but there is also accountability for estimated costs in relation to targets. In other words: The question is not just 'How are cost estimates to be understood compared to a predecessor model?' but the more insistent question is 'How are cost estimates to be understood compared to a cost target, and why are you not achieving that cost target?' This could make it even more compelling to interact with the ML model to understand the cost estimation results.

Table 4. Further descriptive statistics[a]

	Work experience			ML experience[b]			Gender		Sub-department		
	Participants with least work experience (mean 4.7 years)	Participants with median work experience (7.0 years)	Participants with most work experience (mean 17.1 years)	Participants with least ML experience (mean 1.0)	Participants with median ML experience (2.0)	Participants with most ML experience (mean 3.7)	Male	Female	Complete vehicle	Assembly	Parts
# participants	17	5	18	17	13	10	35	5	10	10	20
EtA^S Mean[c] (standard deviation)	5.69 (0.96)	4.46 (0.66)	6.02 (1.36)	5.62 (1.43)	5.58 (1.07)	5.92 (1.06)	5.84 (1.18)	4.56 (0.80)	5.67 (1.04)	5.8 (1.39)	5.61 (1.26)
ANOVA or T-Test	F-statistic 3.663, p-value .035[d]			F-statistic 0.243, p-value .785			T-stat -2.331, p-value .025		F-statistic 0.129, p-value .880		

[a]The basic descriptive statistics for these measured participant variables are provided in Section 4.1. Descriptive information about the manipulated variables (such as Project Unpredictability) are in Table 2, and descriptive statistics about the dependent variable EtA^S are in Section 5.2 and in Figure 4.
[b]Measured on a 7-point Likert scale.
[c]The mean importance of explainability to accuracy, based on the direct measurement on a 10-point Likert scale (EtA^S) over all observations is 5.68. This is differentiated here for the level of each control variable (for the different categories of a control variable, or by splitting the 40 participants in three groups that are below, at, and above the median).
[d]The Anova shows that the mean values of EtA^S are statistically different for these three groups. However, there is no linear association going from least to median to most work experience, which explains why in the regression analysis results in Table 5, the coefficient of work experience is not statistically significant.

Table 5. Random effects regression results (t-statistics in parentheses). The dependent variable is the importance of explainability to accuracy[a]

	Expected sign	EtA^S	EtA^S	EtA^M	EtA^M
Constant		**5.6844*****	**4.5286*****	**1.3861*****	**1.2118*****
		(29.845)	(6.991)	(18.134)	(4.613)
Project Unpredictability high (H1)[b]	+	**0.2063***	**0.2063***	0.0276	0.0276
		(2.261)	(2.252)	(0.984)	(0.980)
Product Cost Granularity component (H2)	−	**− 0.1438**+	**− 0.1438**+	− 0.0388	− 0.0388
		(− 1.805)	(− 1.798)	(− 1.402)	(− 1.396)
Predecessor Product Availability yes (H3)	+	− 0.0969	− 0.0969	**− 0.0915***	**− 0.0915***
		(− 1.095)	(− 1.091)	(− 2.365)	(− 2.356)
Target Cost Gap high (H4)	+	**0.4625*****	**0.4625*****	**0.1192****	**0.1192****
		(3.597)	(3.582)	(2.749)	(2.739)
Product Development Phase late (H5)	−	**− 0.4563*****	**− 0.4563*****	**− 0.1279****	**− 0.1279****
		(− 3.693)	(− 3.679)	(− 2.990)	(− 2.979)
Work Experience			0.0208		− 0.0014
			(0.663)		(− 0.125)
Machine Learning Experience			0.0350		0.0454
			(0.220)		(0.494)
Gender (m = 1, f = 0)			**1.2785*****		0.0463
			(3.394)		(0.278)
Complete Vehicle Controlling Dept.			− 0.3266		− 0.0103
			(− 0.700)		(− 0.069)
Parts Controlling Department			− 0.3435		0.1215
			(− 0.784)		(0.770)
Observations		640	640	640	640
R^2		12.0%	12.7%	7.9%	8.1%

[a]The dependent variable relative importance of explainability to accuracy (*EtA*) is measured with a direct, perceptual method using a Likert scale (EtA^S) and with an indirect method that is based on model selection (EtA^M). Hypotheses are tested on the basis of the first measure.
[b]The coefficients are the same comparing the models without and with control variables (only the t-statistics vary somewhat) because the controls and variables of interest are orthogonal by design in this specific case. The random effects model without the control variables estimates individual effects for each participant. Since the control variables are always the same across all observations for a participant, these are also individual effects in our study. Therefore, including the control variables enables refining the estimation of these individual effects on the basis of a participant's scores on the individual control variables, but the slopes of the main effects remain the same in this specific case.
$^+p < .1$; $^*p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Except for gender, the coefficients of the control variables are not statistically significant. Work experience, ML experience or the sub-department do not affect the importance of explainability. For gender, we find a positive, statistically significant coefficient, indicating that the importance of explainability relative to accuracy is greater for male participants (coded with 1, female participants coded with 0), which is consistent with Table 4.

As mentioned in Section 4, we also measured the relative importance of explainability to accuracy indirectly, based on model selection (EtA^M). Both measures are highly correlated, see Section H in the Online Appendix. Regression results for the indirect measure are in Table 5. Despite the high correlation, several substantive results are different. The signs of the coefficients are similar, but these are typically smaller and less often statistically significant with the indirect measurement. It seems that participants are more outspoken about the importance of explainability relative to accuracy when they are asked directly, compared to when they selected particular ML models. Only the strongest results are substantively the same, namely for H4 and H5. The results for H3 are an exception: the small, negative coefficient that is opposite to H3 is now statistically significant. We will discuss the results for H3 in more detail in the next section.

5.3. Discussion of Qualitative Results

As described in Section 4, we asked participants why they considered explainability or accuracy more important. In this section, we discuss their comments, which reiterate our arguments for the hypotheses or provide additional insights as to why they chose particular ML models.

Some of the comments provide support for H1 that explainability is more important when project unpredictability is high, which increases the confidence in the validity of the quantitative results:

> In case of 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)
> 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)
> There is no need to calculate down on the last penny when everything will look different in two months. (Assembly controller 28)
> 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)

However, an argument contrary to H1 was also suggested:

> 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)

Despite having a low explainability, GBR may be preferred when project unpredictability is high because it is more suitable to efficiently estimate a large number of different product designs. Manual adjustments would often be required when applying the CBR and MLR models, which are not always necessary for the GBR model. If many scenarios or different product designs need to be evaluated, the GBR model may therefore be more practical. This corresponds to Deng and Yeh (2010), who state that ML models are beneficial when conducting design-to-cost approaches if they can quickly update cost estimates when the product design is changed.

H2 stated that explainability is more important at the product level (low-cost granularity). The following examples of participants' comments support this hypothesis:

> 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. (Vehicle controller 13)
> Explainability is important for the reconciliation and comparison with other projects with internal and external benchmarks. (Parts controller 36)
> Since the granularity is higher on assembly level there is the expectation of a higher accuracy due to the reduced complexity. (Vehicle controller 7)

A further insight, supporting H2, is that a minimum level of accuracy always needs to be achieved and explainability becomes less important, if there are doubts about accuracy. This was mentioned as an argument why accuracy is more important at the component level (more granular):

> 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)

More granular cost estimates are typically less accurate (which is also observable in Table C1 in the Online Appendix) and the larger prediction error at the component level was considered relevant in the model selection process. The notion of a minimally required level of accuracy was also mentioned when discussing the role of a predecessor product availability (H3):

> 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)

When no predecessor product is available, choosing a more accurate ML model was considered important due to the relatively low level of accuracy to start with. The effect is consistent with H3. These comments suggest that for the trade-off between explainability and accuracy, a minimum level of accuracy is required to ensure the practical applicability of cost estimates (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 (Qian & Ben-Arieh, 2008) and the error compared to alternative methods (i.e., manual calculations) (Caputo & Pelagagge, 2008). A model with a lower explainability but higher accuracy would be prioritized as long as other models do not achieve a minimally required accuracy.

H3 stated that explainability is more important if a predecessor product is available and the following comments provide support for this:

> 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)
> One is already more accurate in case of successor projects. Explainability is more important here. (Parts controller 16)
> In radical projects, accuracy is more important. Nobody can challenge that I might be wrong in this case. (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)

Some comments suggested another effect of predecessor product availability, contrary to H3. When a predecessor product is not available, there is less information and greater unpredictability, which would make explainability more important (which would counter the hypothesized effect):

> 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)
> 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)

These comments suggest that in the case of a more radical innovation when no predecessor product is available, there is more focus on conceptual work and more assumptions are required that need to be explained (Gudem et al., 2014; Herrmann et al., 2007; Norman & Verganti, 2014). Overall, the qualitative evidence for the effects of the availability of a predecessor product are mixed, which could explain the quantitative results for H3 described in the previous section.

Several comments provide support for H4, which stated that explainability is more important when the target cost gap is large:

> In the case of 20% I need explainability. There are cost workshops where each part and even special equipment is questioned. (Parts controller 18)
> 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)
> 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)
> 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)

These arguments concern *actionability*, so how the cost estimate influences decisions and actions. A large target cost gap raises many questions about what is going on and what can be done, which makes explainability the priority. A small target cost gap makes accuracy more important, because there are fewer questions, but it must be substantiated that the target cost has, indeed (almost) been achieved and little more needs to be done.

An additional argument for H4 could rest on *goal commitment*. With high target cost gaps, explainability becomes more important, because employees accept cost targets less and question the cost estimates more.

> 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 [large] deviations. (Parts controller 3)

When the target cost gap is larger, it is more difficult and requires more effort to reach the cost targets, so goal difficulty is also greater. Goal difficulty can be defined as an increase in the required effort of a given task during a given time period (Bonner & Sprinkle, 2002; Campbell & Ilgen, 1976; Erez & Zidon, 1984; Feichter et al., 2018; Locke & Latham, 2002; Matějka, 2018). There is strong evidence for a positive relationship between goal difficulty and task performance, but only given sufficient commitment of individuals (Erez & Zidon, 1984; Locke & Latham, 2002). Thus, *goal commitment* plays a role in the relationship between goal difficulty and task performance. Studies in accounting also show that the relationship between goal difficulty and performance is often complex (Bonner & Sprinkle, 2002; Feichter et al., 2018; Matějka, 2018). Drawing on these ideas, when the target cost gap is larger and goal difficulty is greater, goal commitment becomes more important. To stimulate goal commitment in this situation, explainability may be more important, so people are more likely to trust the model (Alonso et al., 2015), accept that the target cost gap is actually large, and are more committed to try and close the gap.

Finally according to H5, explainability is more important in the early phase. Several of the participants' comments provide support this hypothesis:

> The early phase is about product ideas where I decide based on rough estimates. (Vehicle controller 24)
>
> In the early phase, I have big levers [for cost improvement], which do not need to be calculated in detail. (Vehicle controller 15)
>
> In the late phase I need accuracy. I need to demonstrate profitability, since I substantiate and materialize the concepts. (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)
>
> In the early 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)

6. Conclusion

This paper investigates the relative importance of explainability to the accuracy of ML models. This is important because this is an inevitable trade-off, but hardly any studies have empirically investigated when people prefer accuracy and are willing to sacrifice some explainability, or when they prioritize explainability at the expense of some accuracy. We focus on ML models that are used in the context of cost estimation during new product development. This is a particularly relevant context for investigating this issue, because cost estimation provides a suitable application of ML models in accounting, and cost estimation is particularly important but also especially difficult during product development. This seems to be an area where accountants would like to improve cost estimation with ML models.

Overall, this paper contributes to the literature by providing mixed-methods evidence on the trade-off between explainability and accuracy of ML models. First, we empirically support that people perceive a trade-off between accuracy and explainability. The cost estimation literature often assumes an inverse relationship between accuracy and explainability with limited empirical evaluation (Bertomeu, 2020; Cavalieri et al., 2004; Loyer et al., 2016; Ranta et al., 2023; Verlinden et al., 2008). Our study shows that both attributes are indeed significantly negatively related, and the relationship matches the relationship suggested by Barredo Arrieta et al. (2020). We also show that this trade-off matters for the selection of ML models. This has been suggested (Adadi & Berrada, 2018; Alonso et al., 2015; Barczak & McDonough, 2003, july 20; Baryannis et al., 2019; Bonsón et al., 2021; Cho et al., 2020; Commerford et al., 2022; Frias-Martinez et al., 2005; Ylinen & Ranta, 2024), but few empirical studies investigated this (Baryannis et al., 2019; Rey et al., 2017).

Second, our study challenges the current research trend of using complex ML models for product cost estimation (Caputo & Pelagagge, 2008; Cavalieri et al., 2004; Chou & Tsai, 2012; Deng & Yeh, 2011; Loyer et al., 2016; Verlinden et al., 2008). We show that despite having a higher measured as well as perceived accuracy, the complex GBR model was preferred in few situations. Although prior research suggests that the applicability of complex ML models in accounting could be restricted by the ability to provide explainable results (Arnaboldi et al., 2022; Bertomeu, 2020; Chye Koh & Kee Low, 2004; Lehner et al., 2022; Ranta et al., 2023), there is almost no empirical research supporting this directly. Also specifically in the cost estimation literature, explainability is expected to be a limiting factor of ML models for users (Cavalieri et al., 2004; Loyer et al., 2016), but empirical evidence is lacking (Verlinden et al., 2008). Our research shows that lack of explainability may indeed be a limitation for ML adoption for cost estimation.

Third, a key contribution of this study to the management accounting literature is the theoretical idea and empirical test of task uncertainty in relation to the use of ML models. Due to greater task uncertainty, more information needs to be processed during task execution to achieve a particular level of performance (Galbraith, 1973). In the context of using ML models, we suggest this means that task uncertainty creates a need for interaction with the ML model, which makes explainability more important relative to accuracy. We identify several factors that would create task uncertainty in the context of product cost estimation during new product development: project unpredictability, product cost granularity, predecessor product availability, target cost gap, and product development phase. We find that several of these are indeed associated with a greater importance of explainability relative to accuracy. Task uncertainty plays an important role in the design and use of management accounting systems (e.g., Chapman, 1997; Chong & Johnson, 2007; Davila, 2000; van Triest et al., 2023; Williams & Seaman, 2002; Ylinen & Gullkvist, 2012). We contribute by highlighting how it may also influence the use of ML models in accounting.

Finally, this study provides a comparison of two measurements of the relative importance of explainability and accuracy. The direct measurement is more straightforward and requires fewer assumptions, but it may not always be feasible or wanted. The indirect measurement followed Agarwal and Prasad (1998) and was based on a participant's perception of the explainability and accuracy of various ML models and that participant's selection of a particular ML model for a particular task. Our results suggest that the substantive results may differ, cautioning the use of the indirect measurement.

As always, the study has limitations. The variables we investigated to capture task uncertainty in the context of cost estimation during product development were quite specific to this context and close to practice. This provided a strength in terms of realism for the research participants, who were actually working on in this context. Practice-defined variables have the advantage that they speak the language of practitioners and are often considered relevant. However, the use of the practice-defined variables can make it more difficult to denote multiple phenomena. Second, the task came very close to the actual work of the research participants, but the case company did not have much experience with ML applications. Therefore, the selection between models might be different in other environments with more experienced ML users. Third, we assumed that the measured accuracy scores of the three models are comparable in each experimental situation (accuracy of CBR $<$ MLR $<$ GBR). Due to a lack of data, we could not fully analyze this assumption. Still, the perceived accuracy was fully consistent with this assumption.

This study opens some promising avenues for further research. Future research could investigate further factors that could play a role for the relative importance of explainability to accuracy. For example, particular ML models might be preferred because they fit existing cost estimation practices in the organization. It was mentioned that a similar approach to the CBR method was normally used when predecessors were available. The fit of the CBR method to the approach commonly applied could be one explanation why that method was chosen. Existing practices for the use of particular models, such as CBR in this case, may influence the use of models because experience with information technology has a positive impact on system usage through habit formation (Burton-Jones & Hubona, 2006). Future research could address the impact of existing methods on the selection of ML models and how this could hamper the selection of ML models that would be more suitable for the task in terms of the accuracy or explainability.

Finally, it would be interesting to analyze the relative importance of explainability to the accuracy of ML models in accounting in other areas than new product development. Also, it would be interesting to study more broadly the use of ML models in product development, because diverse kinds of management controls are used in product development and these are often interdependent (Carlsson-Wall et al., 2021; Strauss et al., 2024). Future research could address the use of ML models for conducting financial feasibility studies, budgeting development costs, analyzing cost variances, formulating targets for product development projects, or modeling how product development choices drive revenues and product profitability (Akroyd & Maguire, 2011; Davila et al., 2009; Grabner & Speckbacher, 2016; Henri & Wouters, 2020).

Acknowledgements

The authors thank participants of the 13th Conference on New Directions in Management Accounting, participants of the research seminar at the University of Amsterdam, as well as the reviewers and associate editor of the journal for their valuable comments and suggestions. They also thank the research participants for their partaking in this study and the case study organization for their financial support.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the case study organization.

Supplemental Data and Research Materials

Supplemental data for this article can be accessed on the Taylor & Francis website, https://doi.org/10.1080/09638180.2025.2463961.

ORCID

Marc Wouters http://orcid.org/0000-0001-8250-3290

References

Abernethy, M., Chin, K., Nair, S., & Soderstrom, N. (2023). Can technology-enabled advanced monitoring systems influence individual performance and team dynamics? *European Accounting Review*, *32*(3), 577–605. https://doi.org/10.1080/09638180.2022.2060845

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

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

Aguinis, H., Ramani, R. S., & Alabduljader, N. (2018). What you see is what you get? Enhancing methodological transparency in management research. *Academy of Management Annals*, *12*(1), 83–110. https://doi.org/10.5465/annals.2016.0011

Akroyd, C., & Maguire, W. (2011). The roles of management control in a product development setting. *Qualitative Research in Accounting & Management*, *8*(3), 212–237. https://doi.org/10.1108/11766091111162061

Al-Qady, M., & El-Helbawy, S. (2016). Integrating target costing and resource consumption accounting. *Journal of Applied Management Accounting Research*, *14*(1), 1–39.

Ali, A. (1994). Pioneering versus incremental innovation: Review and research propositions. *Journal of Product Innovation Management*, *11*(1), 46–61. https://doi.org/10.1111/1540-5885.1110046

Ali, A., Kalwani, M. U., & Kovenock, D. (1993). Selecting product development projects: Pioneering versus incremental innovation strategies. *Management Science*, *39*(3), 255–274. https://doi.org/10.1287/mnsc.39.3.255

Allaire, Y., & Firsirotu, M. E. (1989). Coping with strategic uncertainty. *MIT Sloan Management Review*, *30*(3), 7–16.

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. 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

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

Ansari, S., Bell, J., & Okano, H. (2006). Target costing: Uncharted research territory. In C. S. Chapman, A. G. Hopwood, & M. Shields (Eds.), *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*, 29–44. https://doi.org/10.1016/j.accinf.2017.03.003

Arnaboldi, M., de Bruijn, H., Steccolini, I., & van der Voort, H. (2022). On humans, algorithms and data. *Qualitative Research in Accounting & Management*, *19*(3), 241–254. https://doi.org/10.1108/QRAM-01-2022-0005

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

Barczak, G., & McDonough, E. F. (2003, July 20). The role of trust and project commitment in new product development teams. In D. F. Kocaoglu & T. R. Anderson (Eds.), *PICMET '03: Portland International Conference on Management of Engineering and Technology Management for Reshaping the World, 2003* (pp. 274–281). Portland State Univ. https://doi.org/10.1109/PICMET.2003.1222804

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

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

Bertomeu, J. (2020). Machine learning improves accounting: Discussion, implementation and research opportunities. *Review of Accounting Studies*, *25*(3), 1135–1155. https://doi.org/10.1007/s11142-020-09554-9

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

Bodendorf, F., Merkl, P., & Franke, J. (2021). Intelligent cost estimation by machine learning in supply management: A structured literature review. *Computers & Industrial Engineering*, *160*, 107601. https://doi.org/10.1016/j.cie.2021.107601

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/10.1007/978-3-031-01869-5

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

Bonsón, E., Lavorato, D., Lamboglia, R., & Mancini, D. (2021). Artificial intelligence activities and ethical approaches in leading listed companies in the European Union. *International Journal of Accounting Information Systems*, *43*, 100535. https://doi.org/10.1016/j.accinf.2021.100535

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

Bracci, E. (2023). The loopholes of algorithmic public services: An "intelligent" accountability research agenda. *Accounting, Auditing & Accountability Journal*, *36*(2), 739–763. https://doi.org/10.1108/AAAJ-06-2022-5856

Brown, S. L., & Eisenhardt, K. M. (1995). Product development: Past research, present findings, and future directions. *The Academy of Management Review*, *20*(2), 343. https://doi.org/10.2307/258850

Brüggen, A., Grabner, I., & Sedatole, K. L. (2021). The folly of forecasting: The effects of a disaggregated demand forecasting system on forecast error, forecast positive bias, and inventory levels. *The Accounting Review*, *96*(2), 127–152. https://doi.org/10.2308/tar-2018-0559

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

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

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

Carlsson-Wall, M., Goretzki, L., Kraus, K., & Lind, J. (2021). Exploring the role of management control anchor practices in new product development. *European Accounting Review*, *30*(2), 251–276. https://doi.org/10.1080/09638180.2020.1763187

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

Chapman, C. S. (1997). Reflections on a contingent view of accounting. *Accounting, Organizations and Society*, *22*(2), 189–205. https://doi.org/10.1016/S0361-3682(97)00001-9

Chen, C. X., Hudgins, R., & Wright, W. F. (2022). The effect of advice valence on the perceived credibility of data analytics. *Journal of Management Accounting Research*, *34*(2), 97–116. https://doi.org/10.2308/JMAR-2020-015

Chenhall, R. H., & Moers, F. (2015). The role of innovation in the evolution of management accounting and its integration into management control. *Accounting, Organizations and Society*, *47*, 1–13. https://doi.org/10.1016/j.aos.2015.10.002

Cho, S., Vasarhelyi, M. A., Sun, T., & Zhang, C. (2020). Learning from machine learning in accounting and assurance. *Journal of Emerging Technologies in Accounting*, *17*(1), 1–10. https://doi.org/10.2308/jeta-10718

Chong, V. K., & Johnson, D. M. (2007). Testing a model of the antecedents and consequences of budgetary participation on job performance. *Accounting and Business Research*, *37*(1), 3–19. https://doi.org/10.1080/00014788.2007.9730055

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

Chye Koh, H., & Kee Low, C. (2004). Going concern prediction using data mining techniques. *Managerial Auditing Journal*, *19*(3), 462–476. https://doi.org/10.1108/02686900410524436

Commerford, B. P., Dennis, S. A., Joe, J. R., & Ulla, J. W. (2022). Man versus machine: Complex estimates and auditor reliance on artificial intelligence. *Journal of Accounting Research*, *60*(1), 171–201. https://doi.org/10.1111/1475-679X.12407

Cooper, R., & Slagmulder, R. (2004). 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., 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

Davila, A. (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, 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

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

Dikolli, S. S., & Sedatole, K. L. (2004). Delta's new song: A case on cost estimation in the airline industry. *Issues in Accounting Education*, *19*(3), 345–358. https://doi.org/10.2308/iace.2004.19.3.345

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.

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

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), 1–13. 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., 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

Fehrenbacher, D. D., Ghio, A., & Weisner, M. (2023). Advice utilization from predictive analytics tools: The trend is your friend. *European Accounting Review*, *32*(3), 637–662. https://doi.org/10.1080/09638180.2022.2138934

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

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

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

Galbraith, J. R. (1973). *Designing complex organizations. Organization development*. Addison-Wesley.

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

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

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

Grabner, I., Posch, A., & Wabnegg, M. (2018). Materializing innovation capability: A management control perspective. *Journal of Management Accounting Research*, *30*(2), 163–185. https://doi.org/10.2308/jmar-52062

Grabner, I., & Speckbacher, G. (2016). The cost of creativity: A control perspective. *Accounting, Organizations and Society*, *48*, 31–42. https://doi.org/10.1016/j.aos.2015.11.001

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

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 program. *AI Magazine*, *40*(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

Guo, B., Paraskevopoulou, E., & Santamaría Sánchez, L. (2019). Disentangling the role of management control systems for product and process innovation in different contexts. *European Accounting Review*, *28*(4), 681–712. https://doi.org/10.1080/09638180.2018.1528168

Hall, P., & Gill, N. (2018). *Introduction to machine learning interpretability* (2nd ed.). O'Reilly.

Hammann, D. (2024). Big data and machine learning in cost estimation: An automotive case study. *International Journal of Production Economics*, *269*, 109137. https://doi.org/10.1016/j.ijpe.2023.109137

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

Herschung, F., Mahlendorf, M. D., & Weber, J. (2002). Mapping quantitative management accounting research. *Journal of Management Accounting Research*, *30*(1), 73–141. https://doi.org/10.2308/jmar-51745

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), 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), 79. https://doi.org/10.1504/JDR.2007.015564

H'mida, F., Martin, P., & Vernadat, F. (2006). Cost estimation in mechanical production: The cost entity approach applied to integrated product engineering. *International Journal of Production Economics*, *103*(1), 17–35. https://doi.org/10.1016/j.ijpe.2005.02.016

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

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

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

Janka, M., & Guenther, T. W. (2018). Management control of new product development and perceived environmental uncertainty: Exploring heterogeneity using a finite mixture approach. *Journal of Management Accounting Research*, *30*(2), 131–161. https://doi.org/10.2308/jmar-52019

Jørgensen, B., & Messner, M. (2010). Accounting and strategising: A case study from new product development. *Accounting, Organizations and Society*, *35*(2), 184–204. https://doi.org/10.1016/j.aos.2009.04.001

Kadir, A. Z. A., Yusof, Y., & Wahab, M. S. (2020). Additive manufacturing cost estimation models—A classification review. *The International Journal of Advanced Manufacturing Technology*, *107*(9-10), 4033–4053. https://doi.org/10.1007/s00170-020-05262-5

Khurana, A. (1998). Towards holistic "front ends" in new product development. *Journal of Product Innovation Management*, *15*(1), 57–74. https://doi.org/10.1016/S0737-6782(97)00066-0

Kietzmann, J., & Pitt, L. F. (2020). Artificial intelligence and machine learning: What managers need to know. *Business Horizons*, *63*(2), 131–133. https://doi.org/10.1016/j.bushor.2019.11.005

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

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

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

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), 21–31.

Lehner, O. M., Ittonen, K., Silvola, H., Ström, E., & Wührleitner, A. (2022). Artificial intelligence based decision-making in accounting and auditing: Ethical challenges and normative thinking. *Accounting, Auditing & Accountability Journal*, *35*(9), 109–135. https://doi.org/10.1108/AAAJ-09-2020-4934

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

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

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

Luft, J., & Shields, M. D.. (2003). Mapping management accounting: graphics and guidelines for theory-consistent empirical research. Accounting. *Organizations and Society*, *28*(2-3), 169–249. https://doi.org/10.1016/S0361-3682(02)00026-0

Mahlendorf, M. D., Martin, M. A., & Smith, D. (2023). Innovative data – use-cases in management accounting research and practice. *European Accounting Review*, *32*(3), 547–576. https://doi.org/10.1080/09638180.2023.2213258

Margolin, B. H. (1969). Resolution Iv fractional factorial designs. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *31*(3), 514–523. https://doi.org/10.1111/j.2517-6161.1969.tb00811.x

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

Moenaert, R. K., de Meyer, A., 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. (2021). A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Transactions on Interactive Intelligent Systems*, *11*(3-4), 1–45. https://doi.org/10.1145/3387166

Moll, J. (2015). Editorial: Special issue on innovation and product development. *Management Accounting Research*, *28*, 2–11. https://doi.org/10.1016/j.mar.2015.05.003

Newnes, L. B., Mileham, A. R., Cheung, W. M., Marsh, R., Lanham, J. D., Saravi, M. E., & Bradbery, R. W.. (2008). Predicting the whole-life cost of a product at the conceptual design stage. *Journal of Engineering Design*, *19*(2), 99–112. https://doi.org/10.1080/09544820701803061

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

Nielsen, S. (2022). Management accounting and the concepts of exploratory data analysis and unsupervised machine learning: A literature study and future directions. *Journal of Accounting & Organizational Change*, *18*(5), 811–853. https://doi.org/10.1108/JAOC-08-2020-0107

Nixon, B. (1998). Research and development performance measurement: A case study. *Management Accounting Research*, *9*(3), 329–355. https://doi.org/10.1006/mare.1998.0079

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

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

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

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

Qian, L., & 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

Quattrone, P. (2016). Management accounting goes digital: Will the move make it wiser? *Management Accounting Research*, *31*, 118–122. https://doi.org/10.1016/j.mar.2016.01.003

Ranta, M., Ylinen, M., & Järvenpää, M. (2023). Machine learning in management accounting research: Literature review and pathways for the future. *European Accounting Review*, *32*(3), 607–636. https://doi.org/10.1080/09638180.2022.2137221

Reid, S. E., & de Brentani, U. (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

Relich, M., & Pawlewski, P. (2018). A case-based reasoning approach to cost estimation of new product development. *Neurocomputing*, *272*, 40–45. https://doi.org/10.1016/j.neucom.2017.05.092

Repenning, N., Löhlein, L., & Schäffer, U. (2022). Emotions in accounting: A review to bridge the paradigmatic divide. *European Accounting Review*, *31*(1), 241–267. https://doi.org/10.1080/09638180.2021.1908906

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

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

Román, F. J. (2011). A case study on cost estimation and profitability analysis at continental airlines. *Issues in Accounting Education*, *26*(1), 181–200. https://doi.org/10.2308/iace.2011.26.1.181

Roy, R. (2003). *Cost engineering: Why, what and how? Decision engineering report series*. Cranfield University.

Rush, C., & Roy, R. (2000). *Analysis of cost estimating processes used within a concurrent engineering environment throughout a product life cycle*. 7th ISPE International Conference on Concurrent Engineering: Research and Applications, 19-21 July, Lyon, France.

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

Shin, Y. (2015). Application of boosting regression trees to preliminary cost estimation in building construction projects. *Computational Intelligence and Neuroscience*, *2015*, 1–9. https://doi.org/10.1155/2015/149702

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

Strauss, E., Malz, S., & Weber, J. (2024). The role of management controls in new product development: Codifying a collective source of creativity. *European Accounting Review*, *33*(1), 251–277. https://doi.org/10.1080/09638180.2022.2082994

Sutton, S. G., Holt, M., & Arnold, V. (2016). "The reports of my death are greatly exaggerated"—Artificial intelligence research in accounting. *International Journal of Accounting Information Systems*, *22*, 60–73. https://doi.org/10.1016/j.accinf.2016.07.005

Taipaleenmäki, J. (2014). Absence and variant modes of presence of management accounting in new product development – Theoretical refinement and some empirical evidence. *European Accounting Review*, *23*(2), 291–334. https://doi.org/10.1080/09638180.2013.811065

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

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. *The Academy of Management Review*, *3*(3), 613–624. https://doi.org/10.2307/257550

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 Meer-Kooistra, J., & Scapens, R. W. (2015). Governing product co-development projects: The role of minimal structures. *Management Accounting Research*, *28*, 68–91. https://doi.org/10.1016/j.mar.2015.05.001

van Triest, S., Kloosterman, H., & Groen, B. A. (2023). Under which circumstances are enabling control and control extensiveness related to employee performance? *Management Accounting Research*, *59*, 100831. https://doi.org/10.1016/j.mar.2023.100831

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

Voltolini, R., Vasconcelos, K., Borsato, M., & Peruzzini, M. (2019). Product development cost estimation through ontological models – A literature review. *Journal of Management Analytics*, *6*(2), 209–229. https://doi.org/10.1080/23270012.2019.1598899

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

Williams, J. J., & Seaman, A. E. (2002). Management accounting systems change and departmental performance: The influence of managerial information and task uncertainty. *Management Accounting Research*, *13*(4), 419–445. https://doi.org/10.1006/mare.2002.0199

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* (Vol. 26, pp. 139–274). Emerald. https://doi.org/10.1108/S1474-787120150000026005

Ylinen, M., & Gullkvist, B. (2012). The effects of tolerance for ambiguity and task uncertainty on the balanced and combined use of project controls. *European Accounting Review*, *21*(2), 395–415. https://doi.org/10.1080/09638180.2011.631733

Ylinen, M., & Ranta, M. (2024). Employer ratings in social media and firm performance: Evidence from an explainable machine learning approach. *Accounting & Finance*, *64*(1), 247–276. https://doi.org/10.1111/acfi.13146

Zhang, C. (2019). Intelligent process automation in audit. *Journal of Emerging Technologies in Accounting*, *16*(2), 69–88. https://doi.org/10.2308/jeta-52653

Zhang, C., Cho, S., & Vasarhelyi, M. (2022). Explainable artificial intelligence (XAI) in auditing. *International Journal of Accounting Information Systems*, *46*, 100572. https://doi.org/10.1016/j.accinf.2022.100572

Zhang, X., Wang, Y., & Li, Z. (2023). RETRACTED: User acceptance of machine learning models – Integrating several important external variables with technology acceptance model. *International Journal of Electrical Engineering & Education*, *60*(1_suppl), 3986–4005. https://doi.org/10.1177/00207209211005271

Appendix. Questionnaire and steps in the 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) of explainability

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) of explainability

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 the measured accuracy data for each machine learning model, based on the archival data of AutomotiveCompany

(7) Model rating (CBR, MLR, GBR) of 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) Explanation of the 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) Evaluation (16 treatment conditions)

a. Relative importance of explainability to accuracy in this scenario.*Accuracy* ☐ – ☐ – ☐ – ☐ – ☐ – ☐ – ☐ – ☐ – ☐ – ☐ *Explainability*
b. In this situation, 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?