LAPPEENRANTA-LAHTI UNIVERSITY OF TECHNOLOGY LUT
LUT School of Energy Systems
LUT Mechanical Engineering

Niko Hyytiäinen

APPLICABILITY OF MLPP ANALYSIS BASED ON TECHNICAL CHARACTERISTICS OF HYDRAULIC CYLINDERS

Jyväskylä 13.6.2021

Examiners: Docent Harri Eskelinen
Olli Harkoma

TIIVISTELMÄ

Lappeenrannan-Lahden teknillinen yliopisto LUT
LUT School of Energy Systems
LUT Mechanical Engineering

Niko Hyytiäinen

MLPP ANALYYSIN SOVELTUVUUS HYDRAULIIKKASYLINTEREILLE PERUSTUEN NIIDEN TEKNISIIN OMINAISUUKSIIN

Diplomityö

2021

111 sivua, 58 kuvaa, 25 taulukkoa ja 16 liitettä

Tarkastajat: Dosentti Harri Eskelinen
Olli Harkoma

Hakusanat: performance pricing, multilineaarinen regressioanalyysi, should costing, hydrauliikkasylinterit

Työssä tutkitaan multilineaarisen regressioanalyysimenetelmän (MLPP) soveltuvuutta hydrauliikkasylintereiden hinnoittelun analysointiin. Tutkimus toteutettiin vertailemalla useisiin eri hydrauliikkasylintereiden teknisiin ominaisuuksiin perustuvia malleja. MLPP:n avulla suurien osanumeromäärien ja toimittajien vertailu tehostuisi. MLPP:tä voitaisiin hyödyntää myös tavoitekustannusten määrittämisessä uusien tuoteprojekteiden aikana.

Analyysissa käytettävien kustannustekijöiden ja arvoajureiden valinnan perustana hyödynnettiin kirjallisuustutkimusta, yksityiskohtaisia kustannuslaskelmia, olemassa olevia kustannusrakenteita ja haastatteluita. Kustannustekijöiden keskinäisiä riippuvaisuuksia ja niiden suhdetta hinnoitteluun tutkittiin korrelaatiomatriisien avulla, ja tilastolliset validoinnit tehtiin sekä regressiofunktioille että yksittäisille regressiokertoimille. Mallien validiteettia arvioitaessa hyödynnettiin myös yksityiskohtaisia alhaalta ylöspäin- periaatteella tehtyjä kustannusanalyyseja. Osana mallien arviointia, eri mallivaihtojen kuvaajien havaintopisteiden sijainteja ja poikkeamia analysoitiin.

Työn tuloksena valittiin kaksi parasta mallia perustuen tilastollisiin mittareihin ja mallin käytettävyyteen. Malleista toinen soveltuu erityisesti sarjatuotteiden- ja toinen uusien tuotteiden analysointiin. Oston näkökulmasta, tutkimus nosti esiin useita potentiaalisia tuotteita jatkotutkimuksia varten, ja tutkimuksessa luotujen kuvaajien avulla on mahdollista tutkia toimittajien kilpailukyvykkyyttä. Tämä työ toimii MLPP:n ohjedokumenttina ja antaa vakaan pohjan MLPP:n toteuttamisstrategialle yrityksessä. Johtopäätöksenä, MLPP soveltuu hydrauliikkasylintereiden hinnoittelun analysointiin etenkin silloin, kun sitä käytetään yhdessä yksityiskohtaisempien kustannusanalysointimenetelmien kanssa. MLPP:n avulla voidaan kartoittaa säästöpotentiaalia suuremmasta joukosta tuotteita ja muodostaa tuoteryhmän kokonaiskuva perustuen hintoihin ja tuotteiden tekniseen vaativuuteen.

ABSTRACT

Lappeenranta-Lahti University of Technology LUT
LUT School of Energy Systems
LUT Mechanical Engineering

Niko Hyytiäinen

APPLICABILITY OF MLPP ANALYSIS BASED ON TECHNICAL CHARACTERISTICS OF HYDRAULIC CYLINDERS

Master's Thesis

2021

111 pages, 58 figures, 25 tables and 16 appendices

Examiners: Docent Harri Eskelinen
Olli Harkoma

Keywords: performance pricing, multilinear regression, should costing, hydraulic cylinders

The objective of this master's thesis was to study the applicability of Multidimensional Linear Performance Pricing (MLPP) for hydraulic cylinders. The research was carried out by comparing various combinations of value drivers based on technical characteristics of hydraulic cylinders. By means of MLPP, the analysis of a high number of part numbers and comparison of suppliers would be enhanced. MLPP could also be used as a supportive analysis method for target costing during new product introduction (NPI).

The basis for the value driver selection was established by utilizing literature research, bottom-up calculations, existing cost structures and the information received during the interviews. The interdependencies among potential value drivers as well as the relationships between value drivers and prices were studied through correlation matrices. Statistical validations were performed for regression functions as well as for the individual regression coefficients. The detailed bottom-up calculations were created for checking the validity of the models. As part of the model evaluation, the scatter plot analysis of different model options was done by analyzing the residuals and locations of individual data points.

As a result, the two best models were selected based on the statistical indicators and usability of the models. One of the models is suitable for the analysis of serial products and the other for new products. From a purchasing perspective, this research helped to identify multiple potential products for further investigation and the created scatter plots facilitate to study the competitiveness of the suppliers. This research works as instruction document for MLPP and forms a solid base for further implementation strategy of MLPP inside the corporation. In conclusion, MLPP is an applicable price analysis method for hydraulic cylinders, especially when used in parallel with more detailed methods like bottom-up analysis. MLPP enables the identification of saving potential among a higher number of products and facilitates to form an overall picture of a product group based on prices and technical value of products.

ACKNOWLEDGEMENTS

I want to thank Valtra and AGCO Corporation for this opportunity and all the support during my studies. The last couple of years have been very busy, but at the same time I have gained a lot of professional experience and new perspectives to my daily work. Combining work, studies and house building project definitely showed me the meaning of planning and time management. I also want to thank Harri Eskelinen from LUT for all the support during this master's thesis. Special thanks go to my supervisor and our team director Olli Harkoma for the valuable support and encouragement during this thesis and my studies. Last but not least, I want to thank Karoliina, my family and friends for enduring and supporting me during my studies.

Cheers

Niko Hyytiäinen

Jyväskylä, 13.6.2021

TABLE OF CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

Symbols

A_p	Piston area [m^2]
A_r	Rod-side area [m^2]
Adjusted R^2	Modification of coefficient of determination
b	Regression coefficient
[B]	Regression coefficient matrix
d	Number of variables
[D]	Error or unexplained variation in the dependent variable
F	Force [N]
F_critical	F value indicator for statistical deviations
F (statistical)	Statistical predictability
L	Length [m]
n	Number of data points [pcs]
P	Pressure [Pa]
P_c	Price of the product (current) [€]
P_s	Price of the product (standardized) [€]
Q	Flow rate [m^3/s]
Q^2	Stone Geisser Criterion
R^2	Coefficient of determination
RM_{pc}	Raw material price (current) [€/kg or €/ton]
RM_{pry}	Raw material price (reference year) [€/kg or €/ton]
RM_s	Raw material share [%]
signif F	Significance level for F
v	Piston speed [m/s]
V	Volume [m^3]
W	Weight [kg]
x	Independent variable (value driver)
[X]	Matrix for data points
y	Dependent variable (the outcome of the regression)
[Y]	Price vector of each data point
$[y_0]$	Common Y-axis intersection

Abbreviations

APT	AGCO Purchasing Tool
BCC	Best-Cost-Country
CSA	Cost Structure Analysis
GCM	Global Commodity Manager
EME	Europe & Middle East
ERP	Enterprise Resource Planning
LPP	Single Linear Performance Pricing
LTA	Long Term Agreement
MB51	Material Document List transaction (SAP)
MLPP	Multidimensional Linear Performance Pricing
MLR	Multilinear Regression
NLPP	Non-Linear Performance Pricing
OEM	Original Equipment Manufacturer
PCM	Product Cost Management (Siemens software)
PLM	Product Lifecycle Management
PLS	Partial Least Squares
PP	Performance Pricing
R&D	Research & Development
RSQ	Square of the Pearson product moment correlation coefficient
RSS	Residual Sum of Squares
SAP	ERP Software
SCA	Supplier Cost Analysis
SR	Skived and Roller Burnished delivery condition
TCA	True Cost Analysis
TSS	Total Sum of Squares

1 INTRODUCTION

Statistical methods like regression-based top-down value analysis are nowadays relatively widely used among companies in purchasing and during early phases of new projects. The main benefit of top-down analysis is that it allows to analyze a higher number of products or material groups in a relatively short time. Instead of detailed cost elements, which are typically analyzed with a bottom-up analysis method, in top-down analysis the aim is to find the most important value-related criteria (e.g. characteristics or performance of the product), called as value drivers, to explain the pricing of the product. (VDI 2018, p. 3). Top-down models facilitate to understand the influence of different cost parameters on pricing and may be used for instance with products, for which it is challenging or time demanding to build a detailed bottom-up models. However, because of the statistical nature of the analysis, certain preconditions needs to be fulfilled to guarantee reliable basis for the analysis. (Schuh 2017, p. 174). At the same time, due to utilization of several variables in the same analysis for prediction of the price, a sufficient technical knowledge of product properties as well as the understanding of principles used in statistical analysis are required in order to validate and evaluate the results of the analysis.

1.1 Background and motivation

This thesis was done in global purchasing organization of agricultural machinery and equipment manufacturer AGCO, whose main brands are Challenger, Fendt, Massey Ferguson, Valtra and GSI. (AGCO 2021) Hydraulic cylinders are widely used among different brands, vehicles and equipment. MLPP (Multidimensional Linear Performance Pricing) has been used inside AGCO Supplier Cost Analysis (SCA) team for several years, especially for exploring the pricing of products with high quantity of part numbers and with similar technical characteristics. Typical approach is to identify possible saving potential through MLPP analysis and perform more detailed analysis for selected data points by using more detailed methods like bottom-up calculation. The suitability of MLPP has not been tested before for hydraulic cylinders. From AGCO purchasing point of view, hydraulic cylinders are categorized under the main commodity “Linkages, Cylinders, Gears and Machined Components”, and the purchasing spend of hydraulic cylinders forms 20 % of the total spend of the commodity. In terms quantity of parts, there are roughly 400 different

cylinder part numbers with annual purchasing spend of >5 k$. Before this study, the coverage of should cost analysis (including bottom-up and MLPP analysis) done by SCA was 23 % in relation to total spend of cylinders. The global supplier base consists of 21 different hydraulic cylinder suppliers, which are delivering cylinders to 19 different AGCO production locations.

1.2 Research problem

The research problem of this thesis is to study the applicability of MLPP analysis for different types of hydraulic cylinders used in tractors. For creation of the MLPP model, it is necessary to clarify the main factors influencing on costs, which are typically related to technical properties of hydraulic cylinders. Thus, identification of cost drivers is also one key area of the research problem. Cost drivers of hydraulic cylinders will also work as a framework for this work.

1.3 Objective and research questions

The main objective is to create a MLPP model for hydraulic cylinders by comparing different model options, which consist of different combinations of value drivers. By means of MLPP, it would be possible to increase the analyzed purchasing spend and identify potential part numbers for further studies and negotiations with the suppliers. At the same time, in case MLPP turns out to be an applicable tool for hydraulic cylinders, it would probably improve the performance of purchasing, because the identification of unfavorable part numbers from the larger mass would be more efficient. Since the MLPP model scatter plot consists of multiple data points, it is possible to get a better view of the competitiveness of suppliers. MLPP could also be used as a supportive tool for fast analysis in NPI (New Product Introduction) projects for target cost determination. Additionally, it may be utilized as a new tool during negotiations with cylinder suppliers, since the new data repository feature allows also other stakeholders like Global Commodity Managers (GCM) and buyers to utilize existing MLPP models created by SCA in read-only mode for example, during quotation evaluation, price negotiation or new business awards. In general, a thorough study about MLPP process will surely bring benefits for other commodities and MLPP users as well, and

this thesis should work as a comprehensive and detailed instruction document for MLPP model implementation.

In order to solve the research problem and achieve the objective of this study, research questions were created. The main research question for this study is:

- What is the best combination of value drivers and does the technical value based on these value drivers make sense?

Other supportive research questions are listed below:

- What are the cost drivers of hydraulic cylinders?
- What causes possible differences in pricing between different cylinder types and are those in line with technical value?
- What possibly causes differences between suppliers in pricing for cylinders which have similar technical value?
- How statistically accurate model can be built by using cylinder weight as only value driver (this only measures how "well" pricing is following €/kg principle like in traditional regression analysis)?
- Is it possible to build reliable model without using weight of cylinder?

1.4 Research methods

Due to the statistical nature of this study, selected research methods consist mostly of quantitative methods. The main analysis method is the MLPP analysis, which enables the comparison between performance or technical characteristics of the product (combination of value drivers) and prices instead of only one criterion (e.g. only weight in one-dimensional regression analysis) and price (VDI 2018, p. 5). Supportive methods include several different type of analysis like data processing during data collection, price standardization, correlation matrices, analysis of statistical indicators, distribution- and scatter plot analysis and bottom-up calculations. Qualitative methods are used in the form of interviews as a basis for model creation and evaluation of the results. Furthermore, in order to expand the theoretical background of the MLPP process, two benchmarking interviews are performed with Hilti and Siemens. The process chart introduced in figure 1 describes the overall approach of methods used in the model development.

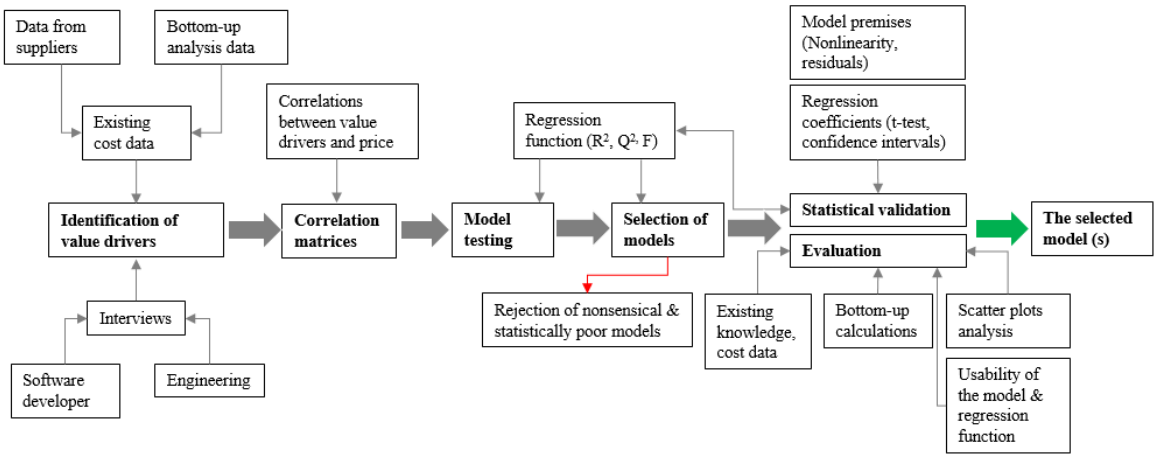


Figure 1. Process chart for methods used during MLPP model development.

1.5 Scope

The scope of this study is reduced to hydraulic cylinders used in tractors on Europe & Middle East (EME) region, which is one of the global purchasing regions alongside Asian/Pacific American (APA), North America (NA) and South America (SA). Only one region was selected to eliminate the fluctuations caused by currencies and economics. However, most (72 %) of the global purchasing spend of hydraulic cylinders is coming from EME region. From technical perspective, with selected cylinders the range of inner diameter is 40-110 mm, whereas the collapsed length is between 220 and 1650 mm. The weight varies between 3 and 46 kg. In terms of purchasing price, the range is from 60 € to 240 €. The cylinders are used mainly for lifting, but the scope includes also cylinders used for suspension of the vehicle. Based on the application, cylinders were categorized into four different types. The collected data includes technical specifications for 103 hydraulic cylinders, of which 80 part numbers were selected to be utilized for MLPP model development after detailed data processing. The supplier base of selected hydraulic cylinders consists of 6 European suppliers. From purchasing spend perspective, this study covers 55 % of the total global spend of hydraulic cylinders like presented in figure 2.

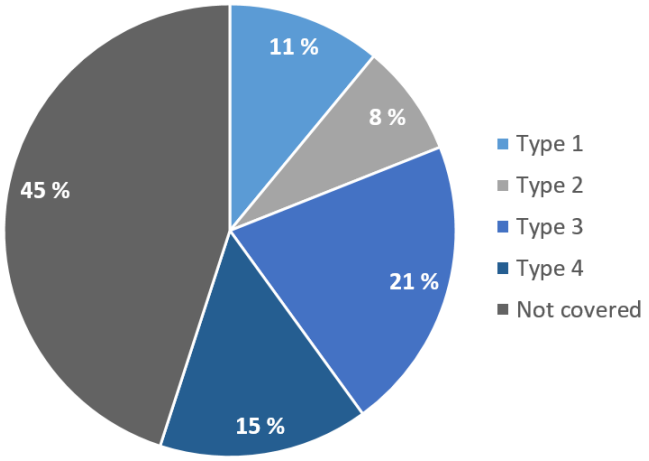


Figure 2. Spend coverage analysis including the cylinder types in scope of the study.

1.6 Contribution

Performance pricing (PP) has been a widely used analysis method for quite a long time, especially within the automotive industry. According to Sherefkin 2003, all car manufacturers use some variation of PP analysis method. However, depsite the widespread usage, there is a limited amount of discussion about its practical applicability (Coşkun & Yilmaz A 2021). Newman & Krehbiel 2007 also highlights the lack of discussion about PP analysis methods in literature. Based on VDI 2018, PP method is generally used, but the procedures are typically company-specific. The same conclusion was made by Schmidt 2019, who compares the cost management methods for purchased parts, and mentions the limited attention in literature to be in contrast to common usage in industry. Similarly, the amount of studies examining especially the applicability of MLPP analysis for pricing of hydraulic cylinders seems to be relatively limited. For this reason, this research is relevant to both scientific community and client corporation. At the same time, it discusses the cost drivers of hydraulic cylinders as well as provides a detailed and practical example of MLPP model development, validation and evaluation process.

2 MULTIDIMENSIONAL LINEAR PERFORMANCE PRICING (MLPP)

MLPP (Multidimensional- or multilinear performance pricing) is an analysis of the correlation between the purchasing price and selected parameters, like characteristics or performance of the product, called as value drivers. (VDI 2018, p. 5-6) A traditional linear regression analysis where relationships between a dependent and independent variable are analyzed, is flexible and one of the most used statistical analysis methods (Backhaus et al. 2018, p. 58). Perhaps the best known one-dimensional performance pricing method in industrial practices is the unit-kilogram price method, in which the weight of the product is the only value driver (Möller 2009, p. 26). However, for most research purposes, it is necessary to include more than one independent variable in the model (Backhaus et al. 2018, p. 72). Especially with complex products, it may be challenging to explain the pricing of the product based only on one value driver like in one-dimensional regression analysis. With MLPP, multiple value drivers may be utilized in the same regression analysis, which improves the quality of the analysis while it reflects the correlation between multiple value drivers. (Processbench 2017) This chapter discusses MLPP partially based on guidelines provided by Processbench, since it is the developer of the statistical MLPP software utilized in this study. In addition to MLPP, there are several other available variants of PP analysis like Single Linear Performance Pricing (LPP) and Non-Linear Performance Pricing (NLPP) (Schmidt 2019). One commercial software example is Saphirion, which is based on NLPP (Non-Linear Performance Pricing) method (Saphirion). Despite the PP method, the main idea of PP analysis is to benchmark the price of the product by taking into account only the selected performance or technical characteristic-based value drivers and compare these parameters to price (Münch 2015).

2.1 MLPP as a statistical method

MLPP is basically a multilinear regression (MLR) analysis, which based on literature research is a widely used method especially in the field of medicine, epidemiology, biostatistics, pharmaceutical research and chemometrics (Valveny & Gilliver 2016; Wold et al. 2001; Harrel 2001, p. 3). In one-dimensional regression analysis, one value driver is used to explain the price of a product, which may be efficient for very simple group of products. It basically reflects only one parameter (e.g. weight of the product) and the relationship

between selected parameter and purchasing price. The representation and formula of one-dimensional regression analysis is introduced on the left side of figure 3. (Processbench 2017)

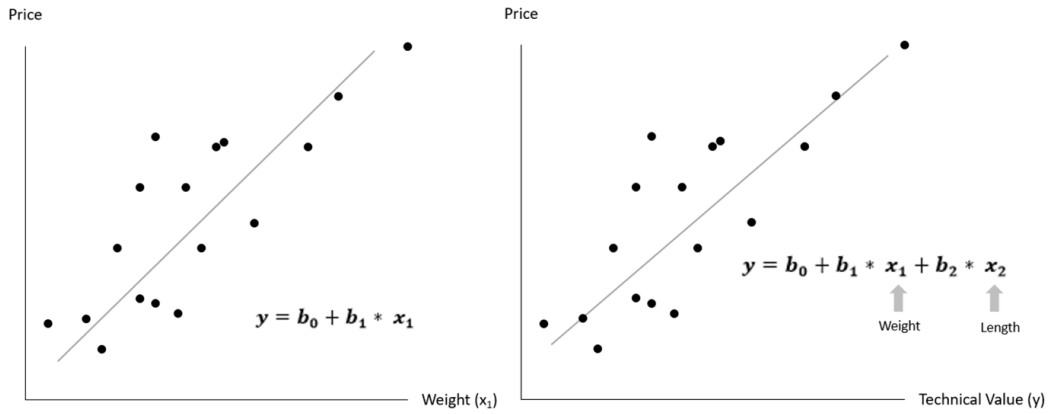


Figure 3. One- and multidimensional regression (modified) (Backhaus et al. 2008, p. 64 & 72; Processbench 2017).

However, when the complexity of product increases, the pricing of the product cannot be easily explained based on only one value driver. Through multiple linear regression analysis, several parameters may be included in the same analysis. The difference between one and multidimensional regression analysis basic formula can be seen by comparing formulas introduced in figure 3. Based on the regression (average) line formula, length (x_2) is additionally chosen as a second value driver of analysis. The technical value (y) is statistically calculated based on selected value drivers. (Processbench 2017) In terms of mathematics, the regression coefficients b_0, b_1, b_2 etc. are unknown parameters whereas x_1, x_2 etc. are independent, predicting variables. The variable y is called target, response or dependent variable and hence usually the most interesting variable in terms of understanding or modelling. (Chatterjee & Simonoff 2013, p. 3-4)

MLPP mathematically combines several value drivers into a technical value, to explain the one dependent variable, which is typically purchasing price (Processbench 2017). Compared to bottom-up calculation, MLPP works other way around by approaching the pricing from top-down analysis perspective. This enables relatively fast investigation for many components with significantly less effort. (Soltau 2018) As a mathematical analysis tool, MLPP works as a good support by offering a base for further and more detailed methods like

cost structure analysis. The usual procedure is to place the technical value of the product on the x-axis and price on the y-axis. The regression (average) line will then be placed based on locations of each data points in the scatter plot. Depending on application, y-axis can sometimes also be used for other parameters like product features instead of price. (Soltau 2018; VDI 2018, p. 43-44)

The terminology in this area is a bit inconsistent, since especially the terms multivariate-, multivariable-, multiple linear- and multilinear regression are often used interchangeably. In terms of statistical analysis, multiple linear, multivariable and multilinear model mean a statistical model, which has multiple independent (a.k.a. predictor) variables, but only one dependent variable (a.k.a. response or outcome). In case of performance pricing the dependent variable is usually the purchasing price. Thus, multiple linear modelling is based on equation, in which multiple variables can be found from the right side of the equation. (Hidalgo & Goodman 2013) The multivariate model differs from the multivariable by having more than one dependent variable on the left side of the equation (Ryan 1997, p. 118). The multivariate model assesses the relationship between multiple independent variables with two or more dependent variables (outcomes) at the same time (Valveny & Gilliver 2016).

In terms of visualization, it is effortless to create a 2D visualization for the linear model with one independent variable. However, in case of multilinear regression, visualization of the model gets more challenging, especially if the number of independent variables is higher than 2, due to the fact that we can only use a 3D environment for the visualization. (Shin 2018) The figure 4 represents an example of visualization for multiple linear regression with two independent variables. The plane describes the fitted least squares relationship. (Chatterjee & Simonoff 2013, p. 7) The method of least squares is a standard mathematical approach, in which the idea is to select the unknown parameter b in a way that the sum of the squares of the residuals (difference between observed and fitted value) is minimized, which improves the data- and curve fitting. (Mellin 2006, p. 260) The lengths of the vertical lines correspond to the residuals (solid line=positive residual, dashed line=negative residual). The plane is positioned to minimize the sum of squares of the residuals. (Chatterjee & Simonoff 2013, p. 7)

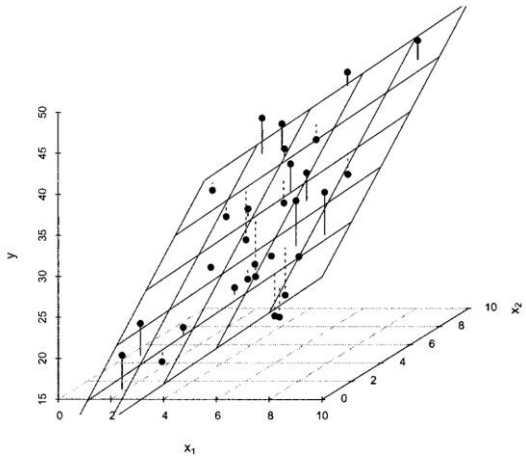


Figure 4. Example for visualization of multiple linear regression analysis with two variables (x_1 and x_2) (Chatterjee & Simonoff 2013).

2.2 Partial least squares regression

MLPP uses partial least squares (PLS) regression method (Processbench 2021: Wold et al. 2001), originally developed by Herman Wold in 1975 and discussed in the article "Soft Modelling: The Basic Design and Some Extensions" (Wold 1982). The basic principle of PLS regression is that it creates a linear regression by projecting *x*-variables (latent variables to a new space and uses these "new" latent variables to predict *Y*. The main benefit of PLS regression is that it uses two-block predictive PLS to model relationships between two matrices, *X* and *Y*, which together with modelling the stucture of *X* and *Y* enables more advanced results, compared to the traditional multiple regression approach. Another benefit of PLS regression is its ability to analyze strongly correlated (collinear) and noisy data as well as high number of *x*-variables, which is not possible with traditional multiple linear regression. PLS regression allows to investigate more complex data and typically the precision of model parameters improves, when the number of relevant observations and variables increases. (Wold 2001) In Processbench user manual, PLS method is simplified and explained with the help of an example for Rubik's cube (see figure 5). The cube represents one data point in the model. The actual analysis is performed in the shadows of the cube, which represents the projection of the model in the 2D vector space. After analysis, the results can be transformed back to 3D world. (Processbench 2021)

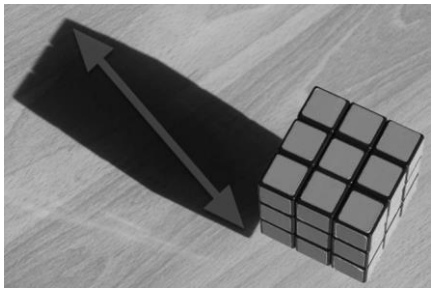


Figure 5. Example of PLS method used in MLPP (Processbench 2021).

2.3 MLPP as a method in purchasing

MLPP is nowdays a widely used method among leading companies especially in purchasing and purchasing-related development (Soltau 2018). The method was introduced to companies with purchasing activities in the 1990's by consultant companies (Güthenke & Möller 2007), and the original purpose was to achieve short-term cost savings in purchasing (Proch et al. 2017). Especially in the early phases, MLPP tools were developed for OEMs (Original equipment manufacturers) of automotive industry in order to enhance the cost efficiency. In general, MLPP method can be utilized for both new products and also for already sourced serial products (Newman & Krehbiel 2007).

Based on Schuh et al. 2017, the "Purchasing Chessboard" describes the position of MLPP analysis as one of the purchasing methods in figure 6. Due to diversity of terminology around this topic, also the positions of other levers related to target pricing like bottom-up calculation and traditional one-dimensional regression analysis are described and clarified in this same figure. (Schuh et al. 2017, p. 50) In the context of this figure, "Cost regression analysis" refers to regression based methods like multiple linear regression model (e.g. MLPP) and "Linear performance pricing" to one-dimensional linear regression analysis (Schuh et al 2017, p. 25). The "Cost-based price modelling" means bottom-up calculation (Schuh et al 2017, p. 24). Based on the location in the lower right-hand corner, products or components on this area have high-demand power, but low supply power. In practice, this means that purchased parts on this area are relatively simple and there is tough competition between several suppliers. In most cases, the purchasing company probably has a pretty strong position in relation to suppliers due to competitive situation in the market. (Schuh et al. 2017, p. 11)

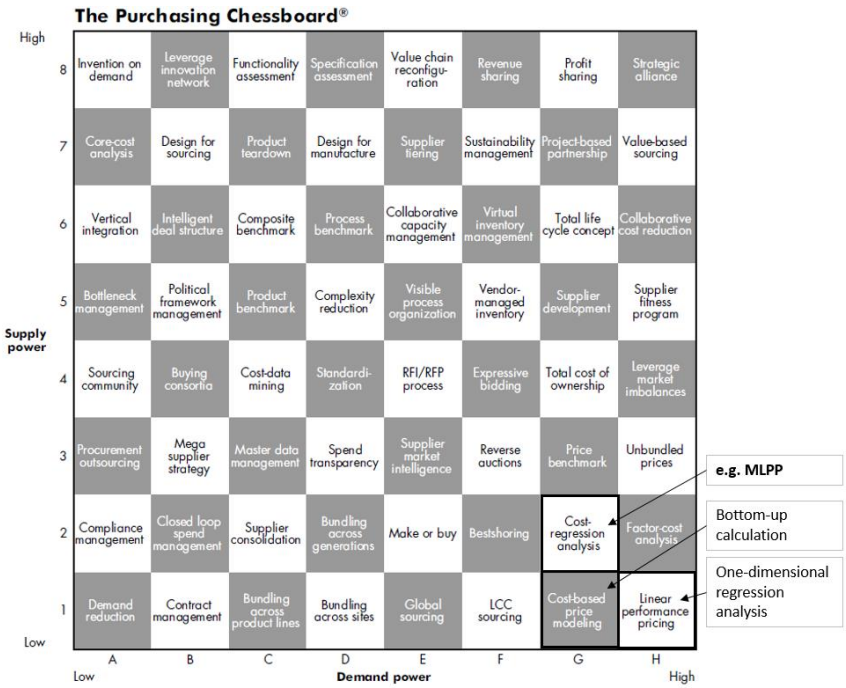


Figure 6. 64 methods of purchasing chessboard (modified) (Schuh et al. 2017, p. 50).

Especially in the automotive industry, the relationship between OEMs and suppliers is typically different compared to other industries, since OEMs have typically a significant power over their supplier base, whereas in other fields of industry the distribution of power is more equal or sometimes even on the supplier side (Newman & Krehbiel 2007). In figure 7, this same chessboard is used together with different purchasing categories of some construction equipment manufacturer. Based on the relationship between supply and demand powers, the selection of the most suitable purchasing methods for the each category may be done. (Schuh et al. 2017, p. 38) For example, steel structures with a high demand power, have many supplier candidates and the construction company was able to offer a very attractive package of products due to its strong position on the market. At the same time, regression based cost methods were utilized to identify the most potential parts for market test. (Schuh et al. 2017, p. 43) MLPP is typically useful with products, which include a lot of added value from suppliers (not raw material intensive products). Furthermore, purchased services have turned out to be an interesting area and their potential should not be underestimated. (Sieben 2021)

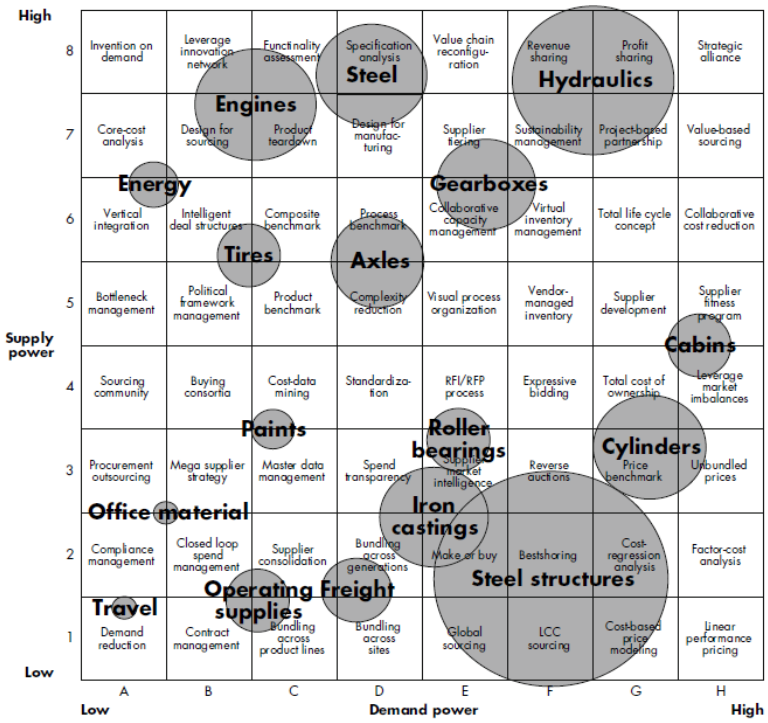


Figure 7. Purchasing chessboard in relation to different product categories (Schuh et al. 2017, p. 38).

2.4 Application strategies and process of MLPP

Based on Processbench e-learning material, application of MLPP can be divided into four different strategies as shown in figure 8. Especially among purchasing, the main benefit of MLPP is the identification of cost reduction potential within a high number of similar products, since MLPP allows users to process large amount of data relatively fast. The analysis results may then be utilized in supplier negotiation. (Processbench 2017) However, based on Leitol (2021), MLPP models are rarely used directly in negotiations. Instead, models are used as a tool to identify the potential parts. The tricky thing with MLPP is that models practically always include always certain amount of cheap parts, which may complicate the negotiation with suppliers. (Leitol 2021) This analysis method can also be used for comparison of products which are similar, but cannot be easily analyzed with traditional cost calculation methods. Additionally, MLPP facilitates recognizing possible issues in purchasing strategy by revealing all data points (prices in relation to technical value) which may support optimization of supplier base by taking the product, region and volume into consideration. Through the analysis, it is possibly easier to recognize possible

strengths and weaknesses of suppliers with certain type of products. (Processbench 2017) Prediction of prices for new products (target and should costing) may also be carried out by MLPP. By means of MLPP, it is also possible to do faster price analysis for new products based on certain characteristics of the product, which may not be completely ready in terms of final design. Additionally, it may be used as a supportive tool in research & development (R&D) for design-to-cost measures. One good example could be the reduction of technical complexity or high number of different product variants, which could possibly be identified with help of MLPP analysis. In some cases, the increase of technical value may also be relevant, if it looks that increased performance or some technical feature could be achievable with similar or slightly higher price. MLPP may also form an excellent basis for product optimization projects between purchasing and R&D, since the identification of favorable and unfavorable designs in terms cost is relatively easy. (Güthenke & Möller 2007; Processbench 2017)

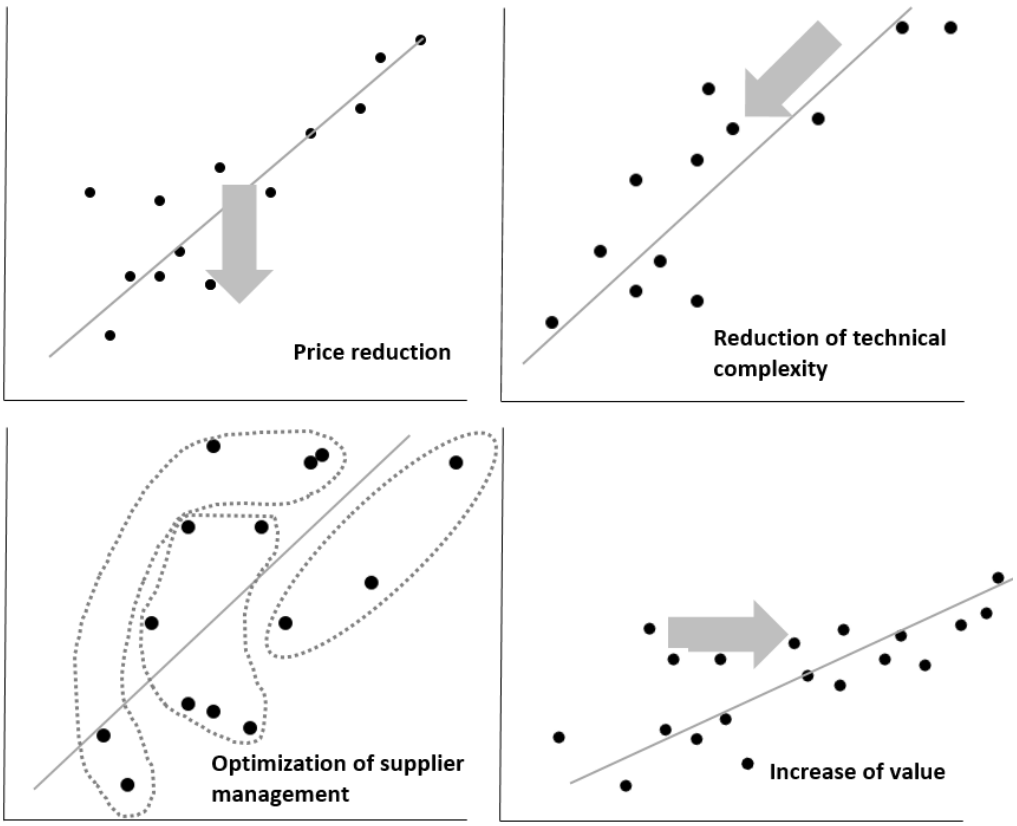


Figure 8. Application strategies of MLPP (modified) (Processbench 2017)

In addition to strategies shown in figure 8, MLPP also enables quick comparison of offered prices for new products as well as evaluation of already determined internal target prices as

presented in figure 9. (VDI 2018, p. 43) In case of hydraulic cylinders purchased by the client company, especially the quotation comparison for new cylinders in the early phase of the project would be useful, since it would enable faster cost estimates based on market benchmark for purchasing and R&D. Even though the design is mainly based on know-how of the suppliers (Mankki 2021, Karjalainen 2021), the early involvement of cost analysis methods like MLPP would likely increase the knowledge about the factors impacting on costs.

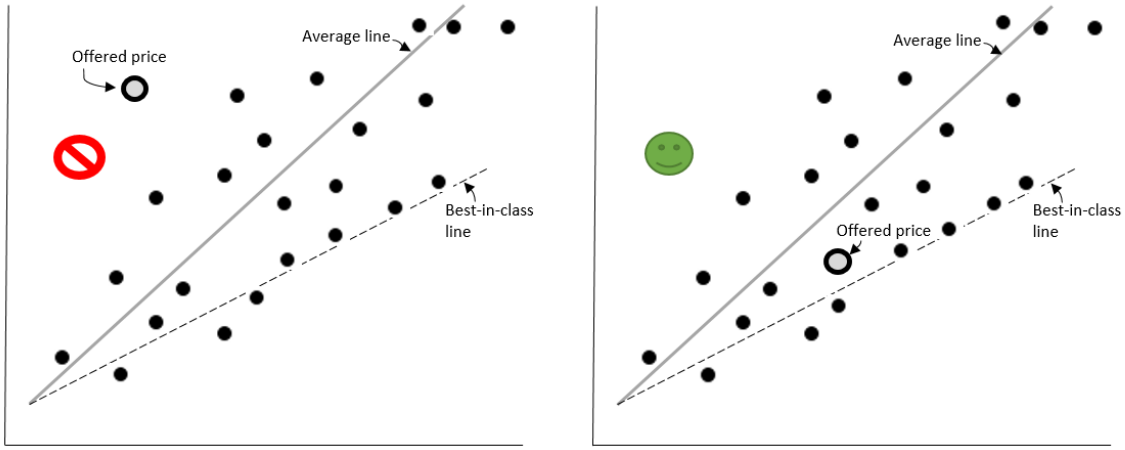


Figure 9. Quotation comparison for new products (modified) (VDI 2018, p.43).

In terms of required user skills, the creation of value graphs based on value drivers and price is relatively easy. However, evaluation and validation of the model (e.g. through bottom-up calculation) requires a lot of technical, product and cost related knowledge. The same applies to supplier negotiations, in which MLPP model is used as an argument for possible price reductions – the user must have comprehensive understanding of all details impacting on the functionality of the model. (Soltau 2018)

According to VDI (2018, p. 7) guideline, the application of MLPP typically consists of seven main phases introduced on the left side of the figure 10. However, despite the separation of the different process steps, an iterative approach is necessary during the whole process (note the arrows in figure 10) (VDI 2018, p. 17). Compared to the main steps determined by VDI 2018, Schuh et al. 2017 presents an application strategy including only four main steps on the right hand side of the figure 10, being a bit more straightforward and purchasing oriented. (Schuh et al. 2017, p 175)

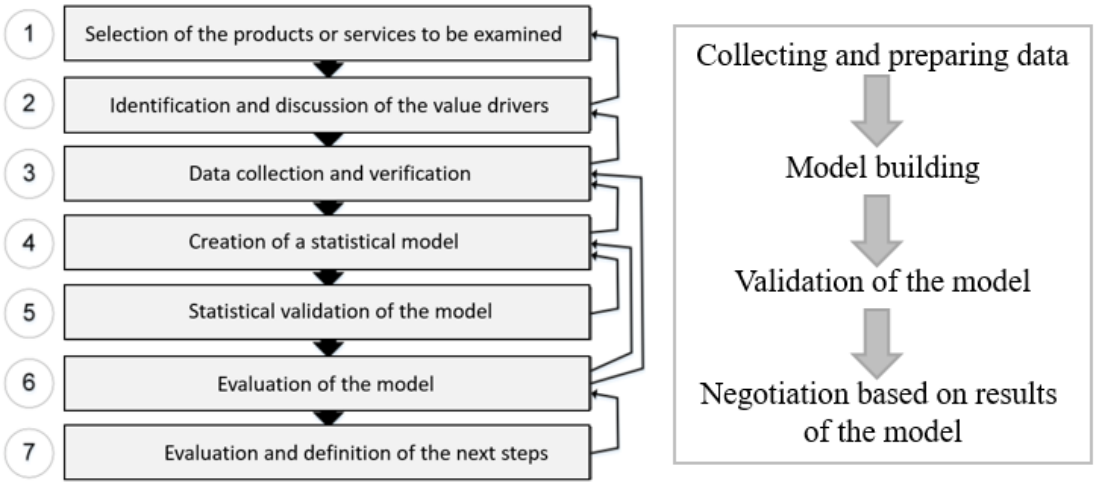


Figure 10. Main steps in application of MLPP (modified) (VDI 2018, p. 7; Schuh et al. 2017, p. 175).

2.5 Identification of value drivers

A value driver is a factor that impacts on the value of a product, component or service (Processbench 2021). At the same time, it describes the product and its application as well as possible manufacturing related aspects. Value drivers are utilized as data-elements (variables) in the regression equations when the model is built. (Hilti 2017; VDI 2018 p. 9) The identification of value drivers is typically a cross-functional exercise between different functions and participation of different functions like product experts, design and manufacturing engineers, finance, sales and marketing is highly recommended (Processbench 2017). In some cases, the interdependencies of value drivers should also evaluated during data collection phase (Güthenke & Möller 2007). Typical value drivers include parameters like length, weight, width, height, as well as thicknesses, material grades, surface treatment and other parameters, which have direct influence on production costs. However, sometimes value drivers may be based on performance, customer, market or regulatory parameters. (VDI 2018, p. 9) According to Güthenke & Möller 2007, the value drivers should take into account both material cost and production cost. Based on Sieben (2021), nowadays also sustainability (e.g. CO^2 emissions during production) has an increasingly important role in value driver identification. Some examples of possible value drivers are presented in figure 11.

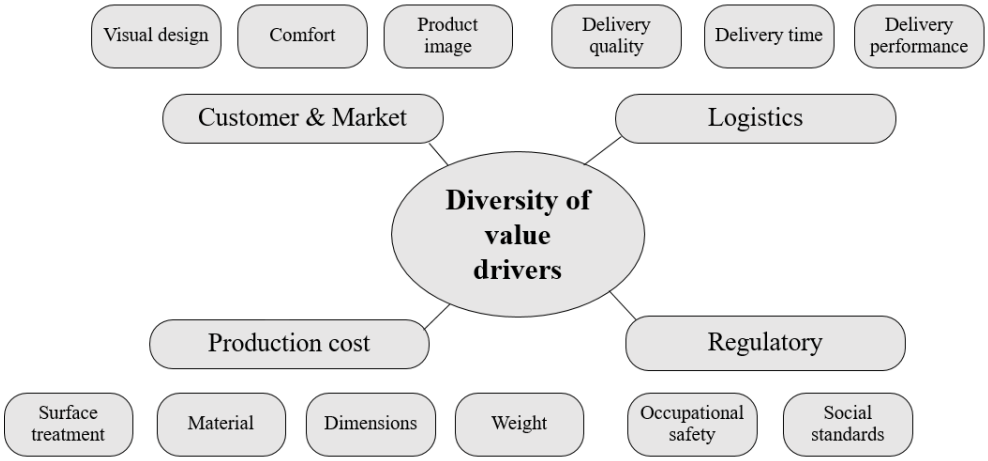


Figure 11. Diversity of value drivers (modified) (VDI 2018, p. 9).

One example of the identification process is presented through the Ishikawa diagram in figure 12, which can be utilized for value driver identification. This diagram also highlights the diversity of perspectives as well as the importance of cross-functional co-operation during identification process. (VDI 2018, p. 9-10). The goal during data collection sessions is to identify as many value drivers as possible (Hilti 2017).

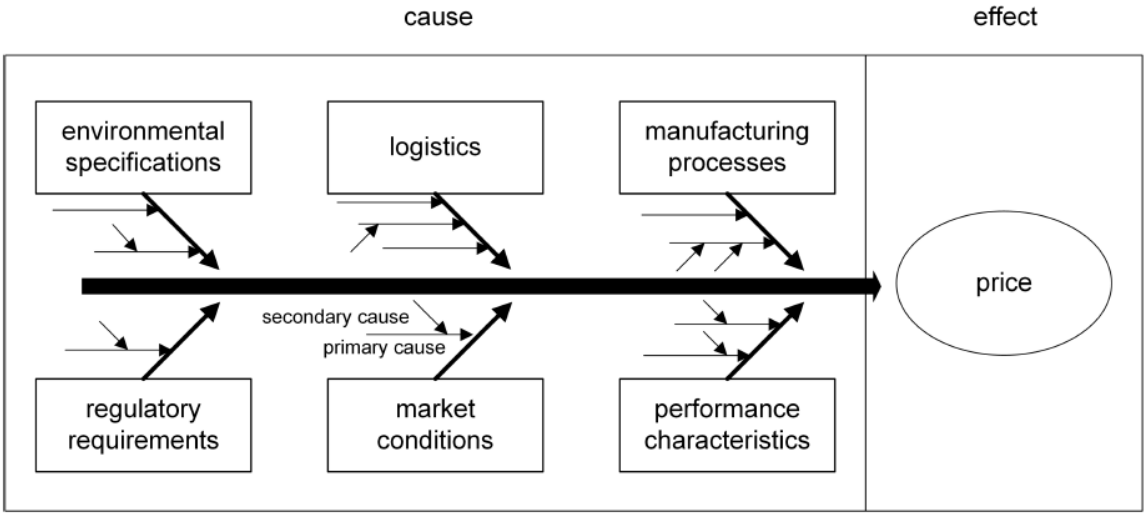


Figure 12. Identification process of value drivers (VDI 2018, p. 10).

In definition of value drives, value driver selection should be done based on criteria presented in figure 13. (Processbench 2021)

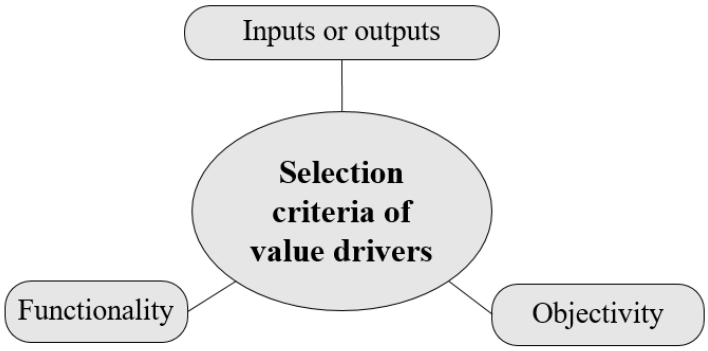


Figure 13. Selection criteria of value drivers (modified) (Processbench 2021).

The criteria of value driver selection is explained more detailed in the list below, including some practical examples from Processbench 2021. Furthermore, this list introduces corresponding examples for hydraulic cylinders:

- Consistent orientation towards either input or output of the product
    - The value refers to the usefulness or functionality of the product, not to the configuration. Example: In the assessment of race car engine, value driver should be based on acceleration or top speed instead of number of cylinders in engine.
    - The value driver is related to the characteristics of the product which has a major impact on costs. Example: The casting process cost is highly dependent on weight of casting → weight to be used as value driver.
    - In case of hydraulic cylinders, the output of the product would most likely be the cylinder force or the operational length. On the other hand, cylinders are relatively raw material intensive, which would perhaps justify the utilization of cylinder weight as a value driver.
- Functionality
    - The value driver selection to be done based on the major functionality of the product instead of auxiliary function. Example: The central processor power impacts significantly on overall performance of computer, but the maximum volume of speakers is just an additional feature → CPU power to be used as value driver instead of maximum volume.
    - With hydraulic cylinders, the similar example could be built around bearings- or grease nipples and cylinder tube inner diameter. Bearings

and grease nipples may be considered as additional feature in relation to inner diameter, since inner diameter impacts on the force produced by the cylinder.

- Objectivity
  - The value driver should be clearly measurable. Example. It is challenging to perform objective quantitative measuring on design of air conditioner. However, cooling capacity is measureable and it should be used as value driver instead of design.
  - Similarly, the design of piston rod or assembly method for the cylinder are challenging to measure. Thus, value drivers related to technical dimensions or output of the cylinder are preferable.

Value drivers can be either numerical (e.g. operational length of hydraulic cylinder in mm) or categorical (e.g. type of cylinders used in tractors (type based on application)). In general, numerical value drivers are preferred and excessive usage of categorical value drivers should be avoided. Based on Processbench user guide, the amount of required data points in relation to value drivers should follow instructions shown below in table 1. A general rule of thumb is that a stable model should have at least 30 data points (Soltau 2018). According to Schuh et al. (2017, p 175), the required number of data points should be more than 40, whereas Wakkee et al. (2014) recommends 10 data points for each variable. Based on existing experience and testing among the years, the threshold used in AGCO has been roughly 100 data points or 10 data points for 1 value driver. After all, this is heavily a case and product dependent matter.

Table 1. Number of required data points in relation to selected value drivers (modified) (Processbench 2021).

Value driver	**Number of data points**
Value driver 1	5
Value driver 2	5
Value driver 3	5
+5 additional data points	5
Total number of data points	20

Often finding the appropriate group of variables (in this context of value drivers) is called a variable selection problem (Montgomery et al. 2012 p. 328). A general principle in value driver identification is to select only the necessary value drivers, which still comprehensively explain the pricing of the products in the model. In case possible identified value drivers correlate with each other (e.g. volume of battery and capacity), either of these should be removed to avoid the distortion of the mathematical model. (Processbench 2021) Güthenke & Möller (2007) describes the identification of value drivers as an optimization problem, in which the aim is to achieve high regression with low number of value drivers.

In terms of analysis, categorical value drivers are a bit trickier compared to numerical value drivers. Perhaps the most important thing to keep in mind with categorical value drivers is that selecting the categorical value driver usually rapidly increases the internal number of value drivers. The increased number of value drivers again increases the number of required data points in the model (as already presented in table 1 above). With categorical value drivers, ordering (e.g. "iOS is more than Android"), numbering (e.g. iOS = 1, Android = 2 etc.) or summing of the values is not recommendable, because it will most probably distort the model. Thus, in case it is obvious that for example "iOS is more than Android 4", the physical or obvious property should be found to support this statement. In other words, numerical variables should be utilized to explain differences between two different operating system. However, in some cases it may be challenging to do quantification for categorical value drivers. If the price impact of certain property or feature is obvious, during data processing it can be handled by filling in "1" for those products which have this specific property, and "0" for the ones without it. (Processbench 2021)

One practical method for categorical value driver processing is the utilization of value driver folding approach. In practice, matrix type data input with selected categorical value drivers and related numerical value drivers like shown in figure 14. Typically, matrix input improves the quality of the regression and most importantly, the number of required value drivers is reduced (in this example from 4 to 3). This method is usable when categorical- and numerical value drivers have mutual influence on regression. (VDI 2018, p. 11-12)

	Numerical	Categorical
	Weight (kg)	Material
Component 1	1,2	***Aluminium***
Component 2	3,5	***Steel***
Component 3	0,8	***Copper***

= 1 numerical + 3 categorical value drivers

	Numerical	Numerical	Numerical
	Aluminium	***Steel***	***Copper***
Component 1	1,2	0	0
Component 2	0	3,5	0
Component 3	0	0	0,8

= 3 numerical value drivers

Figure 14. Matrix type value driver bundling (modified) (VDI 2018, p. 13-14).

Another practical approach for value driver processing is an aggregation of numerical value drivers, which also reduces the number of value drivers and may also increase the statistical quality of the model. For instance, length and diameter of a cylindrical component can be aggregated into volume. (VDI 2018, p. 11) The more value drivers are identified, the more extensive is the data collection work (VDI 2018, p. 10). Additionally, selection of unnecessary variables typically add statistical "noise" to the model and degrees of freedom will be wasted. (Faraway 2002, p. 117). In order to reduce the number of value drivers already in the identification phase, it is recommendable to check the relevance of an identified value driver in relation to the product and the possible dependencies between value drivers. In addition, the value driver should also be evaluated from a required data collecting effort perspective as well. (VDI 2018, p. 10) It is pretty common with practical regression problems, that the respective analyst recognizes a high number of potential variables, of which only a few are really important in terms of regression function. (Montgomery et al. 2012, p. 328)

2.6 Technical value

Technical value is the result of statistical mathematical calculations based on the selected combination of value drivers, and it is one of the most important outputs of the MLPP

analysis. The 45 degree regression line fit based on technical value (x-axis) and price (y-axis) describes the relationship between technical value and price. Even though the technical value is unique for each data point in the model, the technical value is calculated by taking into account the influence of other products in the same analysis. In terms of mathematics, the technical value is calculated by the software based on the PLS regression method and the optimum weighting of each value driver in the calculation of the technical value is done by the software. The higher the technical value is, the more complex the product is. In case the product is located at the regression line, it means that pricing of the product in relation to selected value drivers (technical value) and other data points is on the average level. On the other hand, all the data points on top of the average line are relatively expensive in relation to technical value, and data points below the average line are good value for the money, as illustrated in figure 15. (Processbench 2017)

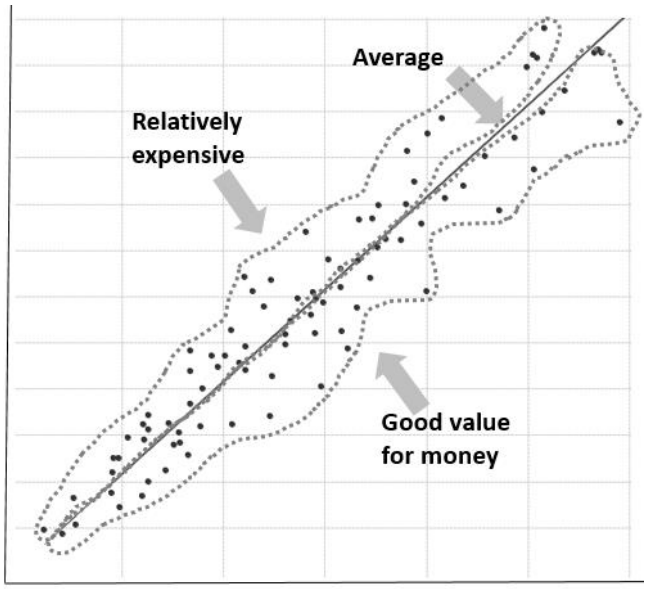


Figure 15. Data points based on technical value (AGCO MLPP model for forgings). (SCA 2021)

In cases where the price does not seem to be on reasonable level, this information may be utilized for example in price negotiations with suppliers and it may help to turn the discussion from prices to actual production costs. It may also facilitate to understand possible pricing principles for certain product portfolios, when facts impacting on technical value are known. One practical example of technical value formula is presented below (Processbench 2017):

$$Technical\ value = -369 + 3{,}74 * [Width]\ mm + 13{,}96\ [Display\ diameter]\ inch + \mathbf{127{,}24} * [\boldsymbol{CPU\ clock\ rate}]\ GHz + 3{,}11 * [Internal\ memory]\ GB + 0{,}01 * [RAM\ size] MB + 43{,}36 * \left[\frac{3G}{UMTS}\right] y/n \quad (1)$$

The example formula (1) includes 6 different value drivers of which 5 are numerical (width, display diameter, CPU Clock rate, internal memory, RAM size) and one categorical (3G/UMTS). Each of these value drivers has their own unit and for example one additional GHz for CPU clock rate costs 127,24 € (bolded in equation 1). In overall, this technical value formula is the base for the regression line and the first term (-366,69) of the formula determines the intersection of the y-axis. (Processbench 2017)

2.7 Data collection for MLPP model

Comprehensive and high-quality data collection is the basis of a good MLPP model, and like in all data-based tools, the usability of the MLPP depends highly on the quality of the collected data (Newman & Krehbiel 2007). Based on Harrel (2001, p. 6), with help of properly planned data collection, the statistical predictive model is more accurate and useful. Data collection is perhaps the most critical part of the MLPP process (Leitol 2021). Thus, it is necessary to ensure collected information is correct and also make sure that it has been entered properly into the system or input file. In terms of reliable model results, the overall data quality is in a key role. As in traditional data quality evaluation, metrics like free-of-error, completeness and consistency apply to MLPP data collection as well. (Pipino et al. 2002, p. 213; VDI 2018, p. 12-13) The effort needed for data collection depends highly on the number of identified value drivers and especially on the source of the data. The data collection made directly from the IT system (e.g. enterprise resource planning (ERP)) can be done efficiently and relatively fast. However, the collection of technical specifications and characteristics must usually be done manually based on technical drawings or 3D data, which significantly increases the required effort and time. (Schuh et al. 2017, p. 175; VDI 2018, p. 10) For the most part, the data collection must be typically carried out manually. However, there are certain products (e.g. simple sheet metal components), for which the data collection is possible to do automatically. (Leitol 2021)

During data collection, the semantic and syntactic aspects should be used. Especially with numerical value drivers, the correctness of numerical data must be checked with regard to their meaning. Sometimes certain numerical values may have a non-numerical meaning (e.g. DIN strength classifications for hardware parts). In order to ensure comparability of all products in the model, the input should follow consistent principles and it is necessary to use the same approach with separators like periods, apostrophes, commas and decimal separators during number formatting. The same requirement applies to all units like dimensions and currencies which should be adjusted in the same manner. The impact of value drivers is highly based on input data and the unit of the selected value driver. In case of categorical value drivers, a consistent spelling is essential to make sure the model works properly. If the model includes products with incomplete information, those should be cleaned up or excluded from the analysis, since for example inserting "0" to replace missing data leads most probably to an incorrect calculation result. Duplicates should also be avoided in order to prevent distortion of the calculation. (VDI 2018, p. 14-15) With MLPP for hydraulic cylinders, especially the pairs (left and right side) should be analyzed carefully, since in many cases, the design and pricing of both sides are similar, which could over strengthen the regression (Berz 2021). In terms of part numbers, MLPP software does not allow to use an input data file with duplicates, but in case of pairs the checking must be done manually based on technical specifications and prices. Another important point is related to objectivity of data, since collected data should not be depended on the subjective assessment of the user. This also enables reproducibility of the data. In addition to data related to possible value drivers, it may also be wise to collect additional data related to products in the analysis, which may be used as supportive data during the evaluation of the statistical model. Possible additional data options are introduced in figure 16. (VDI 2018, p. 15-16)

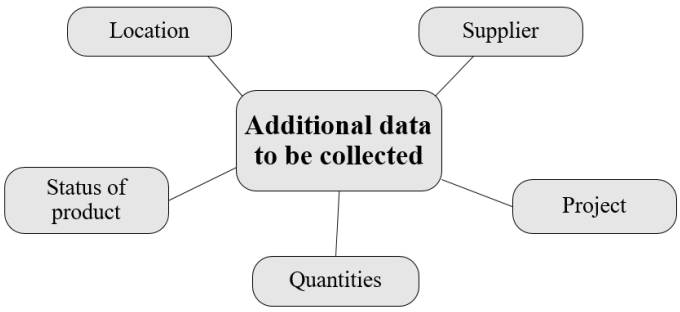


Figure 16. Additional data to be collected during data collection (modified) (VDI 2018, p. 15-16).

Even though the quantity of the products does not necessarily represent a value driver, it is important additional data as the potential calculation of the model is usually based on annual quantities. Depending on product, the product price may be heavily linked to the status of the product like pre-series, series and post-series prices. Thus, it is generally advisable to collect the status information as well. In most cases models should be based on series prices to ensure the quality level of the model. (VDI 2018, p. 15-16) However, when collecting pricing data from the system, it is necessary to check the existing contracts, since it is quite common that there are several different delivery conditions, which may impact on the price of the product. With certain product groups, also development costs may be added as a part of the sales price. (Güthenke & Möller 2007) Sometimes, there may be different kind of price agreements that change or impact on prices (e.g. fixed price reductions, unusual delivery quantities, additional levies, cross subsidization etc.), in which the current cost structure does not correspond real costs of production. This kind of products could be marked with special price status to support the evaluation process of the model. Collection of additional data may also allow to use this additional data as a filter. (VDI 2018, p. 15-16) Filter is basically a selection criteria, which is not a cost- or value-driver, but may offer an interesting view on selected sample in the model (Hilti 2017). Filters can for example, be based on production location (see figure 17), supplier and volume as well as on project names given for certain product groups by the company. In certain conditions, these filters may be utilized later as value drivers as well. (VDI 2018, p. 15-16) Hilti performance pricing training document (2017) recommends, that in case model creator is not sure whether to use certain criteria or property as a value driver, it should be used as a filter only.

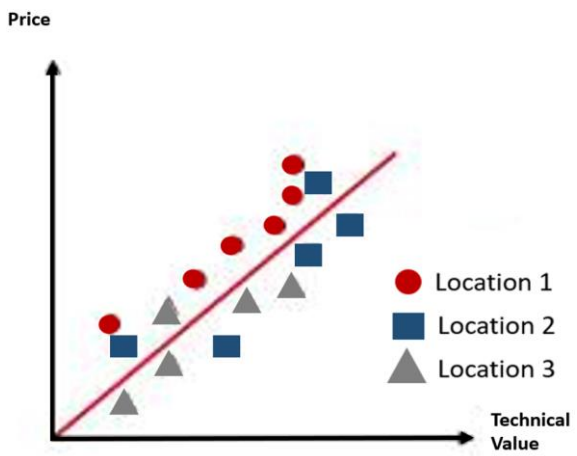


Figure 17. Manufacturing location used as a filter (modified) (Hilti 2017).

With products whose prices are highly depended on volatile cost factors (e.g. raw materials), it may be wise to index them (VDI 2018, p. 8). In practice, it means standardization of purchasing prices due to changes for example in raw material cost. However, in most cases changes caused by raw material are linked to the selected index. The first step is to determine the raw material price for the reference year (when price was agreed). Secondly, the share of the raw material cost must be determined for each product. Then, the comparison is done between the reference year raw material price and current raw material price. Finally, the standardized price will be calculated based on factors mentioned earlier. An example of price standardization is introduced below (VDI 2018, p. 15):

$$P_s = \frac{RM_s * P_c}{\frac{RM_{pc}}{RM_{pry}}} + (1 - RM_s) * P_c \qquad (2)$$

In equation (2) above, P_s = Price of the product (standardized) [€], RM_s = Raw material share [%], P_c = Price of the product (current) [€], RM_{pc} = Raw material price (current) [€/kg or €/ton] and RM_{pry} = Raw material price (reference year) [€/kg or €/ton] (VDI 2018, p. 15). However, perhaps the most practical way to overcome this raw material impact challenge is to select a certain time period (e.g. a certain quarter) and use prices only from this period. This also ensures that the market situation is equal for all products. (Berz 2021)

2.8 Creation of a statistical model

According to Backhaus et al. 2018, the actual creation of statistical regression model consists of five different phases as shown in figure 18 (Backhaus et al. 2018, p. 63). Based on VDI 2018, these five main phases can be divided into two main phases; model development and -validation. (VDI 2018, p. 18) This chapter discusses the main points regarding the model development as well as the coefficients to be followed during different phases of the model creation. The validation of the model is discussed on the next chapter.

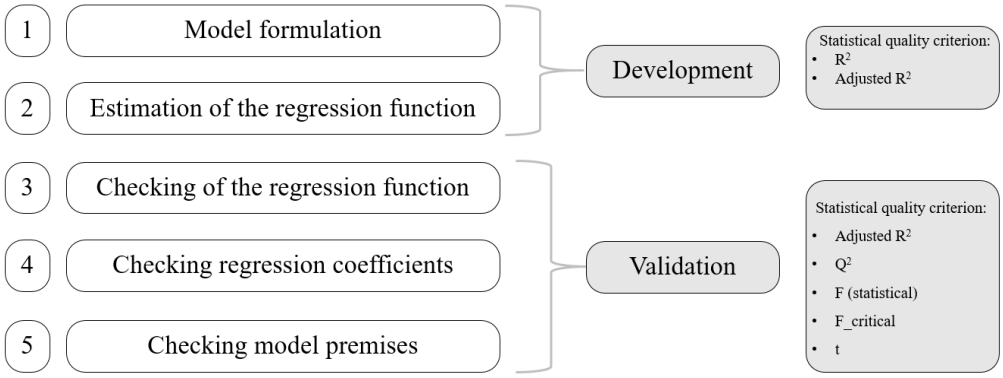


Figure 18. Main phases of regression model creation and indicators to be followed during different phases (modified) (Backhaus et al. 2018, p. 63; VDI 2018, p. 18).

Before actual model development, certain conditions must be met to ensure qualified results. As discussed in chapter 2.5, the amount of data points need to be sufficient, since too low number of parts will probably result in poor quality of regression. Another important point is related to complexity of technical specifications and variation between specifications. The required preconditions are listed to table 2. (Schuh et al. 2017, p. 174)

Table 2. Preconditions to be checked before model creation (modified) (Schuh et al. 2017, p. 174).

Required preconditions:	**Possible consequence if the condition is not met:**
Sufficient number of data points	Too high influence of outliers --> low statistical quality
Enough technical complexity in specifications (value drivers)	Too simple parameters may become dominant --> distortion of the model
Enough variation among the selected specifications (value drivers)	Minor differences among specifications --> model is not able to perform differentiation between products

Furthermore, before value driver selection it would be advisable to consider the overall approach for model building. Based on product nature, the model may be built based on performance- or cost parameters. However, often the model is a mixture of both because in many cases the ignoring of cost parameters may be challenging. For example with hydraulic cylinders, the performance based parameters could be cylinder force, lifetime, stroke or operational length. With certain products, it may also be impossible to build a model based on performance parameters, since the cost of the product may be so heavily driven by cost parameter related characteristics (e.g. weight). Performance parameters are still preferable since those describe the things the customer is usually willing to pay for and which describe

the quality of the product. At the same time, it may allow us to do comparisons of the cost efficiency of performance among the supplier base (identification of best-in-class suppliers). (Berz 2021)

Before actual model formulation, it is recommended to investigate the correlation among already identified value drives (independent variables). This may be carried out by using a scatter plot, in which one variable is placed on x-axis and the other on y-axis. Based on the positions of the data points, it is possible to do evaluations about correlations between value drivers. In order to get a better understanding of the dependencies between all value drivers, scatter plot matrix may be a useful tool. An example of scatter plot matrix is introduced in figure 19, in which the dependencies between three potential value drivers (x_1, x_2 and x_3) are shown. It also reveals the relationships between price (P) and each value driver, which is probably the most interesting thing to be examined in the beginning of model creation. (VDI 2018, p. 17) However, although the relationship between independent variables and price offers valuable information of potential value drivers, it should be noted, that sometimes variables with low correlation with the price may actually be important, when used in the full model with other variables. (Wakkee et al. 2014) At the same time, scatter plot matrix enables us to explore interdependencies between variables, which helps to exclude certain combinations of variables to be used as a base for technical value. (VDI 2018, p. 17)

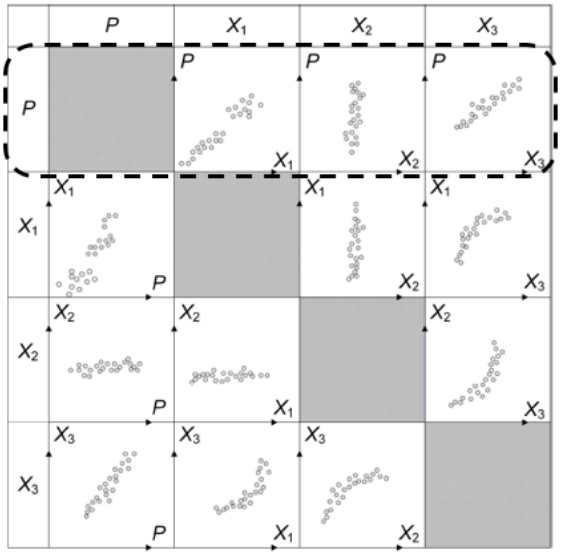


Figure 19. Scatter plot matrix indicating the interdependencies between potential value drivers and also price (P) (see top row) (modified) (VDI 2018, p. 19).

A simple practical example of a scatter plot for a cube is presented in figure 20, where length (*L*), volume (*V*) and weight (*W*) are selected to be possible value drivers. Based on scatter plots, it is obvious that both volume and weight are potential value drivers, due to linear correlation with price. However, only one of these may be used as a value driver due to high interdependency between volume and weight of the cube. (VDI 2018, p. 17)

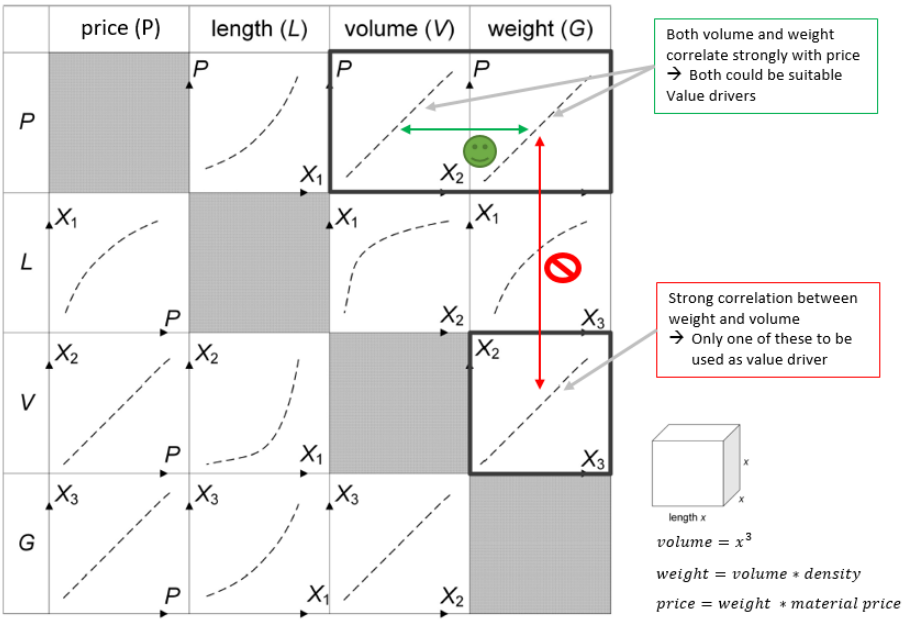


Figure 20. Scatter plot matrix (modified) (VDI 2018, p. 19).

The results of the scatter plot matrix may also be analyzed through the coefficient of determination, and use this as a selection criteria for value drivers (e.g. $R^2 > 0{,}4$ → potential value driver) (VDI 2018, p. 17). Similarly, the correlation coefficients can be presented in a matrix table, which facilitates the observation of dependencies between variables based on numerical values (Shrestha 2020).

However, despite comprehensive comparison of the possible dependencies between variables, sometimes there may still be linear dependencies among variables that cannot be detected by checking the correlation coefficients between variables only. This is called as multicollinearity. In practice, collinearity means that too many variables are trying to do the same job (Rydén 2014; Faraway 2002, p. 117). For detection of multicollinearity, there are also more advanced methods available. In addition to the correlation coefficient technique, Shestra 2020 introduces two other primary techniques, which are variance inflation factor

and eigenvalue methods. (Shestra 2020) However, with models using PLS regression the collinearity among variables is usually not a problem (Wold 2001).

The model building and estimation of regression function is usually started by selecting the variable with highest adjusted R^2 to be used as first value driver. Based on adjusted R^2, additional value drivers are then added to the regression step by step. (VDI 2018, p. 19 & 20) At this point, working with new variables may be based on forward selection (a.k.a. forward addition), backward selection (backward elimination) or the stepwise selection. In practice, forward selection means adding new variables sequentially, whereas backward selection is the opposite of this (all possible variables are entered in the first place and then removed one by one). Stepwise selection is basically the combination of forward- and backward selection, since it includes both adding- and removal of variables. Typically, adjusted R^2 is used as a criterion when selecting and deselecting the variables. Based on experience, the best results are achieved through iterative working manners with stepwise selection method. (Yan & Gang Su 2009, p. 171-173; IBM) According to recommendations from statisticians, backward selection is typically the most reliable method especially with standard software (Heinze & Dunkler 2007). Depending on the selection method, sometimes t-values may also be utilized as criterion during variable selection in which case the selection of the value drivers would be done based on magnitude of t-statistic. (Ryan 1997, p. 228) Based on the interview with Dr. Berz before MLPP model creation for hydraulic cylinders, R^2 and Q^2 may be used as main criteria during model development, whereas F and F critical should be used as supportive indicators (Berz 2021).

2.9 Statistical validation of the model

As presented earlier in figure 18, the statistical validation of the model can be separated into three main phases; checking the regression function, regression coefficients and model premises. Usually, the first step in regression function validation consists of evaluating R^2 and adjusted R^2. (VDI 2018, p. 22) Mathematically, R^2 describes the relationship between residual (a.k.a. regression) sum of squares (RSS) and total sum of squares (TSS) like shown below (Ruppert & Matteson 2015, p. 228; James et. al 2013, p. 69-70):

$$R^2 = 1 - \frac{RSS}{TSS} \tag{3}$$

In equation 3, RSS measures the predictable variation whereas TSS measures the total amount of variability inherent in the response (dependent variable) before the regression is performed. (Ruppert & Matteson 2015, p. 228; James et. al 2013, p. 69-70) In other words, R^2 measures the proportion of variability in Y that can be explained by using x (James et al. 2013, p. 70). Within performance pricing analysis, the dependent variable (Y) is usually purchasing price and independent variables (x) are the selected value drivers (Saphirion 2016). An R^2 value close to 0 means that regression is not able to explain much of the variability, whereas R^2 values close to 1 indicates that big portion of the variability is explained. In case R^2 is 0.6, it means that roughly two-thirds of the variability is explained by the regression. However, sometimes it may be challenging to determine what the sufficient threshold is for R^2 and this usually is highly dependent on application. (James et al. 2013, p. 70) Even though the model with higher R^2 generally represents a better model, this is not always the case and thus R^2 and adjusted R^2 should be carefully used for the evaluation of the model (Bedre 2021). Based on interviews with Leitol (2021) and Sieben (2021), there is not any determined threshold for adjusted R^2. A model with relatively low adjusted R^2 may bring up potential data points, and the basic idea is to identify outliers from the model (Sieben 2021). An adjusted R^2 differs from R^2 by taking into account also the number of data points as well as the variables (value drivers in case of MLPP). Especially in cases, when models get larger in terms increased selected variables, the R^2 value tends to increase, even though this does not correspond to reality. Thus, adjusted R^2 is more advisable value to be used for comparison, especially for models with different amount of variables. In principle, the value of adjusted R^2 is always lower compared to R^2 (Bedre 2021). A Mathematical explanation for adjusted R^2 is shown below (James et al. 2013, p. 212):

$$Adjusted\ R^2 = 1 - \frac{RSS/(n-d-1)}{TSS/(n-1)} \qquad (4)$$

In equation 4, n=number of data points and d=number of variables. (James et al. 2013, p. 212) However, the coefficient of determination is not the only factor which impacts on validity or quality of the model. (VDI 2018, p. 22) With help of F (statistical), it is possible to identify, how much of the price deviation can be explained through the price function (Backhaus 2018, p. 80). F takes into account the sample size and the variance. The greater the F value is, the more likely the selected combination of value drivers describe the pricing of the products. In general, F value could be described as an indicator, which reveals, how

realistic the regression function is. (VDI 2018, p. 22) In case F is lower than F critical, the model can be found to be meaningless (Sieben 2021). The third value for regression function validation is Q^2, which describes the model's stability and robustness. (VDI 2018, p. 22) Basically, the calculation of Q^2 is based on removing individual data points from the model and following how the price function is acting. This method is also known as "Predictive sample reuse technique" (Geisser, 1975) and "Cross-validatory method" (Stone, 1974). It was initially developed by Stone and Geisser and is thus nowadays called as Stone Geisser criterion (Chin 1998). According to Sieben (2021), Q^2 value below zero indicates that something is fundamentally wrong with the model.

For validation of regression coefficients, VDI 2018 presents two different methods; t-test and confidence intervals. In practice, this phase of validation may facilitate to evaluate the necessity of certain value drivers. (VDI 2018, p. 22) When validating individual regression coefficients, the special attention should be paid to possible negative coefficients, since typically negative coefficient indicates that the increasing technical functionality leads to lower price (Güthenke & Möller 2007). Generally, t-test describes the relationship between "signal" (typically a mean) and "noise" (measure of standard error) related to certain variables (Berkman & Reise, p. 52). In validation of regression coefficients, t-value is calculated by dividing the regression coefficient by its standard error. T-test is targeted for individual coefficients, whereas F-tests describes the relevance of whole model. (Backhaus 2018, p. 84; VDI 2018, p 23). In case of MLPP software used in this study, t-test (t-value) is not shown by default, but it must be enabled manually in "Preferences" → "Experimental features". The confidence interval describes the variation around the selected regression coefficient and it may be an important factor, especially if the sign (+ or -) changes among the selected coefficient (VDI 2018, p. 24).

As a first part of checking the model premises, the residuals must be checked (VDI 2018, p. 24). In practice, residuals mean the difference between the regression line and the selected data point as introduced in figure 21 (Gareth et al. 2013, p. 62). In most cases, the sum of positive and negative residuals should be zero and those are canceling each other out. However, in case residuals have a correlation with each other, among value drivers or residuals seem to increase with more expensive products, the pricing formula and/or value

drivers should be checked. Residuals should remain constant through the whole model, despite the price level of the product. (VDI 2018, p. 24)

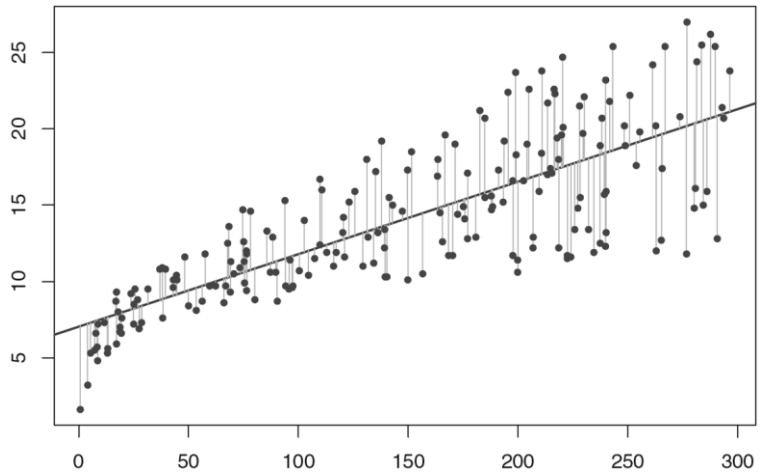


Figure 21. Example of regression residuals (Gareth et al. 2013, p. 62).

In case the formula of technical value is negative (negative y-axis intersection), it may indicate that model includes nonlinear relationships. An example of nonlinearity and negative technical value are shown on the left side in figure 22. (VDI 2018, p. 25)

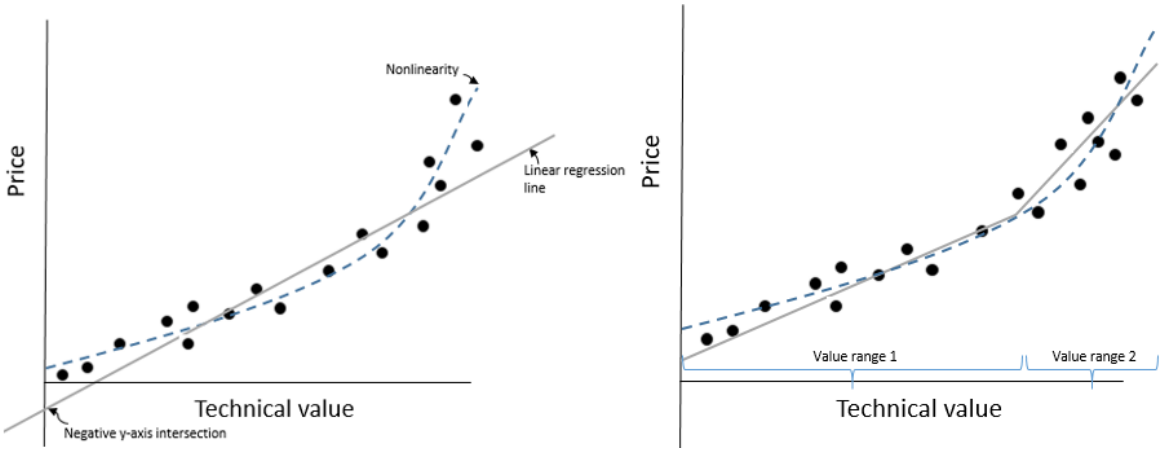


Figure 22. Nonlinearity with negative technical value (left) and model's split (right) (modified) (VDI 2018, p. 26 & 28).

VDI 2018 presents several possible reasons for nonlinearity. First of all, the nonlinear interdependencies among value drivers should be checked. In some cases, value drivers may actually include nonlinear values, which can be the case e.g. with products manufactured with tooling. In case the tooling is amortized in piece price, the proportion of tooling cost is

higher with lower volume products, which probably causes nonlinearity and distorts the model. If the model includes products manufactured with different manufacturing processes (high tech vs. low specification components), the correlation between products may be nonlinear. In this kind of cases, the nonlinearity may be supplier dependent as well. Market situation (e.g. supply and demand) as well as other price fluctuations like raw material price changes may lead to nonlinearities. However, by selecting the specific time period or utilization of price standardization equation (2) introduced earlier, market-related fluctuations may be reduced. (Berz 2021; VDI 2018, p. 25) Figure 23 presents some typical reasons for nonlinearity.

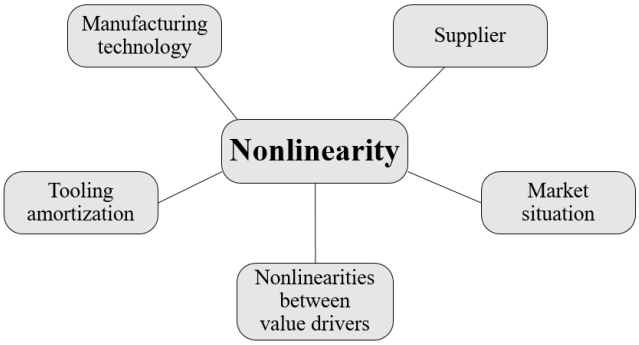


Figure 23. Possible reasons for nonlinearity (modified) (VDI 2018, p. 25).

There are certain methods and tricks available, which may facilitate to handle nonlinearities in the model. In general, modification of value drivers is advisable, e.g. usage of areas or volumes instead of lengths or widths. In some cases, multiplicative inverses may be used to convert nonlinear model into linear. (VDI 2018, p. 27) One practical example of the nonlinearity is related to MLPP model for hydraulic hoses (see figure 24), in which clear nonlinearities were identified, when the hose diameter and length were initially selected to be used numerical value drivers. In order to avoid nonlinearity, quadratic values of the inner diameter as well as utilization of categorical value drivers were tested. This did not solve the nonlinearity, but after all, the multiplication of the inner diameter and length provided linear results, and improved the statistical quality of the model as well (Processbench & AGCO 2017).

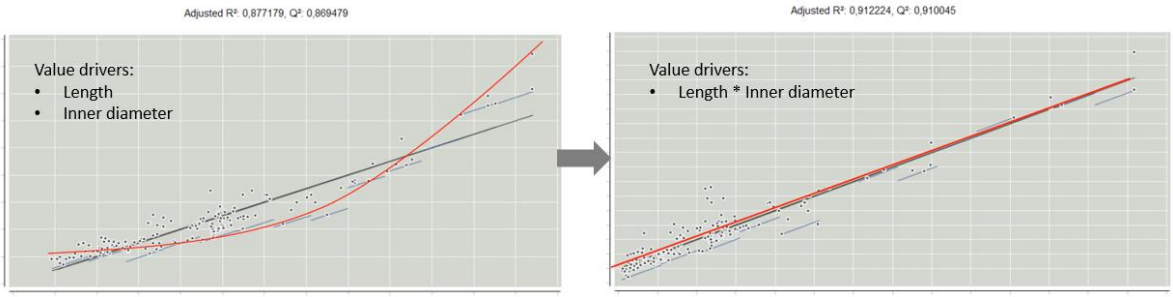


Figure 24. Practical example of dealing with nonlinearity. MLPP model created in AGCO for hydraulic hoses (Processbench & AGCO 2017).

In some cases, one option is to split the model in half and use two different value ranges with their own slope coefficients as shown in figure 22 (right side). After all, the most important thing is to use MLPP as a tool which approximates the price equation for industrial applications, instead of building too complex mathematical equations. (VDI 2018, p. 27-28)

2.10 Evaluation of the model

According to VDI 2018, the next phase after statistical validation is the evaluation of the model based on application context. In practice, this phase evaluates the model's ability to explain the pricing in relation to existing information, expectations and more detailed cost calculation models. In this phase of evaluation, the main focus is on value graph and locations of data points. Of course, the assessment of price equation (technical value) is a crucial part of evaluation as well. (VDI 2018, p. 28-36) In practice, this means critical evaluation of selected value drivers. Even though certain combination of value drivers would show high statistical relevance to price, but the value driver cannot be justified from the technical perspective, this kind of value drivers should be excluded. (Güthenke & Möller 2007). Based on practically approved methods presented by VDI 2018, the table 3 summarizes the areas of evaluations with required actions.

Table 3. Factors to be checked during evaluation of the model based on VDI 2018 (modified).

Area of evaluation		Action	
Price equation	Check if price equation, coefficients and signs (+ / -) make sense	Yes	No actions needed
		No	Review the model
Extreme values of the data points	Check the data points with highest and lowest technical values in relation to expectations	Yes	No actions needed
		No	If the location does not make sense, review the model
Leverage points	Check if point(s) with significantly higher technical value fits to the model.	Yes	Include the data point (increased validity range)
		No	Exclude the data point
Point clouds	Check if model clearly includes individual point clouds	Yes	Split the model
		No	No actions needed
Supplier	Compare if the recognized price level of suppliers to model results	Yes	No actions needed
		No	Review the model
Bottom-up calculation (true cost analysis (TCA) or cost structure analysis (CSA)	Calculate e.g. 3 data points to check if bottom-up calculation supports model results	Yes	No actions needed
		Systematically higher/lower	Check the reason
		Not identical or systematical with the model	Make a comprehensive review on the model
Volume	Compare the ratio between manufacturing volume and price	Perform this especially for products, in which volume has significant impact on pricing	Test volume as one value driver
Purchasing rules & agreements	Check the commercial factors which could distort the model	Tooling amortization, development costs, ex-rates, volume-based pricing agreements etc.	However, these factors should be checked already in data collection phase!

The figure 25 represents the evaluation of extreme values (on the left), point clouds (middle) and points to be evaluated through bottom-up calculation (right). As proposed in table 3, in case the model includes clear point clouds, the model should be split and analyzed separately (Güthenke & Möller 2007). With bottom-up calculations, it is advisable to enter respective bottom-up calculation results to the value graph by using same technical value with the analyzed data point, but the result of bottom-up calculation will be used as a price (VDI 2018, p. 31).

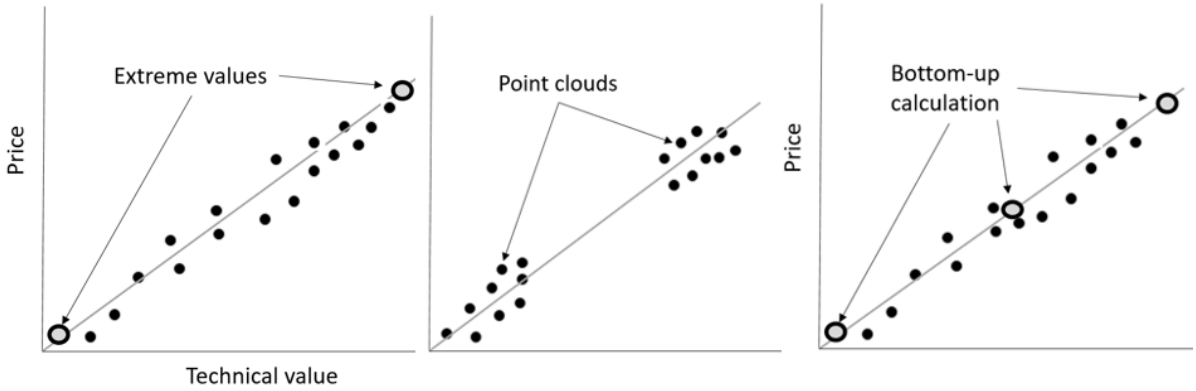


Figure 25. Evaluation of extreme values and point clouds as well as the points to be evaluated through bottom-up calculation (modified) (VDI 2018, p. 29-31).

As introduced in this chapter, the creation of MLPP model consist of multiple steps, of which each plays a significant role in the outcome of the process. The model creator should have a comprehensive statistical as well as technical understanding, which are both in the key position in terms of successful model creation. (Schuh 2017, p. 2017) Furthermore, in addition to statistical and technical aspects, there are multiple other factors, which may have affect pricing and should be considered as potential value driver. Thus, the cross-functional co-operation should be utilized, especially in the phase of value driver identification and selection. (Güthenke & Möller 2007) In case of hydraulic cylinders, due relatively high technical variability among cylinders, there are several possible factors, which may have an impact on pricing. (Zhou, Chen & Gao 2013) For that reason, it is necessary to have a sufficient understanding about all the technical details and properties, which should be taken into account during actual MLPP model creation.

3 HYDRAULIC CYLINDERS

Hydraulic cylinders are one of the most common actuating devices used in fluid power systems (Mobley 2000, p, 80). Cylinders are actuators which convert the hydraulic power into mechanical power during the linear motion of the piston and piston rod. The amount of mechanical power delivered to the load depends on fluid pressure and flow rate, which are controlled through hydraulic control valves. (Galal Rabie, p. 251) This chapter covers the main technical features and characteristics of the cylinders involved in the study.

3.1 Classification

Hydraulic cylinders are typically classified based on their functionality as follows: single-acting, double-acting, tandem three position and telescopic (Galal Rabie 2009, p. 259). Single-acting cylinder is the simplest type having only one oil port, which means that cylinder can be driven hydraulically only in one direction like shown in figure 26. Single-acting cylinders are typically applicable when high extend force is needed (e.g. lifting heavy loads), but the required retract force is minimal (Childs 2019, p. 870; Parr 2011, p. 118). The retraction happens usually with help of external force (load itself will retract the piston) or in some cases with compression spring at the end of the rod (Childs 2019, p. 870; Parr 2011, p. 118). In some applications, the piston rod works also as a piston when piston diameter is equal with piston rod as also shown in figure 26 (Galal Rabie 2009, p. 259). This type of cylinder is called as plunger- or ram-type cylinder and it is typically practical with high loads and long strokes (e.g. elevators, jacks). (Doddannavar & Barnard 2005, p. 83)

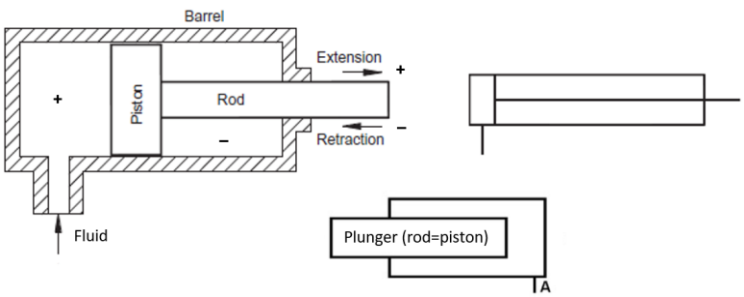


Figure 26. Single-acting (1-way) hydraulic cylinder (modified) (Childs 2019, p. 871; Galal Rabie 2009, p. 259).

Double-acting cylinders (a.k.a. differential cylinders) have two oil ports which enables both extension and retraction with help of hydraulic power. Double-acting cylinders are the most widely used cylinders in hydraulic applications. Due to the different effective areas on piston (+) and rod side (-), more fluid is needed to fill the piston side, which means that extension (+ motion) is slightly slower compared to retraction. On the other hand, by means of larger piston side area, more force may be generated. (Doddannavar & Barnard 2005, p. 85) The most common designs are single-rod, twin-rod symmetrical and twin-rod asymmetrical like presented in figure 27. (Galal Rabie 2009, p. 259)

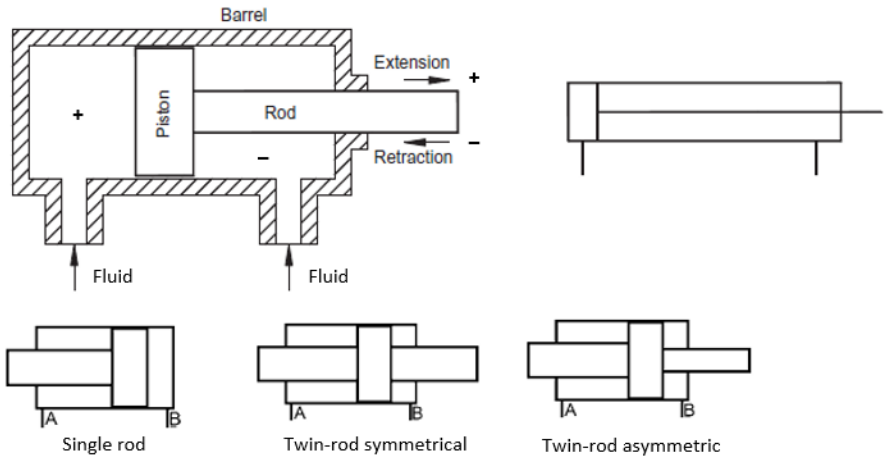


Figure 27. Double-acting (2-way) hydraulic cylinder (modified) (Childs 2019, p. 871; Galal Rabie 2009, p. 259).

3.2 Construction

Construction of hydraulic cylinders is typically categorized into tie-rod and mill-type based on the assembly method of the cylinder. In a tie-rod cylinder, both cylinder head and –cap are assembled together with tie-rods, which ensure the robust structure of cylinder together with seals located typically at the contact surfaces of cylinder tube, -head and- cap as introduced in figure 28. (Galal Rabie 2009, p. 253)

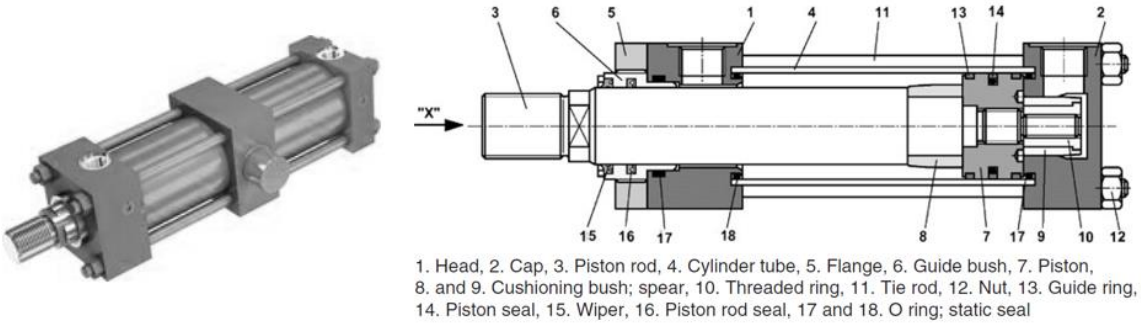


Figure 28. Tie-rod cylinder (modified) (Bosch Rexroth; Galal Rabie 2009, p. 253).

Mill-type cylinders may be assembled in various ways. However, cylinder cap (located at piston side) is usually welded to cylinder tube (in some cases it may be screwed as well). For cylinder head assembly, there are several different options available:

- Cylinder head is bolted to an additional flange, which is connected to cylinder tube by screwing, option a) in figure 29. This cylinder is also equipped with cushioning equipment (on both piston- and rod side). The purpose of cushioning is to reduce the kinetic energy of moving parts and help to control the deceleration speed of piston.
- Cylinder head is assembled into tube by screwing (threads in both cylinder tube and-head, option b) in figure 29).

  (Galal Rabie 2009, p. 253-255; Parr 2011, p. 120)

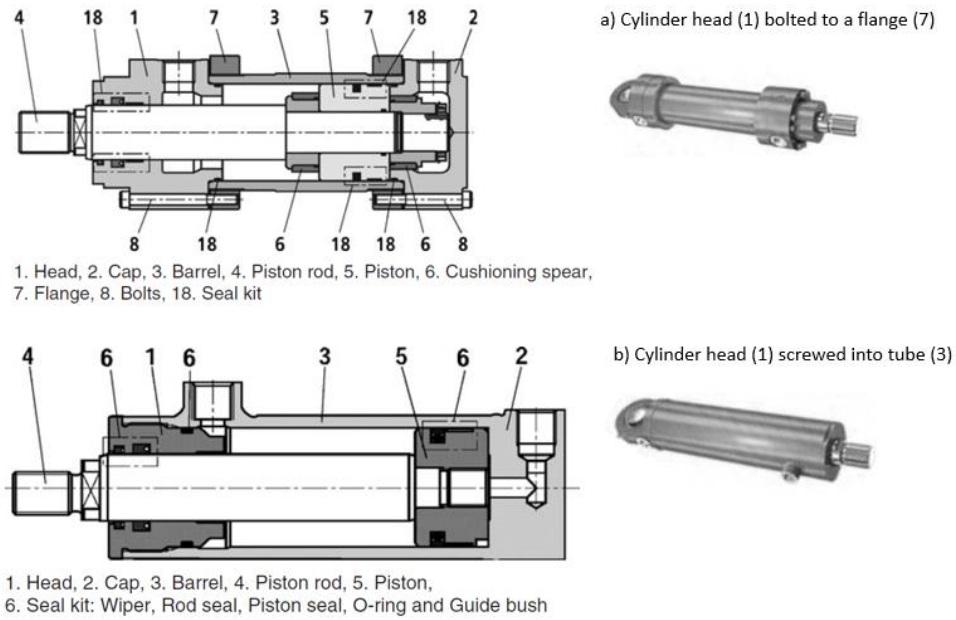


Figure 29. Mill-type cylinder with different assembly options for cylinder head (modified) (Bosch Rexroth; Galal Rabie 2009, p. 253).

- Cylinder head may also be assembled with help of wire- and locking rings or set screws like shown in figure 30 below. In a wire ring method, there is no need for additional thread in cylinder tube or –head. The assembly of the cylinder head is carried out by utilizing inner and external rings, which are placed into machined grooves. (Degelman 2016)

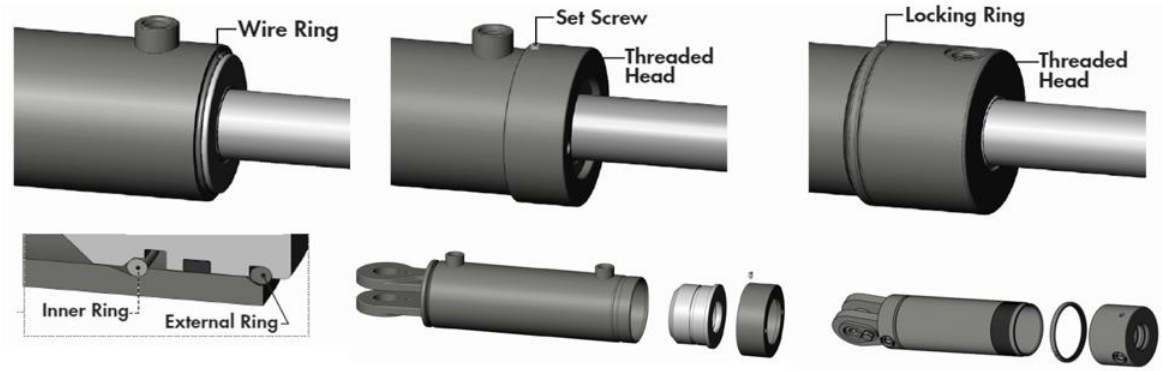


Figure 30. Comparison of wire ring-, set screw- and locking ring style hydraulic cylinders (modified) (Degelman 2016).

3.3 Main components

As highlighted in figure 31 below, there are five main components in a hydraulic cylinder; cylinder tube (barrel), rod, piston and two end caps (called as “top nut” and "tube end” in figure 31). In addition to these main components, also other components like seals, oil ports and bearings play a key role in order to ensure the proper functionality and resistance against internal- and external leakage (Galal Rabie 2009, p. 252).

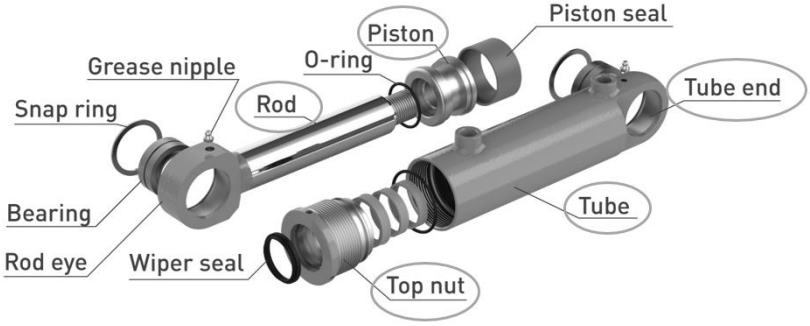


Figure 31. Main components of hydraulic cylinder (modified) (Hydroline 2019).

Generally used material for cylinder tube is drawn steel tube, which is machined or honed from inside due to accuracy requirements. (Parr 2011, p. 121) One generally used raw material grade for cylinder tubes is E355+SR (skived and roller burnished) (Uranie 2018). The surface quality of the inner surface must be smooth in order to avoid wear and leakage. With applications where the cylinder is exposed to corrosive conditions or materials, for example stainless steel, aluminum or brass may be used as material. (Parr 2011, p. 121) Nowadays, it is quite common that manufacturers utilize pre-cut and machined components. (Ovako 2016)

Rods are usually manufactured from heat-treated chromium alloy steels in order to increase the strength properties and corrosion resistance (Parr 2011, p. 122). Example steel grades for rods are C35, C45, 20MnV6, 38MnV6, 42CrMo4 (Uranie 2018; Ovako 2016). Especially, when the rod is extended and extruded through cylinder head, it will be exposed to the surrounding conditions like dirt, moisture and corrosion. In cases where surface quality and corrosion resistance requirements are high, the rod may be chromium plated and polished. (Galal Rabie 2009, p. 252; Parr 2011, p. 122) Usually, the thickness of chromium plating is around 0,25 µm. (Watton 2009, p. 17) Based on Ovako product catalogue, the main driver for costs related to piston rod is the diameter of the bar. In addition to increased usage of raw material, also other manufacturing operations like cutting, machining, chrome plating become more expensive. The general rule of thumb is that in case the rod diameter is reduced by 5mm, total costs will decrease by 15 %. (Ovako 2016)

Pistons are usually made of steel or cast-iron (Doddannavar & Barnard 2005, p. 86). The main function of piston is to transfer the force to the piston rod. Due to its location and function, it also works as a sliding bearing in the cylinder tube. It may also face significant forces since it is in a connection with the rod, which may be subjected to a lateral force. In addition, the hydraulic pressure varies a lot around the piston (piston side vs. rod side), which means that the piston must act as a seal inside the cylinder tube to ensure the internal tightness. (Galal Rabie 2009, p. 252; Parr 2011, p. 122) For that reason, piston seals are typically used between the tube and piston. Sometimes when a small leakage is allowed piston may be used without seals. Piston is typically attached to rod with help of thread or - screwed nut at the end of the rod like shown in figure 32. (SKF; Parr 2011, p. 122)

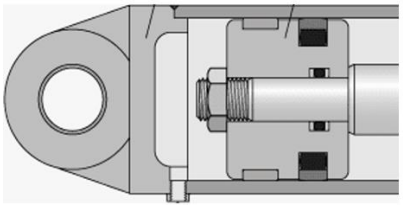


Figure 32. Piston assembly to rod with screwed nut (SKF).

Like the piston and rod, also the cylinder cap and –head must be able to withstand the loads caused by kinetic energy of moving parts and fluid pressure. The cylinder head also prevents external leakage with the help of seals installed on it. The cylinder cap and -head are generally made of cast iron or aluminum. Depending on construction of the cylinder, oil ports may be integrated either into these parts (drilled hole and thread) or into cylinder tube. (Galal Rabie 2009, p. 252; Parr 2011, p. 121)

Depending on the application, there are multiple different ways for mounting of hydraulic cylinders. The mounting style has a remarkable influence on the strength properties of cylinders. (Chapple 2015, p. 30) Mounting methods are divided into eye- or clevis mounting, trunnion mounting, flange mounting and foot mounting as shown in figure 33. (Galal Rabie 2009, p. 258) In general, mounting points on the centerline of the cylinder offer the straight line force transfer and increased lifetime (Heney 2012).

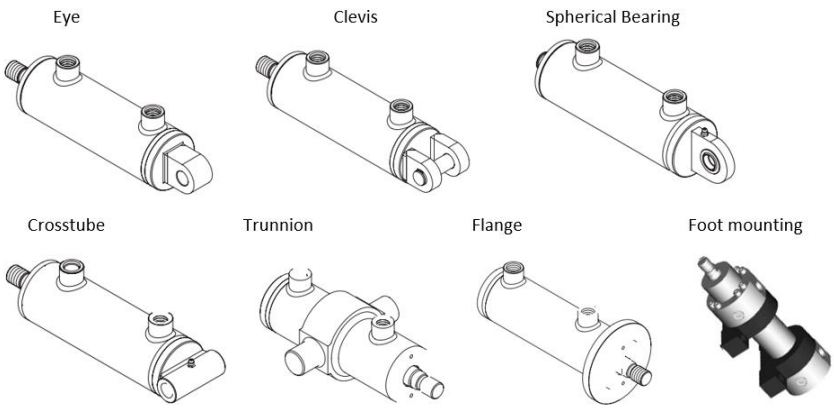


Figure 33. Different mounting options for hydraulic cylinders (modified) (Parker; Galal Rabie 2009, p. 263).

Based on the cylinder properties as part of this study, the eye- or clevis type mounting options are the most used. Typically, especially the end-eye on piston rod side is attached to

the rod by welding. A modern welding process used for this is friction welding technology (Liebherr). Depending on the application, a spherical bearing or end-eye may be assembled into the cylinder end-eye. Figure 34 reveals possible moving directions and pivoting angles of these mounting options. (Galal Rabie 2009, p. 258)

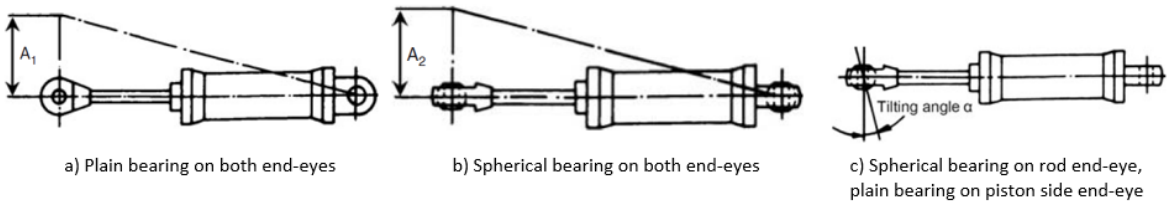


Figure 34. Eye- or clevis type mounting (Galal Rabie 2009, p. 262).

Seals are used to prevent leakage which may cause safety risks, environmental damage, loss of efficiency, increased power usage and temperature rise. Seals are typically classified based on the function as static- (stationary parts) or dynamic (moving parts) seals. It is challenging to manufacture mechanical components like pistons and cylinders with such tight tolerances, which would enable their successful usage without seals. On the other hand, even though required tolerances were achieved, resultant friction would be too high. Thus, selection of the right seals is an important part of the design process. Typically seals are located between the piston and tube, rod and cylinder head (wiper for rod & other seals) and also between cylinder tube- and head. The material of seals is selected based on application, and the selection should be determined according to hydraulic fluid used, working pressure and temperature range. Nowadays, plastic and synthetic rubbers have replaced the earliest materials like leather and cork. Table 4 presents the most common synthetic seal materials. Natural rubber is typically not used in hydraulic systems due to its poor resistance against oil. (Parr 2011, p. 128-130)

Table 4. Comparison of synthetic seal materials (Parr 2011, p. 130).

Material	**Temperature range (°C)**	**Price**	**Other**
Nitrile (Buna-N)	From -50 °C to 100 °C	Cheap	The most common
Silicon	From -100 °C to +250 °C	Expensive	Good temperature resistance, tends to tear
Neoprene	Below 65 °C	Not known	Replaced widely by Nitrile due to poor temperature resistance

3.4 Cylinder forces

Actual forces are established, when the pressure of hydraulic fluid is converted into the force acting on the piston like shown in the figure 35, in which A_p=piston side area [m^2], A_r=Rod-side area [m^2], F=Force [N], P=Pressure [Pa], Q=Flow rate [m^3/s] and v=Piston speed [m/s] (Galal Rabie 2009, p. 251-252).

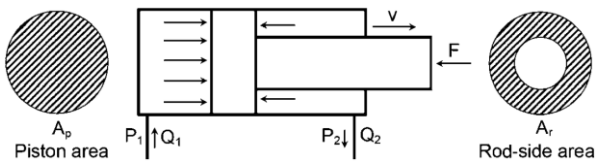


Figure 35. Functionality of double-acting cylinder (Galal Rabie 2009, p. 252).

In case pressure (P_1) is applied to oil port on the piston side, the piston will extend. If pressure (P_2) is applied to piston rod side, the piston will retract. The cylinder force and piston speed are calculated as follows (Galal Rabie 2009, p. 251):

$$F = P_1 A_{p-} P_2 A_r \tag{5}$$

$$v = \frac{Q_1}{A_p} = \frac{Q_2}{A_r} \tag{6}$$

These basic equations (5 and 6) describe the relationship between different factors in hydraulic cylinder design. In this case, the hydraulic cylinder is considered to be leakage-free. In more detailed studies, all possible losses like friction forces and internal leakages should also be taken into account. (Galal Rabie 2009, p. 251) It can be noted that flow rate does not have an impact on cylinder force. Cylinder force is calculated purely based on pressure and piston- and piston rod side areas. In practice, in case the piston area is doubled, the maximum force will be doubled as well, but speed of piston decreases by half. (Parr, p. 120) The maximum achievable forces of hydraulic cylinder are calculated with equations shown below (Parr 2011, p. 117-118):

$$F_{piston\ side} = P_1 A_p \tag{7}$$

$$F_{piston\ rod\ side} = P_2 A_r \tag{8}$$

Equation 7 reveals the maximum force on the piston side during extension (pushing) and equation 8 the maximum force on the piston rod side during retraction (pulling). Due to cylinder structure and existence of piston rod, the maximum retract force is always lower compared to extend force. Typically, the A_p/A_r ratio is roughly 6/5. (Parr 2011, p. 117-118)

3.5 Dimensioning

In addition to cylinder forces, the size of main components and operational length of the hydraulic cylinder needs to be specified based on each application. For piston and piston rod dimensioning, commonly produced dimensions and combinations are introduced in table 5. (Galal Rabie 2009, p. 263-264)

Table 5. Commonly used dimensions for piston and piston rod (modified) (Galal Rabie 2009, p.263-264).

D Piston (mm)	d Rod (mm)	D Piston (mm)	d Rod (mm)
25	12	80	36
	14		45
	16		56
32	18	100	45
	22		50
	25		56
			70
40	16	125	50
	18		56
	22		63
	25		70
			90
50	22	150	63
	25		70
	28		80
	36		100
63	25	160	90
	28		
	36		
	45		

The stroke of the simple hydraulic cylinder must be less than the tube length, and typically the desired relation between extended- and retracted lengths is 2:1. In case space is restricted around the cylinder, telescopic cylinders can be utilized. (Parr 2011, p.123) Figure 36 presents the basic dimensions of double-acting cylinder.

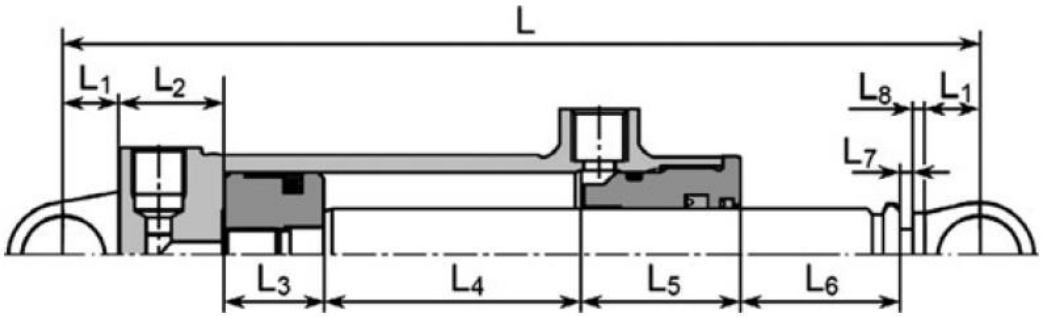


Figure 36. Basic dimensions for stroke and collapsed length (Galal Rabie 2009, p. 258).

In addition to stroke and minimum length (also known as collapsed- or retracted length) of the cylinder, it also shows the wasted lengths related to structure of the cylinder. The collapsed length L is composed of following factors (Galal Rabie 2009, p. 258):

$$L = L_4 + adjustment\ allowance + wasted\ lengths \tag{9}$$

$$L = 2L_1 + L_2 + L_3 + L_4 + L_5 + L_6 + L_7 + L_8 \tag{10}$$

In equations 9 and 10 above, L_1 = Radius of end-eyes [m], L_2 = Cylinder cap (bottom part) [m], L_3 = Piston length [m], L_4 = Stroke [m], L_5= Cylinder head length [m], L_6 = Piston rod length (+thickness of possible locking mechanism) [m], L_7 = Possible allowance of length adjustment [m] and L_8 =Length of end-eyes [m]. (Galal Rabie 2009, p. 258)

4 RESEARCH METHODS

This chapter introduces the research methods utilized in this study. As presented in figure 37, the methods are separated into five different main areas, which are sampling, data collection, interviews, numerical and statistical analysis and indicators. During this chapter, all these methods are described on more detailed level.

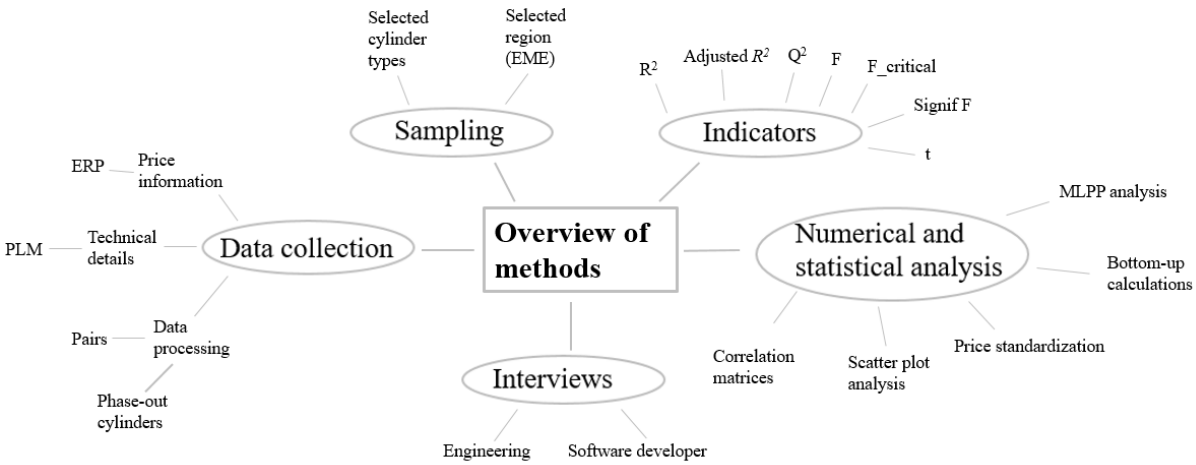


Figure 37. Overview of methods used during this research.

4.1 Sampling, data collection and processing

In the organizational framework, sampling is based on product groups and the selected scope consists of hydraulic cylinders used in EME region tractors and delivered by European suppliers. The regional delimitation is done to minimize the fluctuations caused by currencies and economics. From the cylinder type point of view, four different types of cylinders are included. Due to technical dissimilarities, one cylinder type used in tractors is excluded. The same applies to cylinders equipped with sensor technology, which are removed from data due to significant cost caused by the sensors. Cylinders delivered by sub-suppliers as a part of purchased assemblies are mainly not part of this study due to unknown pricing- and technical data. The pricing data collection is carried out by using the ERP system (SAP) reports. For individual NPI cylinders, the pricing data is collected during supplier selection and business awarding process. Technical drawings and 3D models are downloaded from Product Lifecycle Management (PLM) software PTC Windchill. Additionally, in case of missing technical information, the data was requested directly from

cylinder suppliers. The pricing data is based on selected time period (quarter). During data collection phase, all possible technical characteristics of cylinders were collected to ensure comprehensive comparison of different value drivers during model creation.

In terms of data processing, possible pairs (left/right) are eliminated from data by including only one side as a part of the data. ERP data included also some phased-out items which were still as a part of the scheduling agreements. In this case, the exact phased-out quarter and year are clarified by using MB51 (Material Document List) transaction of SAP due to possible further processing of the data. With a few cylinders, the phase-out information was not available and thus those parts were directly excluded. Respectively, the overlaps among active and phase-out products with similar technical properties are eliminated during the data collection. In total, data was collected for 103 cylinders. After data processing, the sample size is 80 cylinders, of which 56 are active, 6 NPI and 18 phased-out. The more detailed overview of collected data in terms of cylinder types, status of products and suppliers is presented in table 6.

Table 6. Overview of collected and processed data.

	Supplier 1			Supplier 2			Supplier 3		Supplier 4		Supplier 5	Supplier 6			
Type	**Active**	**NPI**	**Phased-out**	**Active**	**NPI**	**Phased-out**	**Active**	**NPI**	**Active**	**NPI**	**Active**	**Active**	**NPI**	**Phased-out**	**Total**
Type 1	2	0	0	10	2	0	2	0	0	0	0	0	0	0	**16**
Type 2	0	0	0	11	0	4	0	0	0	0	0	0	0	0	**15**
Type 3	4	1	2	9	0	6	2	2	1	0	0	3	0	3	**33**
Type 4	2	0	0	6	0	3	0	0	0	0	1	3	1	0	**16**
Total	**8**	**1**	**2**	**36**	**2**	**13**	**4**	**2**	**1**	**0**	**1**	**6**	**1**	**3**	**80**

The figure 38 shows the price distribution in relation to quantity of products for collected data. Based on the price distribution analysis, most of the collected data (>75 %) falls in the price range of 60-140 €.

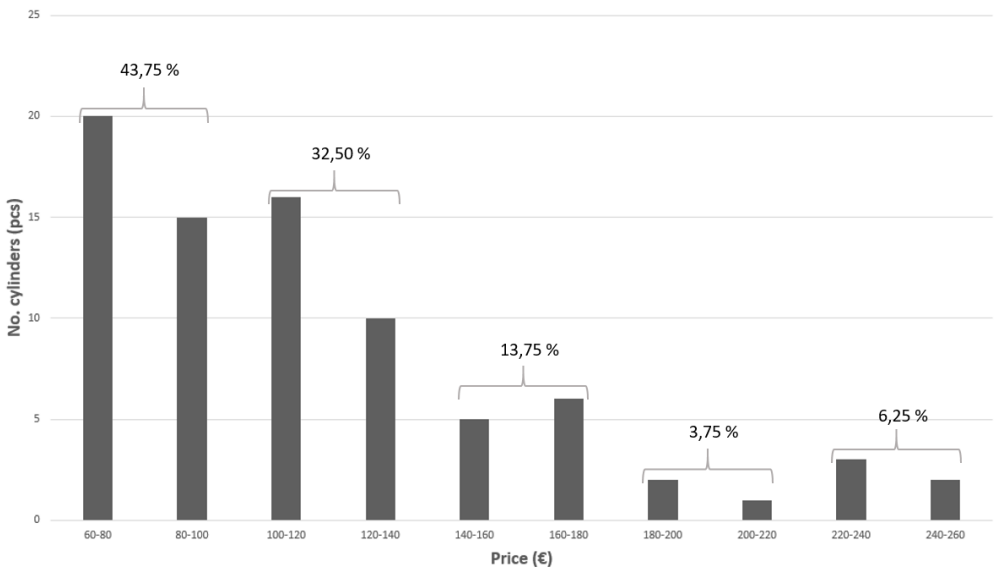


Figure 38. Distribution analysis for purchasing prices and number of products.

4.2 Numerical and statistical analysis

MLPP analysis is the main analysis method utilized in this research. In terms of mathematics, MLPP utilizes PLS regression analysis in order to solve the following matrix equation (Processbench 2021):

$$[Y] = [X] * [B] + [y_0] + [D] \tag{11}$$

This matrix equation 11 includes following elements:

Y = Price vector of each data point.

X = Matrix for data points (1 cell for each data point & value driver combination)

B = Regression coefficient matrix

y_0= Common Y-axis intersection

D = Error or unexplained variation in the dependent variable

The software solves the equation 10 by finding B and y_0. At the same time, D should be minimized. The two main steps of solving the equation are introduced below:

1) X and Y are split and projected into another latent space. In latent space, the number of dimensions is the number of latent variables.
2) Regression coefficient matrix B will be determined.

(Processbench 2021)

The most of the numerical research work is made during MLPP model development, when different combinations of value drivers and technical values are tested and compared. After validation and evaluation of the models, the best models will be selected. Validation is done for both regression function as well as for individual regression coefficients. The evaluation consists mostly of scatter plot analysis, bottom-up calculations and assessment of the model's usability. The figure 39 describes the overall approach used during model development.

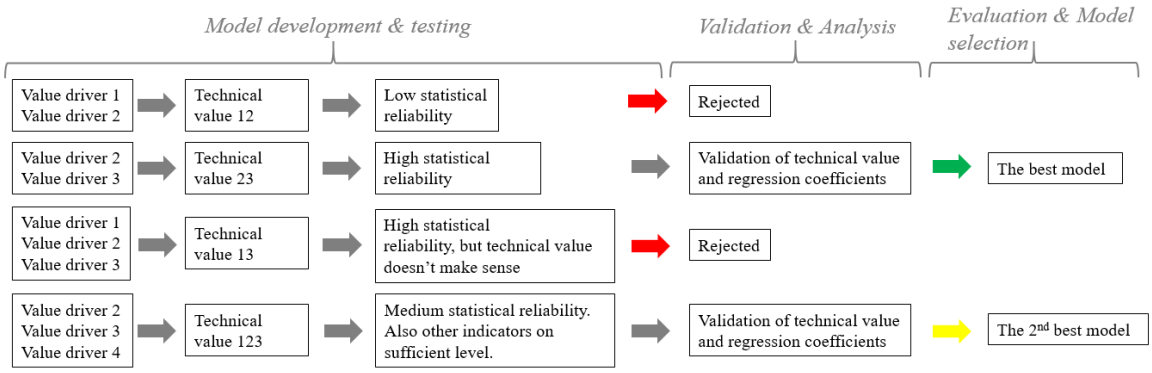


Figure 39. Process chart for MLPP model creation.

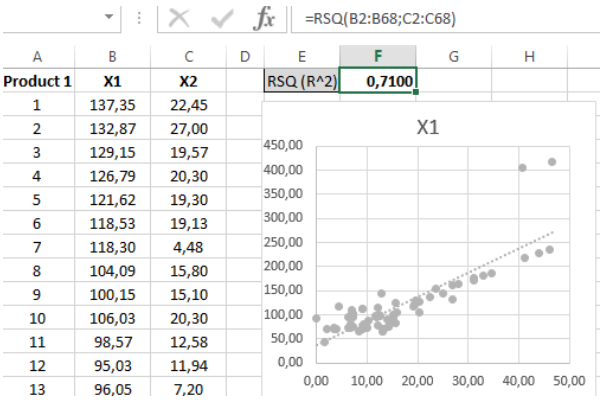
In order to increase the number of data points in the model, the phased-out products are utilized as a part of the models during model testing phase. Due to changes in economics and raw material prices over the years, the slightly modified version of price standardization equation 1 introduced during literature research is utilized to adjust the raw material cost to match the price level of selected quarter. The equation is shown below (VDI 2018, p. 15):

$$P_s = \frac{RM_s * RM_{pc}}{\frac{RM_{pry}}{P_0}} + (1 - RM_s) * P_o \qquad (12)$$

In equation 12 above, P_s = Price of the product (standardized) [€], RM_s = Raw material share [%], P_o = Price of the product (old) [€], RM_{pc} = Raw material price (current) [€/kg or €/ton] and RM_{pry} = Raw material price (reference year) [€/kg or €/ton].

For identification of value drivers, the existing cost information from the market as well as the cost structure data of bottom-up calculations are utilized by identifying the most significant cost drivers. During the multiple linear regression model building, validation and

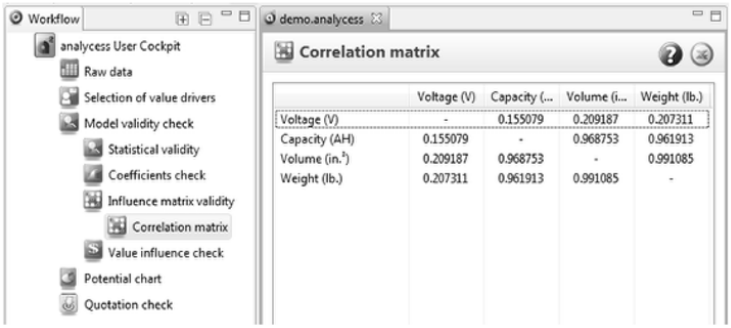
evaluation, the main steps introduced by VDI 2018 are followed. In terms of numerical methods, the correlation of possible value drivers (variables) is studied with help of correlation matrices. At the same time, the relationship between price and selected value drivers may be studied. For visualization, scatter plots are created to make further investigations for the most potential value drivers as shown in figure 40.



Product 1	X1	X2
1	137,35	22,45
2	132,87	27,00
3	129,15	19,57
4	126,79	20,30
5	121,62	19,30
6	118,53	19,13
7	118,30	4,48
8	104,09	15,80
9	100,15	15,10
10	106,03	20,30
11	98,57	12,58
12	95,03	11,94
13	96,05	7,20

Figure 40. RSQ function and scatter plot.

The correlation between value drivers can also be checked in MLPP software by using the "Correlation matrix" feature as shown in figure 41.



	Voltage (V)	Capacity (...	Volume (i...	Weight (lb.)
Voltage (V)	-	0.155079	0.209187	0.207311
Capacity (AH)	0.155079	-	0.968753	0.961913
Volume (in.³)	0.209187	0.968753	-	0.991085
Weight (lb.)	0.207311	0.961913	0.991085	-

Figure 41. Capture of correlation matrix in MLPP software.

In the model evaluation phase, bottom-up calculations are performed for selected data points in order to evaluate the validity of the model within the whole range of technical value. Bottom-up calculations are based the existing expertise and knowledge about cylinder manufacturing as well as suppliers is utilized. In order to increase the validity of bottom-up calculations, most of the chosen bottom-up calculations have already been used during price negotiations with current suppliers or during new supplier selection process. Thus, the input

data used in calculations is adjusted to correspond manufacturing conditions of suppliers. The input data used in bottom-up calculations was adjusted as accurately as possible to match selected quarter's economics conditions as introduced in figure 42.

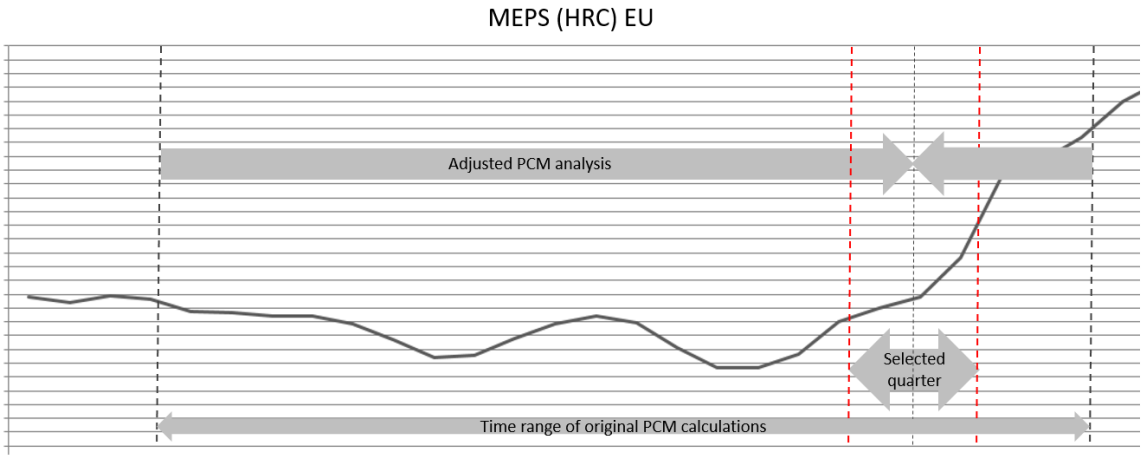


Figure 42. Adjustment of existing bottom-up calculations in relation to MEPS HRC index (modified) (MEPS 2021).

The bottom-up analysis are based on following details:

- The date of calculation adjusted in the middle of selected quarter
- Raw material prices according to SCA material database and existing knowledge
- The most common raw materials used in analysis:
    - Tube raw material (E355 + Skived and roller burnished)
    - Rod raw material (20MnV6 + Induction hardened and chromed, 42CrMo4
    - Cylinder cap and -head, end-eye, piston raw material (S355, C45)
- Production site is based on supplier location
- Machinery and hourly rates are based on existing knowledge and Siemens Product Cost Management (PCM) software database
- Manufacturing strategy (in-house / external) based on existing knowledge. In most cases, the main components are produced in-house and smaller components (e.g. oil ports, seal kits, fixtures) are considered to be bought from external suppliers
- Production volumes according to budgeted volumes taken from ERP system

4.3 Indicators used during model development

The statistical indicators are utilized during whole model creation process. In principle, adjusted R^2 and Q^2 offer good overall guidance during iterative model testing by measuring dependencies of data points and stability. *F*, *signif F* and *F_critical* are used as supportive indicators. When evaluating the necessity of individual value driver, T-test can be utilized. In practice, *t*-value reveals the relationship between regression coefficient and standard error. The most important indicators are presented in table 7. (Processbench 2021; VDI 2018, p. 23)

Table 7. Indicators to be followed during model development (Processbench 2021; VDI 2018, p. 23; Welc & Rodriquez 2018, p. 169).

Indicator	**Description**	**Explanation**	**Criteria**
R^2	Coefficient of determination	Describes the dependency of data points (technical value and price)	In case all data points are located on average line --> R^2 = 1 (very good linear fit).
Adjusted R^2	Modification of coefficient of determination	Takes into consideration both data points and also value drivers	Good model should have > 0.8
Q^2	Stone Geisser Criterion	Measures the model's stability (responsiveness in case individual data points are removed)	Good model should have > 0.8
F	Statistical predictability	*F*=explained variance / unexplained variance	The greater *F* value, the better model (statistically)
signif F	Probability value related to *F* (significance level for *F*)	Indicates whether the model has statistically significant predictive capability	This should be close to 0. In practice 0.05 value for *signif F* means 95 % confidence level
F_critical	*F* value indicator for statistical deviations	States the *F* value at which the probability of statistical deviations is 0,3 % with selected value drivers and data	In practice, in case F_critical < F, null hypothesis can be rejected
t	The result of t-test (relationship between regression coefficient and standard error)	Describes the usefulness of individual value driver	The higher t, the more reliable the regression is about the respective value driver

Another useful indicator to be used during model development is R^2 Sensitivity [%], which allows to evaluate the influence of individual data point in the model. In case the value R^2 Sensitivity [%] is negative, deselection of this particular data point will probably improve the model in terms of Adjusted R^2. For evaluation of technical value, Std. coefficient and

Regression significance [%] are applicable indicators. Std. coefficient tells which value driver has the most significant impact on price, whereas Regression significance [%] reveals the significance of each value driver in relation to regression result.

4.4 Interviews with software developer and engineering

As a basis for quantitative research, also qualitative methods are utilized in the form of interviews. The interviewees are from MLPP software developer Processbench and from AGCO engineering departments. The form of interviews is an open interview and the purpose of the interviews is summarized in figure 43.

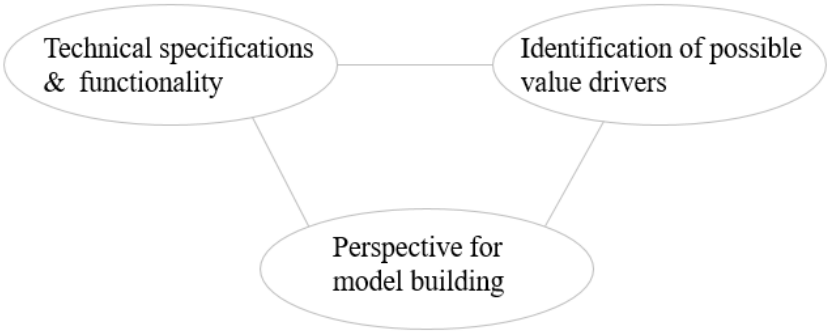


Figure 43. The purpose of interviews.

5 MLPP MODEL FOR HYDRAULIC CYLINDERS

This chapter discusses the MLPP model development for hydraulic cylinders. As a result of model development, two models based on different combination of value drivers are selected for further utilization in the organization.

5.1 Basis for model building

The basis of approach for the model creation was built by interviewing respective stakeholders, which aligns with VDI 2018 proposed strategy of cross-functional co-operation before model creation. Due to relatively complex subject and possible challenges with quantification of the information, the nature of these interviews was an open interview. Another reason for open interviews was that an open interview would more probably bring up new ideas and thoughts around the topic, which could offer valuable perspectives for value driver selection and model building. Interviews were conducted and recorded in Microsoft Teams.

At first, Dr. Berz from Processbench (MLPP software developer) was interviewed in order to get perspective for model creation. Secondly, respective design engineers from R&D departments of two different AGCO sites were interviewed. The purpose of engineering interviews was to get more comprehensive understanding of technical specifications and functionality as well as discuss about potential value drivers to be used. Additionally, more comprehensive technical knowledge would support especially the usage of possible categorical value drivers. In total, interviews were made with 5 design engineers, who are responsible for front- and rear end installations, suspension and hydraulics. Despite the open nature of the interviews, based on the discussion with Dr. Berz, the approach for engineering interviews was to identify possible performance related parameters in hydraulic cylinders instead of technical characteristics or cost drivers only. The main areas of discussion on the technical side were cylinder force, hydraulic pressure, dimensioning, specifications and design comparisons among cylinder types. In the beginning of the interviews with engineering, the operating principle of multiple linear performance pricing was introduced to interviewees. The summarized results of interviews are shown in appendix 1.

5.2 Selection of value drivers

In addition to interviews, existing technical knowledge, bottom-up calculation data and the cost data received from suppliers through cost breakdowns (CBD) was utilized during value driver identification. The utilization of cost data was carried out by collecting the data of 8 cylinders from 3 different suppliers. From cylinder type point of view, 3 of these were type 3-, 2 type 1- and 3 type 4 cylinders. The data was converted into standard format, in which all main cost factors of hydraulic are separated. The data consist purely of production cost figures. Overheads and profits are not considered, because those are not relevant in terms of value driver selection. The principle of cost information processing is introduced in table 8.

Table 8. The utilization principle of cost data during cost and value driver identification.

Hydraulic cylinder	1	2	3	4	5	6	7	8	AVG
Tube	9 %	9 %	11 %	15 %	27 %	26 %	20 %	21 %	17 %
Rod	13 %	11 %	23 %	29 %	11 %	10 %	13 %	10 %	15 %
Machined comp.	29 %	33 %	21 %	25 %	31 %	32 %	34 %	37 %	30 %
Seal kit	9 %	9 %	11 %	8 %	12 %	11 %	16 %	13 %	11 %
Other comp.	14 %	17 %	15 %	6 %	0 %	0 %	0 %	0 %	6 %
Welding	6 %	6 %	6 %	5 %	7 %	7 %	6 %	6 %	6 %
Testing/assembly/washing	15 %	11 %	9 %	7 %	9 %	9 %	8 %	8 %	10 %
Painting	5 %	4 %	4 %	5 %	4 %	4 %	3 %	4 %	4 %

For clarification, tube, rod, machined components includes raw material and also manufacturing costs related to these components. Machined components consist of piston, end-eyes and cylinder head and -cap. Other components includes standard components like bearings, oil ports, retaining rings and grease nipples. The processed cost information is presented in figure 44.

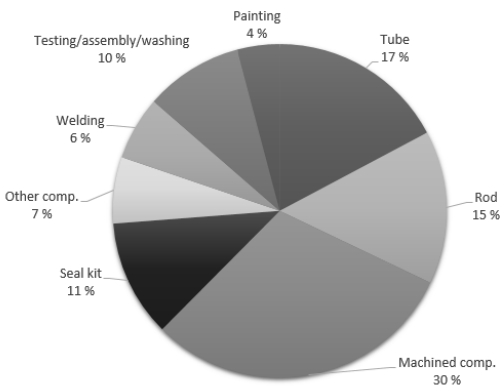


Figure 44. Average %-shares of different cost factors based on supplier data in table 8.

One important perspective for value driver identification was to utilize such technical parameters and characteristics, which are determined by our own engineering during NPI phase. This improves the usability of MLPP model also with new products, taking into account the fact that the design process is mainly done by an external supplier. Based on the interviews with engineering, only the mandatory dimensions and specifications are determined during new product development. Typical specifications are related to cylinder tube and (sometimes) rod diameter (required forces), collapsed length, stroke, pressure, type of mounting, bearings and grease nipples. Thus, for example the size of the piston or exact length of the piston rod are not included in these identified value drivers, even though those would most likely be suitable candidates for value drivers. Additionally, in some cases the dimensions like the exact length of piston, cylinder cap or –head or detailed information of seals are not specified in supplier drawings. It is also quite common, that 3D files do not include sub-part structure and hence this would significantly increase and complicate the measurement effort in data collection phase. However, although it is difficult to quantify such components in cylinder, those can be expected either to increase or decrease in proportion to cylinder tube diameter. In terms of performance pricing, the cylinder force would be one good option to be used as value driver. However, in this case it does not make sense since the working pressure is typically around 200 bar and thus cylinder force would be an unnecessary derivate of area and pressure. In addition to numerical value drivers, cylinder type is selected to be used as categorical value driver in model testing due to technical differences between cylinders used in different applications. The identified value driver (default values) candidates are listed in table 9 below.

Table 9. Identified possible value drivers.

Value driver	**Abbreviation**	**Type**	**Unit**	**Source**	**Effort needed**
Collapsed length	CL	Numerical	mm	Technical drawing	Low
Inner Diameter	ID	Numerical	mm	Technical drawing	Low
Outer Diameter	OD	Numerical	mm	Technical drawing	Low
Wall thickness	WT	Numerical	mm	Technical drawing	Low
Stroke	S	Numerical	mm	Technical drawing	Low
Rod Diameter	RD	Numerical	mm	Technical drawing	Low
Weight	W	Numerical	kg	Technical drawing / suppliers	Medium
No. oil ports	NOB	Numerical	pcs	Technical drawing	Low

No. bearings/bushings	NOB	Numerical	pcs	Technical drawing	Low
No. grease nipples	NOG	Numerical	pcs	Technical drawing	Low
Cylinder type	CT	Categorical	*	Technical drawing	Low

Based on value drivers listed in table 9, it is possible to extend the list of value drivers by creating other derivatives like areas (mm^2) and volumes (mm^3). The basic technical data was mainly available in technical drawings, but some portion of weights are based on information asked directly from suppliers.

The selection of value drivers was started by creating correlation matrices *(R^2)* for numerical value drivers, which describes dependencies among value drivers. The correlation matrices were created for two different samples:

- All cylinders (80 pcs) including active-, NPI- and phased-out parts with standardized prices
- Only active- and NPI cylinders (62 pcs)

At this point, in addition to interdependencies among value drivers, another important fact revealed by the matrix is the correlation between purchasing price and value drivers. The correlation matrices are presented in tables 10 and 11.

Table 10. The correlation matrix based on R^2 (all cylinders 80 pcs).

	P	CL	ID	OD	WT	S	RD	W	NOP	NOB	NOG
P	1,00	0,77	0,04	0,06	0,14	0,60	0,09	0,83	0,00	0,41	0,15
CL	0,77	1,00	0,02	0,02	0,03	0,73	0,00	0,86	0,00	0,34	0,24
ID	0,04	0,02	1,00	0,99	0,11	0,05	0,04	0,18	0,02	0,01	0,00
OD	0,06	0,02	0,99	1,00	0,19	0,06	0,05	0,20	0,02	0,01	0,00
WT	0,14	0,03	0,11	0,19	1,00	0,07	0,08	0,12	0,00	0,00	0,05
S	0,60	0,73	0,05	0,06	0,07	1,00	0,01	0,74	0,00	0,18	0,29
RD	0,09	0,00	0,04	0,05	0,08	0,01	1,00	0,06	0,30	0,04	0,00
W	0,83	0,86	0,18	0,20	0,12	0,74	0,06	1,00	0,01	0,27	0,19
NOP	0,00	0,00	0,02	0,02	0,00	0,00	0,30	0,01	1,00	0,09	0,01
NOB	0,41	0,34	0,01	0,01	0,00	0,18	0,04	0,27	0,09	1,00	0,00
NOG	0,15	0,24	0,00	0,00	0,05	0,29	0,00	0,19	0,01	0,00	1,00

Table 11. The correlation matrix based on R^2 (active and NPI cylinders 62 pcs).

	P	CL	ID	OD	WT	S	RD	W	NOP	NOB	NOG
P	1,00	0,72	0,05	0,07	0,18	0,58	0,16	0,79	0,00	0,34	0,12
CL	0,72	1,00	0,04	0,04	0,05	0,72	0,01	0,83	0,00	0,23	0,18
ID	0,05	0,04	1,00	0,99	0,08	0,06	0,09	0,25	0,02	0,00	0,00
OD	0,07	0,04	0,99	1,00	0,15	0,08	0,11	0,27	0,03	0,00	0,00
WT	0,18	0,05	0,08	0,15	1,00	0,10	0,13	0,17	0,01	0,00	0,10
S	0,58	0,72	0,06	0,08	0,10	1,00	0,02	0,73	0,00	0,12	0,23
RD	0,16	0,01	0,09	0,11	0,13	0,02	1,00	0,13	0,34	0,04	0,00
W	0,79	0,83	0,25	0,27	0,17	0,73	0,13	1,00	0,01	0,17	0,14
NOP	0,00	0,00	0,02	0,03	0,01	0,00	0,34	0,01	1,00	0,11	0,02
NOB	0,34	0,23	0,00	0,00	0,00	0,12	0,04	0,17	0,11	1,00	0,03
NOG	0,12	0,18	0,00	0,00	0,10	0,23	0,00	0,14	0,02	0,03	1,00

As shown in correlation matrices, statistically cylinder weight (0,83 / 0,79), collapsed length (0,77 / 0,72) and stroke (0,60 / 0,58) have the highest correlation in relation to purchasing price. At this stage, diameters (ID and OD) of the cylinder tube as an individual value drivers have really low correlation with purchasing price. Based on the matrix, relationships among value drivers were also checked. For instance, it is obvious that inner and outer diameter should not be used as value drivers in the same model because of high correlation due to geometry. On the other hand, although for example weight and collapsed length seem to correlate with each other, they are not linked to each other in a way which would prevent the utilization of both value drivers in the same model. Finally, in order to compare and check the validity of matrices with two different samples, the differences between these matrices were studied through the change in the degree of explanation of the correlation (%) table 12 presented below.

Table 12. Change in the degree of explanation of the correlation (%) with two different samples.

	P	CL	ID	OD	WT	S	RD	W	NOP	NOB	NOG
P	0 %	5 %	-1 %	-2 %	-4 %	1 %	-7 %	4 %	0 %	8 %	3 %
CL	5 %	0 %	-2 %	-2 %	-2 %	0 %	-1 %	4 %	0 %	11 %	6 %
ID	-1 %	-2 %	0 %	0 %	3 %	-1 %	-5 %	-7 %	-1 %	0 %	0 %
OD	-2 %	-2 %	0 %	0 %	4 %	-2 %	-6 %	-8 %	-1 %	0 %	0 %
WT	-4 %	-2 %	3 %	4 %	0 %	-3 %	-5 %	-4 %	0 %	0 %	-5 %
S	1 %	0 %	-1 %	-2 %	-3 %	0 %	-1 %	1 %	0 %	6 %	6 %
RD	-7 %	-1 %	-5 %	-6 %	-5 %	-1 %	0 %	-7 %	-4 %	0 %	0 %
W	4 %	4 %	-7 %	-8 %	-4 %	1 %	-7 %	0 %	-1 %	10 %	5 %
NOP	0 %	0 %	-1 %	-1 %	0 %	0 %	-4 %	-1 %	0 %	-3 %	0 %
NOB	8 %	11 %	0 %	0 %	0 %	6 %	0 %	10 %	-3 %	0 %	-3 %
NOG	3 %	6 %	0 %	0 %	-5 %	6 %	0 %	5 %	0 %	-3 %	0 %

In order to ensure the comprehensiveness of model testing, some additional derivatives were created, which are listed in the table 13 below. At the same time, in case certain derivative seems potential, it would allow to reduce the number of used value drivers in analysis. Those could also be utilized for handling possible nonlinearities.

Table 13. Potential value drivers (based on derivatives).

Value driver (derivative)	Abbreviation	Type	Unit
Operational length	OL	Numerical	mm
Area (Piston side)	APS	Numerical	mm^2
Area (Rod side)	APR	Numerical	mm^2
Tube end section area	TESA	Numerical	mm^2
Rod end section area	RESA	Numerical	mm^2
Area (Piston side) x Stroke	APSxS	Numerical	mm^3
Area (Piston side) x Collapsed length	APSxCL	Numerical	mm^3
Area (Piston side) x Operational length	APSxOL	Numerical	mm^3
Tube end section area x Stroke	TESAxS	Numerical	mm^3
Tube end section area x Collapsed length	TESAxCL	Numerical	mm^2
Tube end section area x Operational length	TESAxOL	Numerical	mm^3
Rod end section area x Stroke	RESAxS	Numerical	mm^3
Rod end section area x Collapsed length	RESAxCL	Numerical	mm^3
Rod end section area x Operational length	RESAxOL	Numerical	mm^3
Inner diameter x Operational length	IDxOL	Numerical	mm^2
Rod diameter x Operational length	RDxOL	Numerical	mm^2

The piston rod side area was not further utilized in derivation, since especially in case of type 4 cylinders, piston diameter is typically really close to rod diameter. Hence, the multiplication of piston rod side area would probably distort the model, since presumably thicker piston rod should be more expensive. Similarly, the correlation matrices were created for these additional value drivers (see complete matrices in appendix 2) and also change in the degree of explanation of the correlation (%) was checked (appendix 3). However, the price – value driver relationships for derivative value drivers are shown in table 14.

Table 14. Correlations between derivative value drivers and price.

All cylinders (80 pcs)																
	OL	APS	APR	TESA	RESA	APSx	APSxCL	APSxOL	TESAxS	TESAxCL	TESAxOL	RESAxS	RESAxCL	RESAxOL	IDxOL	RDxOL
P	0,76	0,03	0,01	0,11	0,04	0,50	0,69	0,66	0,60	0,80	0,78	0,49	0,70	0,66	0,76	0,79
Active and NPI cylinders (62 pcs)																
	OL	APS	APR	TESA	RESA	APSx	APSxCL	APSxOL	TESAxS	TESAxCL	TESAxOL	RESAxS	RESAxCL	RESAxOL	IDxOL	RDxOL
P	0,72	0,04	0,01	0,15	0,09	0,51	0,61	0,60	0,60	0,75	0,73	0,55	0,69	0,67	0,71	0,79

As a result of value driver matrices, in total 7 individual- or derivative value drivers are having > 0,70 R^2 correlation with purchasing price like presented in table 15. In addition to weight, collapsed and operational length (the sum of collapsed length and stroke) seem to be a strong explanatory factor in pricing. The scatter plots for weight, collapsed length, stroke and operational length can be found in appendix 4.

Table 15. Value drivers with >0,70 R^2 correlation with purchasing price based on all cylinders (80 pcs).

R^2	**Value driver**	**Abbreviation**
0,83	Weight	W
0,80	Tube end section area x Collapsed length	TESAxCL
0,79	Rod diameter x Operational length	RDxOL
0,78	Tube end section area x Operational length	TESAxOL
0,77	Collapsed length	CL
0,76	Operational length	OL
0,76	Inner diameter x Operational length	IDxOL

5.3 Model testing

The extensive investigation of potential value drivers supported the selection of approach for actual model creating. Based on price – value driver relationships, the weight of the cylinder has clear correlation (0,83) with purchasing price. Additionally, the value drivers related to main dimensions like collapsed length (0,77) and operational length (0,76) seems to have pretty strong correlation with price. Hence, the model testing was started by investigating suitable combinations of the most potential value drivers. In addition to numerical value drivers, also categorical value drivers were included as a part of the model testing. The cylinder weight having strong correlation to price, the scatter plot based on weight as single value driver was studied before actual testing as shown in figure 45, where cylinder type has been used as a filter. Based on locations of data points, at least type 4 and type 2 cylinders can be clearly distinguished, which means that the need for cylinder type as categorical value driver is real.

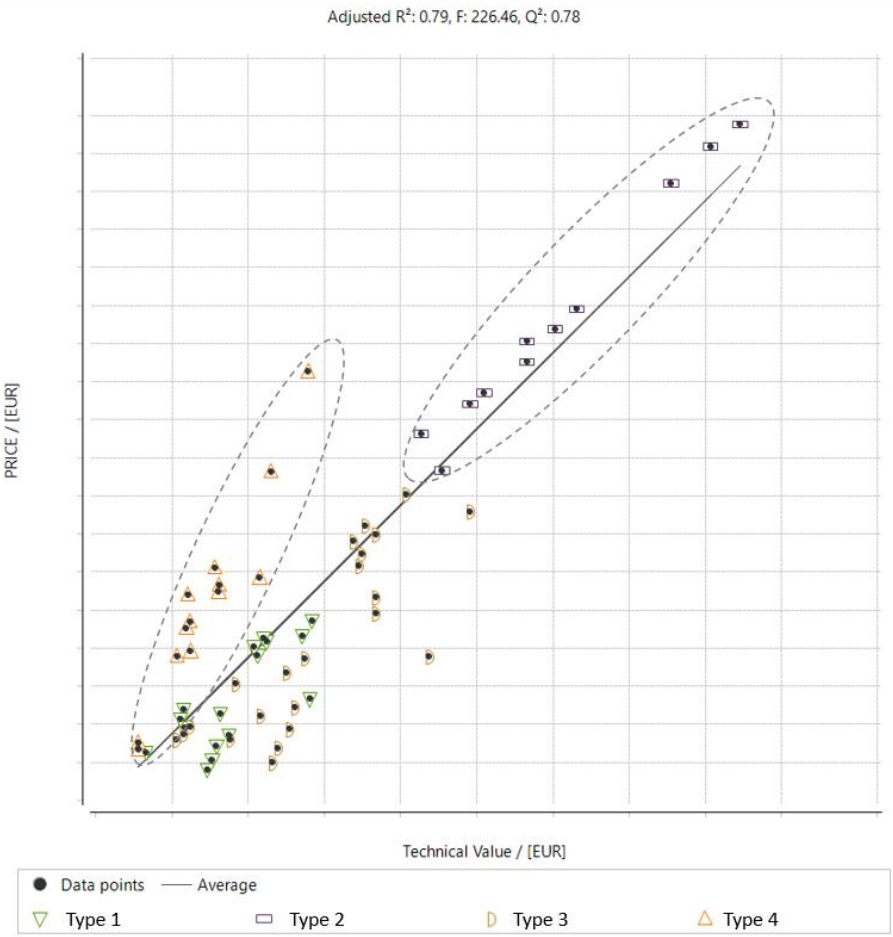


Figure 45. Scatter plot based on weight as single value driver indicates the necessity of categorical value drivers in MLPP analysis (notice the difference especially between types 2 and 4).

Thus, cylinder types were classified with 3 different ways. The first option separates all cylinder types (type 1, 2, 3 and 3), whereas the second option separates type 2 - and type 4 cylinders, but considers both type 1- and type 3 cylinders as one categorical value driver. The third option separates only type 4 cylinders from other cylinders (type 2 cylinders are considered similarly than type 1 and 3).

For model creation, 4 main groups (different combinations of value drivers) were created:

- Cylinder weight and –types
- Cylinder weight and technical characteristics
- Cylinder types and technical characteristics
- Technical characteristics only

Model testing was performed in parallel for two different samples, which were “Active+ NPI + Phased-out parts” (80 data points) and “Active + NPI cylinders” (62 data points). With most of the models, 2 or 3 data points (outliers) were removed from regression due to significant deviation and high R^2 sensitivity [%] with all combinations of value drivers. In general, -10 % value of R^2 Sensitivity [%] was used as threshold for rejection. The removed part numbers were:

- ABC1234628 Type 4 cylinder (NPI)
- ABC1234568 Type 3 cylinder (Active)
- ABC1234580 Type 4 cylinder (Active)

The model testing was performed by using iterative stepwise selection of value drivers and the intention was to find the best combinations in terms of adjusted R^2 and Q^2. In addition to the traditional selection procedure of value drivers, value driver folding was tested with categorical value drivers (cylinder types). In practice, this was done by using matrix-format input for cylinder weights, operational lengths, collapsed lengths and diameters with certain value driver combinations. Dozens of different combinations were tested and 24 combinations were selected for further investigation. The test results of all models can be found from appendices 5-8. In general, as a result of testing, all these combinations indicate relatively high statistical correlation. Especially combinations based on weight and cylinder types have really high statistical reliability. Based on model testing results, 6 models were chosen for further validation and evaluation. At this phase of analysis, the decision was made by using indicators adjusted R^2, Q^2 and F as statistical criteria. Even though statistical indicators give good guidance during model development, perhaps the most important factor in the decision making process was the assessment of whether the technical value of the model makes sense. Secondly, the usability of technical value in practical applications was also an important selection criterion. All nonsensical models were rejected (e.g. technical values in which increasing inner diameter of cylinder tube would decrease the price). To ensure broad perspective, at least one model from each main group was selected. At this point, only models with “Active or NPI cylinders” were chosen for further validation. An overview of selected models for further validation are presented in table 16.

Table 16. Selected models for further validation.

Group	1. Weight + Cylinder types			2. Weight + Technical characteristics	3. Cylinder types + Technical characteristics	4. Technical characteristics only
Number	1.4 (matrix input for weight)	1.8 matrix input for weight)	1.12 (matrix input for weight)	2.2	3.4 (matrix input for OL)	4.4
Data points	62	62	62	62	62	62
Value drivers	Weight + All cylinder types as separate	Weight + (Type 1 and 3) + Type 4 + Type 2	Weight + (Type 1, 2 and 3) + Type 4	Weight + Rod diameter + Operational length	All cylinder types as separate + Inner diameter + Operational length	Operational length + Rod diameter
Number of value drivers	4	3	2	3	5	2
Latent variables	4	3	2	3	5	2
Technical value	48.10 + 3.41 * Type 1 / [kg] + 4.09 * Type 2 / [kg] + 3.27 * Type 3 / [kg] + 7.03 * Type 4 / [kg]	48,63 + 4,07 * Type 2 / [kg] + 3,27 * (Type 1 and 3) / [kg] + 7,04 * Type 4 / [kg]	38.24 + 4.21 * (Type 1, 2 and 3) / [kg] + 8.22 * Type 4 / [kg]	5.06 + 0.04 * Operational length / [mm] + 1.19 * Rod Diameter / [mm] + 1.60 * Weight / [kg]	-24.01 + 0.11 * Type 1 Ol / [mm] + 0.10 * Type 2 OL / [mm] + 0.59 * Inner Diameter / [mm] + 0.10 * Type 3 OL / [mm] + 0.23 * Type 4 OL / [mm]	-8.83 + 0.07 * Operational length / [mm] + 1.56 * Rod Diameter / [mm]
Adjusted R^2	0,95	0,95	0,93	0,89	0,92	0,87
Q^2	0,94	0,94	0,93	0,89	0,91	0,86
F	257,27	356,71	407,81	166,35	136,66	219,18
F critical	4,64	5,32	6,57	5,32	4,20	6,59
Sign F	0	0	0	0	0	0
Parts removed from regression	ABC1234568	ABC1234568	ABC1234568	ABC1234568	ABC1234568	ABC1234568
	ABC1234628	ABC1234628	ABC1234628	ABC1234628	ABC1234628	ABC1234628
	ABC1234580	-	-	ABC1234580	-	ABC1234580

5.4 Validation and evaluation of the model

Before selection of the most suitable model, an extensive validation and evaluation is required to ensure the statistical quality of the model. Moreover, it is necessary to evaluate if the model corresponds to reality. In this study, the validation and evaluation of the model are done in parallel, since simultaneous validation and evaluation of the model allows to assess the model from different perspectives. For example, when model premises are checked based on the scatter plots, it is reasonable to perform the evaluation based on bottom-up calculations at the same time, since bottom-up calculations form an important part of validation process. Another important criteria in model selection is the usability of the model in practical applications (e.g. target costing in NPI phase), and thus it makes sense to do the evaluation of usability at an early stage. The figure 46 presents the main elements of model validation and evaluation process.

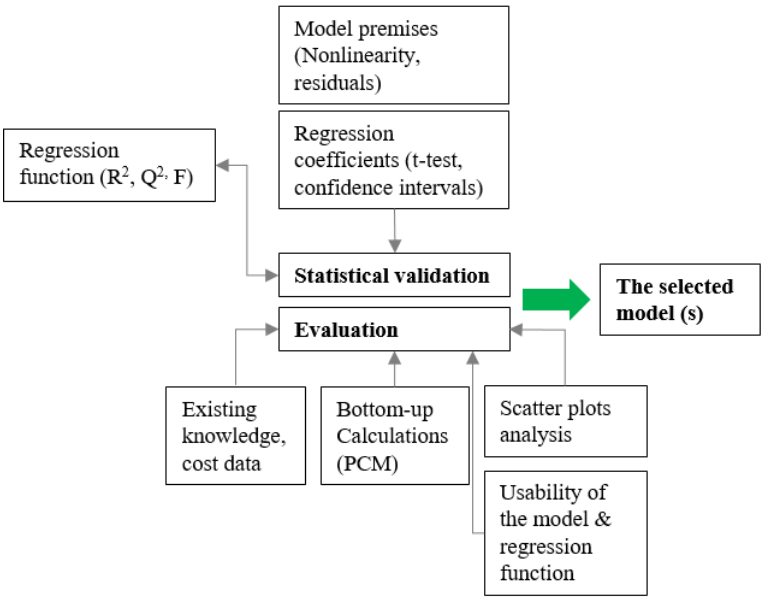


Figure 46. Main elements of validation and evaluation process before model selection.

As introduced in literature research, VDI 2018 proposes three main steps for validation of the regression function (VDI 2018, p. 21-28):

- Checking of the regression function
  - Main criteria R^2, Q^2 and F
- Checking regression coefficients
  - T-test and confidence interval
- Checking model premises
  - Residuals
  - Nonlinearity

In case of this research, checking of the regression function was performed in parallel with model testing phase. In case of overall models, evaluation criteria consisted of R^2, Q^2 and F. Additionally, the regression coefficients were checked as a part of technical value formula in terms reasonability and if the relationships between regression coefficients made sense (e.g. the order of different cylinder types in terms of pricing). Thus, the main actions related to statistical validation of regression coefficients are t-test and checking of confidence intervals. The table 17 presents t-values and maximum- and minimum coefficient values (confidence interval) for each value driver. Additionally, this table includes standardized coefficient- and regression significance values. Standardized coefficient reveals the impact

of individual value driver on purchasing price, whereas the regression significance describes the statistical significance (importance in terms of regression function) of selected value driver. Both of these coefficients are checked as well, to ensure the necessity of each value driver.

Table 17. Validation of individual regression coefficients.

Model	Value driver	Type	Rgression coefficient	Std. coefficient	Regression significance	Max coefficient	Min coefficient	t-value
1.4 (matrix input for weight)	Type 3 (kg)	Numerical	3,27	0,64	100,00 %	3,54	3,00	12,04
	Type 4 (kg)	Numerical	7,03	0,52	100,00 %	7,67	6,40	11,09
	Type 1 (kg)	Numerical	3,41	0,40	100,00 %	3,84	2,98	7,98
	Type 2 (kg)	Numerical	4,09	1,31	100,00 %	4,24	3,94	27,75
1.8 (matrix input for weight)	Type 2 (kg)	Categorical	4,07	1,30	100,00 %	4,21	3,94	30,35
	Type 1 and 3 (kg)	Categorical	3,27	0,61	100,00 %	3,53	3,01	12,59
	Type 4 (kg)	Categorical	7,04	0,57	100,00 %	7,57	6,51	13,33
1.12 (matrix input for weight)	Type 1, 2 and 3 (kg)	Numerical	4,21	1,18	100,00 %	4,36	4,07	28,56
	Type 4 (kg)	Numerical	8,22	0,67	100,00 %	8,73	7,72	16,21
2.2	Operational length (mm)	Numerical	0,04	0,47	99,92 %	0,05	0,03	3,56
	Rod diameter (mm)	Numerical	1,19	0,24	100,00 %	1,44	0,93	4,61
	Weight (kg)	Numerical	1,60	0,39	99,27 %	2,18	1,03	2,79
3.4 (matrix input for OL)	Type 1 (mm)	Numerical	0,11	0,69	100,00 %	0,13	0,08	4,68
	Type 2 (mm)	Numerical	0,10	1,53	100,00 %	0,11	0,09	11,37
	Type 3 (mm)	Numerical	0,10	0,79	99,99 %	0,13	0,08	4,41
	Type 4 (mm)	Numerical	0,23	0,92	100,00 %	0,27	0,20	7,30
	Inner diameter (mm)	Numerical	0,59	0,26	99,99 %	0,73	0,45	4,10
4.4	Operational length (mm)	Numerical	0,07	0,82	100,00 %	0,08	0,07	16,33
	Rod diameter (mm)	Numerical	1,56	0,31	100,00 %	1,79	1,32	6,68

Based on the confidence intervals, all value drivers have the same sign (+) and there is no change between negative and positive values within the coefficient range. In addition to six selected models, the confidence interval check was made for all other test models as well to get more comprehensive understanding of the value driver's behavior. With many models (e.g. 1.1, 1.3, 1.5, 1.7, 1.9, 1.11 and 3.3) including categorical value drivers with normal data input (without value driver folding matrix input format), the sign (+ and -) was heavily changing within the confidence interval, which means that the models are not valid in terms of regression coefficients.

T-value allows to evaluate the importance of each selected value driver for the regression function since it describes the relationship between regression coefficient and standard error (the average of maximum and minimum coefficients). At this stage, for example the

numerical value drivers including both weight and type 2 as a cylinder type (matrix input models 1.4, 1.8 and 3.4) seem to have high t-values, which indicates that the data points located at the upper end of the model have high importance for the regression function.

As a next step of validation, residuals were checked by reviewing the scatter plots. The scatter plots of all 6 models can be found from appendix 9. Based on visual inspection, the distribution of residuals was were ranked from 1-6 (1=the most even, 6= the most uneven) (see table 18). Based on visual evaluation, it was challenging to make a difference between models 1.12 and 3.4 and hence both were ranked to 3rd place. A general observation of scatter plots that applies to all models is that the most of the data points are located at the lower value range of technical value, and the upper end of the model consist mainly of the type 2 cylinders due to relatively different technical characteristics.

Table 18. Assessment table for residuals.

Model	**Distribution of residuals**	**Ranking**
1.4	Even	1
1.8	Even	2
1.12	Uneven with technical values >150	3
2.2	Uneven with technical values >150	5
3.4	Uneven with technical values 110-130	3
4.2	Uneven especially with technical values >100	6

During scatter plot analysis, also possible nonlinearities were checked based on visual check on data point locations. All models based on weight and cylinder type (1.4, 1.8 and 1.12) seem to have a linear fit of regression through the whole value range and the formula of technical value does not refer to nonlinearity. In terms of scatter plots, the higher end of technical value range is pretty similar with models 2.2 and 4.4. In both cases, if the data points with technical value >150 are selected, the slope based on these points seem to differ from model's average line, which could indicate possible nonlinearities as shown in figure 47. In case of model 4.4, also the y-intercept term in technical value is negative.

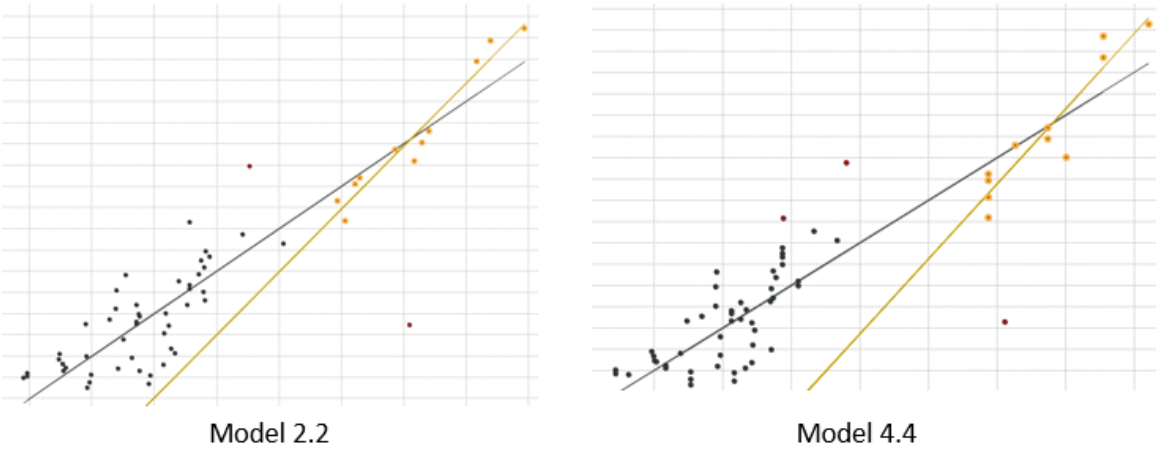


Figure 47. New slopes for data points at the higher end of the model.

The y-intercept of model 3.4 is also negative which could be a sign of nonlinearity. Based on visual investigation of scatter plot, there is not similar nonlinearity among data points with higher technical values like with models 2.2 and 4.4. In case of model 3.4, the possible reason for negative y-intercept may be a consequence of uneven distribution of residuals in technical value range of 110-120 (see figure 48 below).

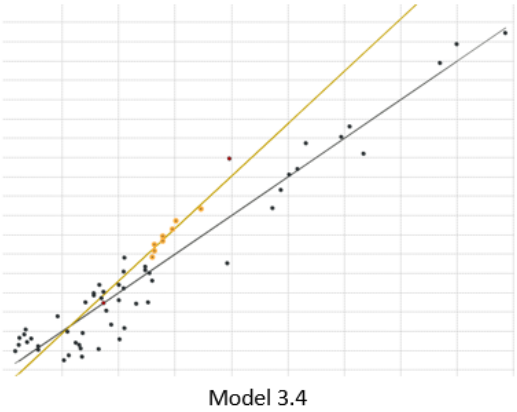


Figure 48. Possible reason for negative y-intercept in the model 3.4.

As part of the scatter plot analysis, the bottom-up calculations were included as a part of the model. In total, calculations were done for 8 different part numbers including 3 type 3, 1 type 1, 1 type 4 and 3 type 2 cylinder calculations, and the intention was to cover the most of the technical value range. The locations of data points created with bottom-up calculations can be found from appendix 9. The figure 49 presents an example of scatter plot analysis with bottom-up calculations based on model 1.4. In this case, 8 bottom-up calculation data points (in green) were added to the scatter plot, but those are excluded from the regression. Blue points are the analyzed purchased data points and red points are data points taken out

from regression. With 6 cylinders, the results of bottom up calculations are below purchasing price and with 2 cylinders the bottom-up calculations are more expensive in relation to purchasing price.

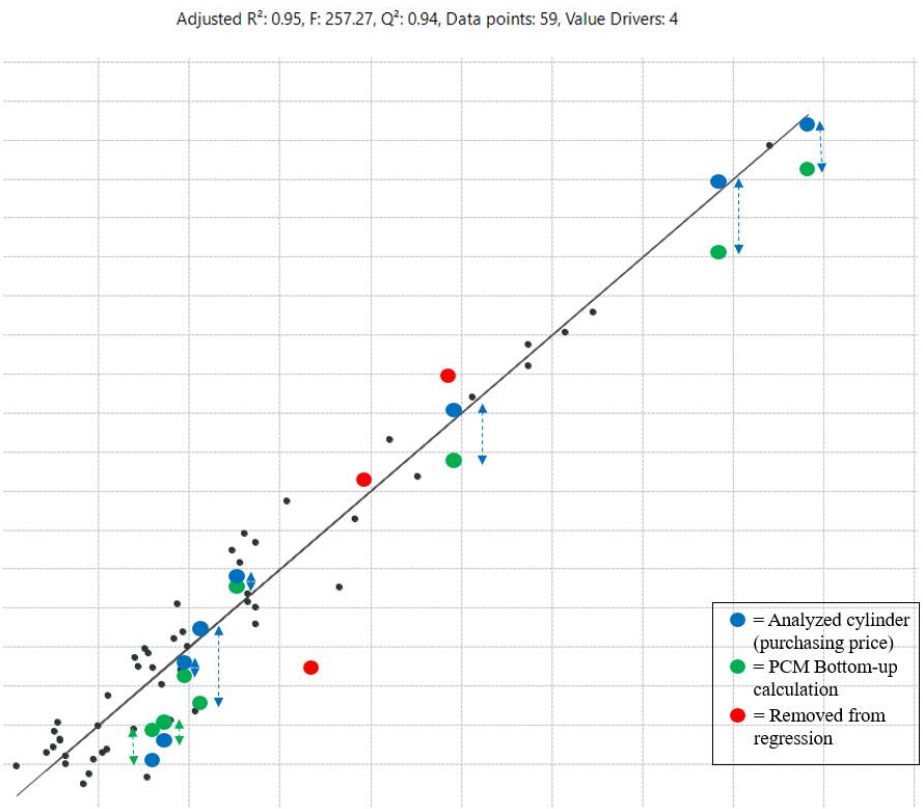


Figure 49. Bottom-up calculation results as part of the scatter plot, but not part of the regression (model 1.4).

When bottom-up calculations were added as a part of the regression, the change of statistical indicators adjusted R^2, Q^2 and F was analyzed (see table 19). A general trend was that including bottom-up calculations improved the values of statistical indicators. Especially the F value was increased in all models, which indicates increased explained variance in relation to unexplained variance. The impact of additional data points on technical values was also checked, but the changes were relatively small. Only with model 2.2, the intercept changes the sign from + to -. However, in case of model 2.2, the y-axis intersection parameter was pretty close to zero (5,06) at the first place with the original data. The appendix 10 includes the more detailed view of changes for all models.

Table 19. The change in adjusted R^2, Q^2 and F after bottom up calculations are added as part of regression function. Green color indicates positive change. Please see appendix 10 for more detailed effects on adding bottom-up calculations to the regression.

Model	**1.4**	**1.8**	**1.12**	**2.2**	**3.4**	**4.4**
Adj. R^2	0 %	0 %	1 %	1 %	1 %	2 %
Q^2	0 %	1 %	0 %	1 %	1 %	3 %
F	46,71	63,19	92,58	28,86	34,54	57,68

In case of this study, an important part of evaluation process is to assess the practical usability of the models for purchasing and R&D purposes. The usability includes two different perspectives; target costing and serial part analysis. The rating scale for the usability evaluation was from 1 to 4 (1 = best, 4 = worst). The ranking method was based on the principle that the same grade should not be given more than once in order to create a clear difference between the models. In the evaluation of usability, statistical indicators are used as a support, since in case it is challenging to make a difference between models, statistically better model gets the better ranking. For instance, it is relatively challenging to choose whether the models in group 1 or 2 are better for target costing. In this case, model group 1 gets the better ranking due to higher statistical values. During usability evaluation, both strengths and weaknesses of the models were assessed. In terms of usability, the structure of technical value formula is a key factor (what does it actually takes into account and how it can be utilized?). Furthermore, the orientation (cost vs. performance) of value drivers were evaluated and performance based value drivers are preferred.

Although statistical indicators like adjusted R^2, Q^2 and F do not directly reveal the best model, those are still giving good indication in terms of correlation, degree of explanation and stability of the models. On the other hand, all selected models already indicate good statistical capabilities (adjusted $R^2 > 0{,}87$ and $Q^2 > 0{,}86$), but there are still clear differences. Thus, at this stage of evaluation, statistical indicators are used as supportive criterion for model selection. The evaluation results for usability are presented in table 20.

Table 20. Usability of the models. The selected models highlighted in green.

Group		1. Weight + Cylinder types			2. Weight + Technical characteristics	3. Cylinder types + Technical characteristics	4. Technical characteristics only
Number		1.4 (matrix input for weight)	1.8 matrix input for weight)	1.12 (matrix input for weight)	2.2	3.4 (matrix input for OL)	4.4
Value drivers		Weight + All cylinder types as separate	Weight + (Type 1 and 3) + Type 4 + Type 2	Weight + (Type 1, 2 and 3) + Type 4	Weight + Rod diameter + Operational length	All cylinder types as separate + Inner diameter + Operational length	Operational length + Rod diameter
Statistical indicators 1 = best 4 = worst		1			3	2	4
Usability 1 = best 4 = worst	Target costing	2			3	1	4
	Serial part analysis	1			3	2	4
Strengths		• Good approach for price analysis, when weight and cylinder type are known • Well suited for evaluating the pricing of serial parts due to highest correlation (*Adjusted* R^2) and model's stability (Q^2)			• Takes into account the OL, which can be considered as performance parameter of the cylinder	• Takes into account both ID and OL, which are perhaps the two most interesting performance parameters of the cylinder	• No clear strengths compared to other models (explaining the pricing based only on technical characteristics seem to be challenging with selected sample)
Weaknesses		• The exact weight of the cylinder is not known at the beginning of the NPI project, since design is done by suppliers • In principle, the weight is not something company is willing to pay for • No performance parameters as value driver			• Absence of cylinder type causes some uncertainty, since cylinder type seem to be an important categorical value driver • Weight not known at the beginning of the NPI project	• Weight is not included, which as an individual value driver seem to have the strongest correlation with price	• Does not directly consider the diameter of the cylinder, which is the major weakness (and value driver combinations of ID and OL are not statistically reliable)

Based on usability- and overall evaluations, models 1.4 and 3.4 were chosen for further investigations. In general, all models in group 1 (Weight + Cylinder type) are statistically valid, and model 1.4 was selected due to most comprehensive set of categorical value drivers, which may facilitate the usability and analysis of the results of the model. The scatter plot of model 1.4 is presented in figure 50. Model 3.4 was selected as the second model, since it was the best model without weight as value driver and hence it is better in terms of usability especially with new NPI projects, when the exact weight of the cylinder is not known.

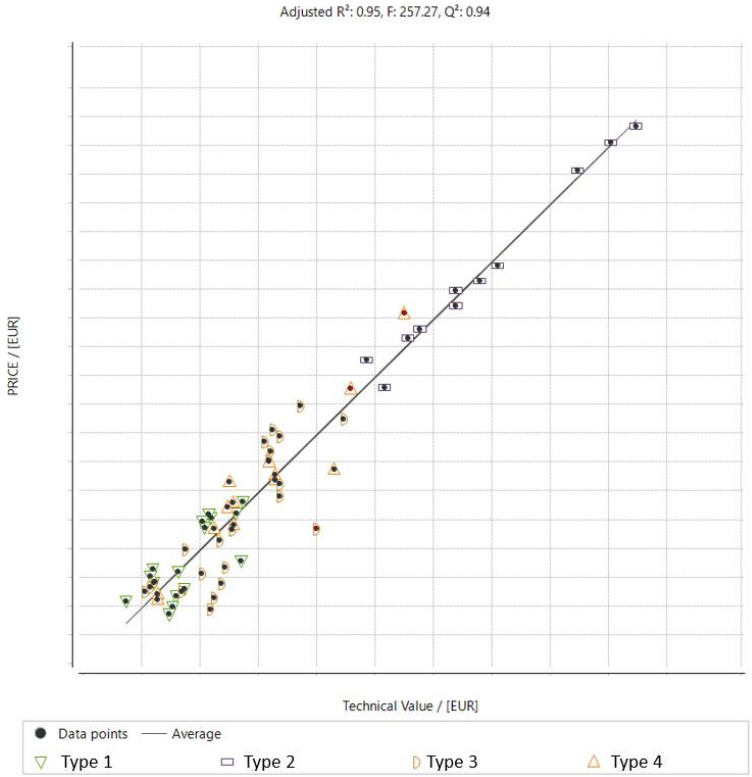


Figure 50. Scatter plot based on model 1.4. Please notice the difference in data point locations compared to figure 45, where weight was used as single value driver.

Additionally, selection of relatively different models in terms of value drivers enables a wider perspective during further studies like analysis of individual data points or -groups. It also allows to compare possible differences resulted by different selection of value drivers. The figure 51 presents the main elements and formulation of technical values for both selected models.

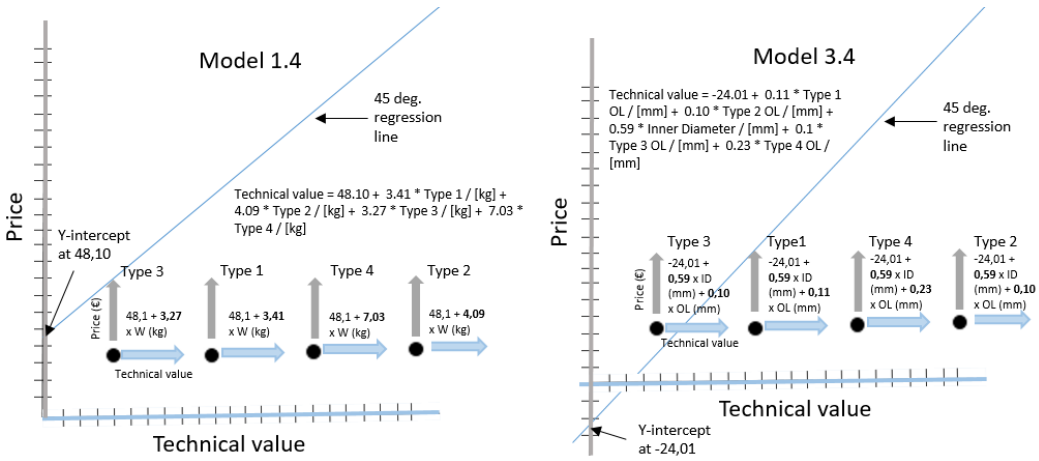


Figure 51. Selected models 1.4 (left) and 3.4 (right).

5.5 Assessment of data points with similar technical values

Although selected models 1.4 and 3.4 seem to be qualified after validation and evaluation procedures, further research on individual data points is required to check, how the model actually works, and why certain data points with similar technical values are located on top- or below of the regression line. For the assessment of data points, certain technical value ranges are selected. During this analysis, the idea is to explore, if there are factors (e.g. technical properties, supplier related factors etc.), which could possibly explain the pricing differences among parts with similar technical value, but are not part of the technical value formula. Especially data points located at the top- or bottom of the value range are the target of interest and more detailed investigations. At this point of analysis, the main focus is on comparison of technical properties.

The analysis of selected data points is carried out by selecting the data points located within the value range. After this, all collected data related to these selected data points is exported by using the “Export”-function of MLPP software, which enables relatively effective comparison between different cylinders. Simultaneously, technical drawings are used as a support to identify possible explanatory factors for higher or lower pricing. The selection of value ranges is done based on visual inspection, and for both models, two different technical value ranges are selected. As identified during scatter plot analysis, the upper end of the value range includes fewer parts compared to lower range, and in both models the data points located on upper end are relatively close to the regression line. Thus, the selected value ranges were chosen from lower end of the value range. With both models, the two selected ranges have technical values roughly 85-105 and 105-125. The selected value ranges for data point assessment are presented in figure 52.

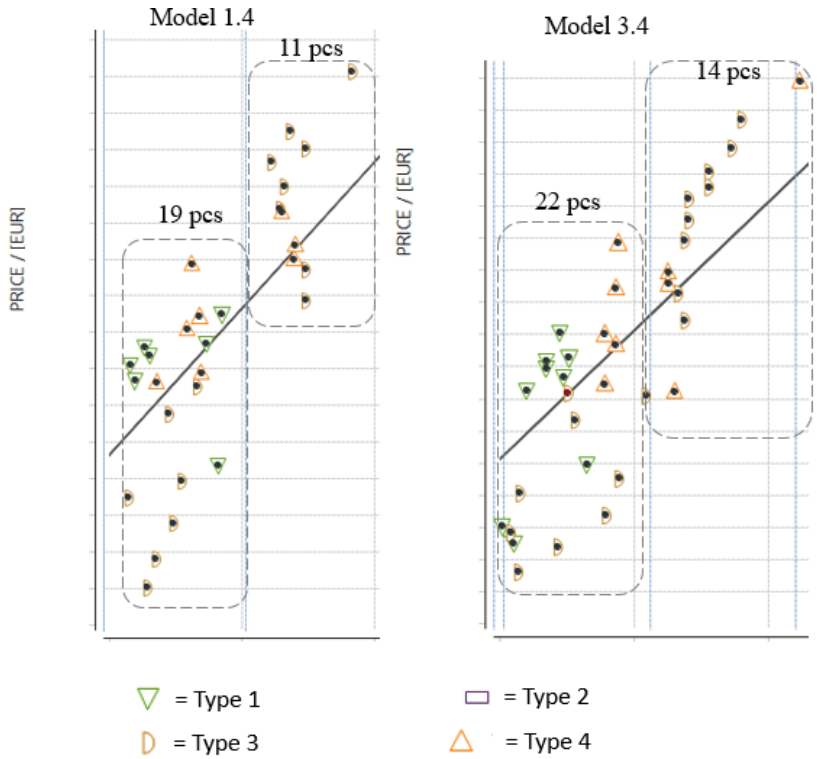


Figure 52. Selected value ranges for assessment of data points (filtered by cylinder type).

As a first step, both models were filtered by cylinder type. Based on earlier investigations during model testing phase, cylinder type has a strong explanatory effect on price. This was evidenced by the clearly improved correlation when the cylinder type was selected as categorical value driver in addition to technical characteristics. Even though MLPP fits different cylinder types based on their "essence" (properties which in this case are not explained by an individual technical or numerical value driver) into the model, a direct comparison of different cylinder types with same technical value is probably not meaningful. In this case and within the selected value ranges, a comparison especially between type 4- and types 1 & 3 cylinders is done separately.

As a second step of analysis, the selected value ranges were studied by adding suppliers to the same scatter plot like presented in figure 53, and possible pricing differences arising from suppliers were analyzed.

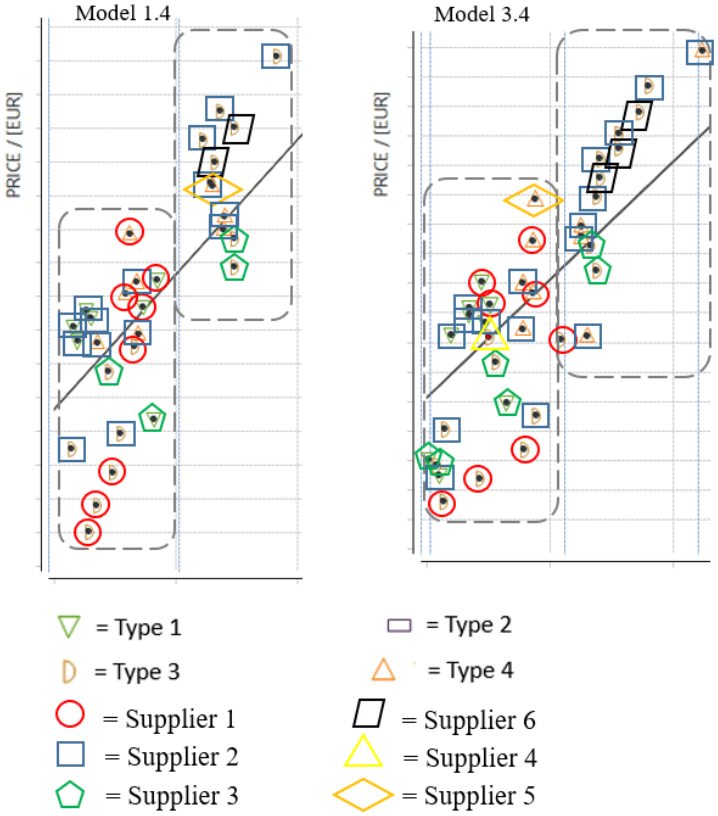


Figure 53. Selected value ranges with cylinder type and supplier information.

The main observations identified during data point analysis are presented in appendices 11 and 12. A general impression after data point analysis is that both models bring out the same favorable and unfavorable cylinders. In most cases, pricing differences identified among parts with similar technical values cannot fully be explained by means of technical specifications not included in technical value. With certain cylinders, the price variation is probably associated with suppliers using different technical solutions (e.g. piston rod design, material grades used etc.).

A relatively interesting finding in both models within value range of 85-125 were the locations of data points for type 3 cylinders. Figure 54 presents the own slope created for type 3 cylinders, which indicates that the price development is actually steeper within this range compared to slope based on all cylinders. However, this value range includes also type 3 cylinders, in which pricing is currently unprofitable, but the loss is compensated with type 1 cylinders. This may partially explain relatively significant difference in the slope. On the other hand, this probably explains also the location of type 1 cylinders on top of the

regression line, because two part numbers within value range 85-105 are currently clearly overpriced due to price compensations.

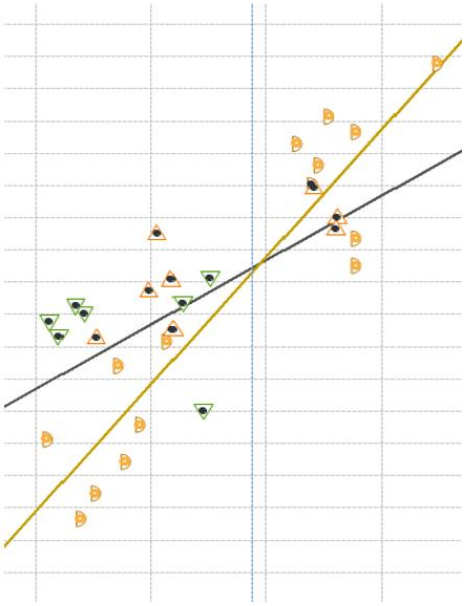


Figure 54. Slope of type 3 cylinders within value range of 85-105 (capture taken from model 1.4).

5.6 Analysis of potential data points and further utilization of the models

In terms of purchasing, the most interesting outcome of MLPP analysis are the unfavorable part numbers and identified saving potential. However, it is important to note that studies based on regression methods practically always provide saving potential since the regression line is placed according to average location of selected data points. In practice, the better is the correlation among data points, the lower is the saving potential due to better fit of the regression. In case of MLPP software used in this study, software allows user to identify interesting data points by using “Toggle visibility for selected data points” feature in value graph window. By default, this feature reveals 20% of the data points, which have the highest price / technical value ratio. In case of this study, all data points on top of the regression line (unfavorable data points) were selected. This enabled to perform comparisons between models 1.4 and 3.4 and study, if the same cylinders are located on top of the regression line in both models. Model 1.4 had 32 data points, whereas model 3.4 had 38 data points on top of the regression line (see appendices 13 and 14). The reason for different amount of parts can be explained by different adjusted R^2 values (0,95 vs. 0,92). When the commonality of

data points was studied, 30 of 32 (94 %) data points in model 1.4 were found from model 3.4 data points as well. Table 21 presents top 10 cylinders based on model 1.4 unfavorable data points. In this same table, the potentials within these data points between models 1.4 and 3.4 are also compared. Volume data is rounded to closest hundred. With NPI cylinders, the exact budgeted volumes are not known yet.

Table 21. Top 10 unfavorable data points of model 1.4 and comparison of potential to model 3.4.

#	Part number	Volume	Technical Value	Potential to AVG line in 1.4 (€)	Potential to AVG line in 3.4 (€)
1	ABC1234579	3100	112,15	17,00	13,66
2	ABC1234574	3700	121,57	15,78	16,94
3	ABC1234623	NPI	109,35	15,58	12,39
4	ABC1234616	300	97,34	13,60	9,25
5	ABC1234628	NPI	157,12	12,38	30,26
6	ABC1234604	1300	114,53	12,26	11,30
7	ABC1234606	1300	111,26	10,36	9,08
8	ABC1234611	1000	71,06	9,69	13,62
9	ABC1234625	NPI	90,22	9,45	8,37
10	ABC1234626	NPI	88,03	9,17	3,38

In terms of further utilization these models, through the new data repository feature, key stakeholders like buyers may use the selected models for evaluation of new quotations from suppliers through “Quotation comparison” feature in MLPP software. To ensure the usability and validity of the selected models over the years, regular updates related to input price data are required. In case of most hydraulic cylinder suppliers, the economics based purchasing price updates are carried out twice a year (based on agreements), which means that created MLPP models need to be updated accordingly to ensure the validity of the models.

6 RESULTS

The results achieved in this study are categorized into three different groups as illustrated in figure 55. The first part of the results consists of selected models and identified cost drivers for hydraulic cylinders, whereas the second part reveals the commercial benefits and opportunities for purchasing as a result of the MLPP analysis. The relevance of this research as an instruction document for MLPP forms the third part of the results.

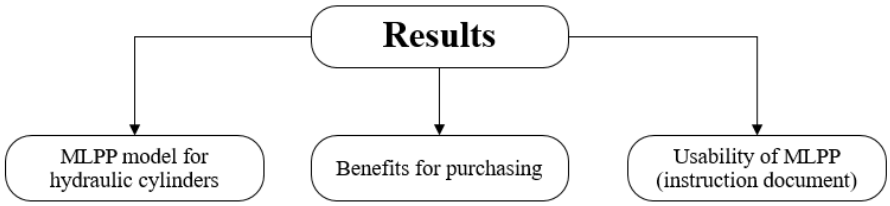


Figure 55. Overview of results

6.1 Selected models

First of all, MLPP turned out to be a useful analysis tool to study the pricing of hydraulic cylinders in relation to technical characteristics. As a result of model testing, two different models were selected after validation and evaluation process as shown in table 22.

Table 22. Selected models 1.4 and 3.4.

Number	1.4 (matrix input for weight)	3.4 (matrix input for OL)
Data points	62	62
Value drivers	Weight + All cylinder types as separate	All cylinder types as separate + Inner diameter + Operational length
Number of value drivers	4	5
Latent variables	4	5
Technical value	48.10 + 3.41 * Type 1 / [kg] + 4.09 * Type 2 / [kg] + 3.27 * Type 3 / [kg] + 7.03 * Type 4 / [kg]	-24.01 + 0.11 * Type 1 Ol / [mm] + 0.10 * Type 2 OL / [mm] + 0.59 * Inner Diameter / [mm] + 0.10 * Type 3OL / [mm] + 0.23 * Type 4 OL / [mm]
Adjusted R^2	0,95	0,92
Q^2	0,94	0,91
F	257,27	136,66
F critical	4,64	4,20
Sign F	0	0
Parts removed from regression	3	2
Primary usage	Serial parts	NPI Projects

Model 1.4 (based on cylinder types and weight) is better especially in terms of statistical indicators adjusted R^2, Q^2 and F, and definitely useful with serial parts, but due to selected

combination of value drivers, not so practical in terms of usability especially in the early stages of NPI projects. Thus, model 3.4 was selected as the second model to be used during NPI, for instance for target costing and concept phase cost evaluations.

As part of the value driver identification process, the major cost drivers were explored by using internal bottom-up calculation data and cost data received from supplier as a basis. Based on correlation matrices, cylinder weight as an individual value driver had the highest R^2 correlation 0,79 with price. Other identified cost drivers were collapsed length (0,72), stroke (0,58) and the operational length (collapsed length + stroke) (0,72). As an individual value driver, the inner diameter of cylinder tube (0,05) or rod diameter (0,16) had a very low correlation within the selected sample. However, when inner diameter is used as an additional value driver together with operational length and cylinder types, it clearly improves the statistical quality of the model. In case of selected scope, cylinder type turned out to be a necessary categorical value driver to be used due to relatively different structure of the cylinders, since especially cylinder types 2 and 4 differ from types 1 and 3 in terms technical properties, which is also reflected in the pricing. The original intention was to try to build a model based purely on performance parameters, but as a result of interviews and review of technical data it became clear, that it makes no sense to use derivative value drivers (e.g. cylinder force), since the hydraulic pressure used in working machines is constant. However, operational length used as value driver in model 3.4 can be considered as performance parameter.

6.2 Benefits for purchasing

In terms of purchasing, this study brought up multiple potential part numbers, which can be taken for further investigations and negotiations with suppliers. If the annual purchasing quantities (with serial parts) are included as a part of the cost potential calculation and the unfavorable data points are compared to regression line, model 1.4 indicates total saving potential of 2,85 % in relation to total purchasing spend of parts in scope. Similarly, the saving potential based on model 3.4 is 3,68 %. The total saving potential calculations on part number level are presented in appendices 15 and 16. Table 23 presents the identified statistical saving potential in relation to average line by supplier. In this context, the reporting of saving potential is done based on model 1.4 due to better statistical quality of the model.

Table 23. Identified statistical saving potential by supplier in relation to average line based on model 1.4.

Supplier	Identified saving potential (%)
Supplier 1	0,85 %
Supplier 2	3,38 %
Supplier 3	0,00 %
Supplier 4	0,00 %
Supplier 5	6,06 %
Supplier 6	2,93 %
Total	**2,85 %**

MLPP software allows user to perform potential identification by means of different calculation methods. One option is to compare the statistical potential by using mixed potential scenario, which means that the price of each data point is compared either to best-in-class line or to favorable line. Based on the location of the data point, the minimum potential out of these two scenarios is used to calculate the mixed potential like presented in figure 56. Table 24 presents the mixed saving potential by supplier based on model 1.4.

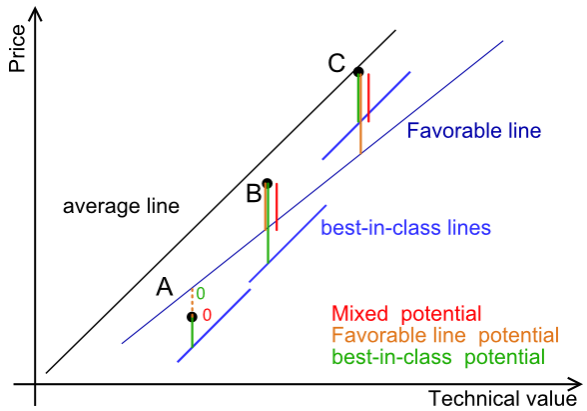


Figure 56. Principle of mixed potential scenario.

Table 24. Identified mixed saving potential by supplier based on model 1.4.

Supplier	Identified mixed saving potential (%)
Supplier 1	13,25 %
Supplier 2	9,85 %
Supplier 3	8,13 %
Supplier 4	0,00 %
Supplier 5	10,09 %
Supplier 6	13,93 %
Total	**9,68 %**

On top of the identified saving potential, the scatter plots created during this study enable to analyze the competitiveness of suppliers in terms of pricing, which can be utilized in future as a support for purchasing strategy. Based on scatter plot analysis, there are clear differences among suppliers. In addition to identified commercial potential, the models created in this research allow purchasing to study the competitiveness of new cylinders with “Quotation comparison” feature in MLPP software.

6.3 Instruction document

As a concrete result, this research works as detailed instruction document for MLPP model development inside the organization. The approach and procedures like regression function- and coefficient validation used during this study for hydraulic cylinders are applicable for other product groups and commodities as well. This creates a solid base for effective and valid MLPP model development, and this study is already considered to be the basis of the strategic initiative “MLPP implementation” among AGCO Supplier Cost Analysis team. Thus, as part of the implementation of this strategic initiative, the findings made during this research may be utilized for example during pilot projects and trainings.

7 DISCUSSION

In this chapter, the validity, reliability and sensitivity of this research are discussed, which form an important part, when the value of this research is evaluated. Moreover, the objectivity as well as the key findings are discussed. This research and the results are also compared to previous research made around this topic. Finally, the novelty value and generalization of results is assessed and possible future research ideas are discussed.

7.1 Validity- and reliability analysis

In terms of validity analysis, the validity of this research is separated into the validity of the indicator and the validity of the results. The main indicator is the technical value equation formed by several selected value drivers, and the internal validity of technical value consists of validity of selected value drivers. The internal validity of the MLPP model results consists mainly of the selected sample and different factors impacting on pricing. The generalizability of the results describes the external validity of the results. The main elements of validity analysis are presented in figure 57.

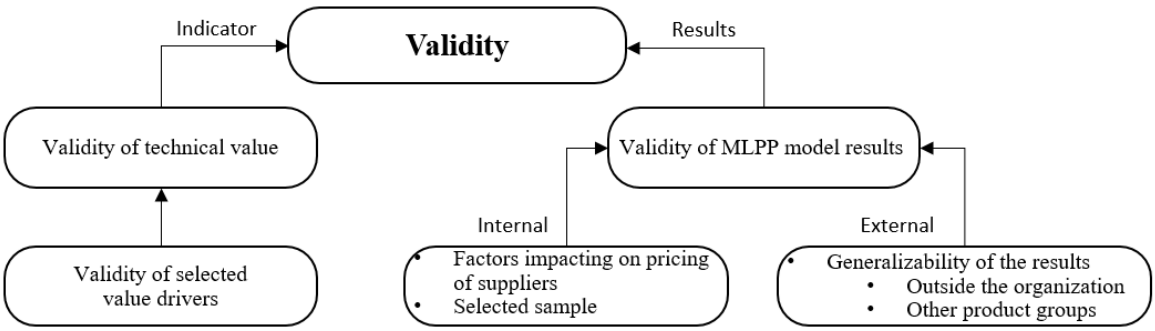


Figure 57. The main elements of validity analysis.

In practice, the high validity of technical value means that technical value corresponds the pricing of the suppliers within the whole value range of technical value. Based on adjusted R^2 values, both selected models indicate excellent results (0,95 and 0,92) in terms of adjusted R^2, which means that the selected combinations of value drivers extensively explain the pricing of selected scope of cylinders. Additionally, bottom-up calculations were performed to verify the technical value. In most cases, the results of bottom-up calculations were

systematically lower, but still in line with the regression line. The results of bottom-up calculations were actually almost in parallel with the average line and close to the favorable line of the model. Only with one selected cylinder, the calculation result was more expensive in relation to the regression line, but still slightly cheaper in relation to purchasing price of this cylinder. With this cylinder, the higher cost is mainly resulted for the forged piston rod – end-eye structure and different raw material grade. In practice, this means that heat treatment-, grinding-, chrome plating- and milling operations are done in-house, which increases the manufacturing costs compared to other cylinder in the model. In most cases, the piston rod material (e.g. 20MnV6) is already heat treated and plated by the raw material supplier, and basically the only manufacturing processes done by supplier are turning and welding process, which is usually more cost-effective solution. The challenge with this kind of supplier related manufacturing differences is that those are not necessarily marked in the technical drawings, and thus cannot be considered during data collection. This particular cylinder is a good example that even though correct value drivers would be selected, there are always factors that cannot be taken into account in the technical value. On the other hand, the goal in the MLPP model creation should be to minimize the usage of value drivers, but still choose the value drivers that explain pricing the most. In overall, the commonalities within two selected models are on high level, which increases the validity of selected value drivers and technical value.

In terms of more detailed analysis for technical value equations, both models indicate that type 4 cylinders are significantly more expensive in relation to other cylinder types. This corresponds the reality and can be easily seen by simply comparing € per kg prices of selected cylinder types. The main reasons for the expensiveness of type 4 cylinders are different piston rod structure (usually heavier in relation to cylinder structure), more complicated mounting arrangements, oil port and tube design as well as different sealing requirements due to more demanding dynamics (higher velocity). Additionally, these cylinders are typically the lightest and usually €/kg- prices decrease when the weight of product increases due to smaller proportion of manufacturing costs. If type 1- and type 3 cylinders are compared, the major difference is the assembly position, since type 3 cylinders are pushing (lifting work is done by piston side), whereas type 1 cylinders are pulling (work done by piston rod side) cylinders. In terms of technical properties, this may slightly impact on design of type 1 cylinder head and sealing. In general, type 3 cylinders are typically

heavier compared to type 1 cylinders. Type 2 cylinders are by far the largest in terms of weight and operational length, and differ from the others by having an additional tube structure at the piston side. Due to relatively different technical properties and significant differences on weight and operational length, the reasonableness analysis of technical value is pretty challenging to do by looking only at the values of regression coefficients, since the fit of the regression and the y-intercept is also part of the formulation of regression coefficients. For example with weight-based model 1.4, the regression coefficients for different cylinder types are not directly comparable to €/kg-prices of each cylinder type. Instead, validity analysis should be based on assessment of data point locations and comparisons of the overall results of the models. In that sense, both selected models give really similar results in terms of favorable and unfavorable parts, since almost all (94 %) of the unfavorable data points of model 1.4 were also found from unfavorable data points of model 3.4.

When evaluating the internal validity of the results, it is good to keep in mind that the results of the MLPP model are as valid as the pricing of suppliers. In addition to suppliers, there are also other factors, which are not directly dependent on the supplier, but still have an effect on pricing of cylinders. The table 25 presents 10 possible factors, which are challenging to take into account in model creation, and which may impact on validity of internal results. The order of identified factors is defined from the perspective of this research and the selected sample (1 = higher impact on internal validity, 10 = lower impact). These factors are also categorized into different categories (technical, commercial, location, manufacturing, quantity and quality), which emphasizes the diversity of identified factors.

Table 25. Possible identified factors impacting on validity of internal results.

1.	Differences on technical properties which are not specified in drawing (e.g. raw material grades, forging vs. bar material, surface quality- and tolerances of machined components, testing processes, quality of seals etc.).	Technical
2.	Sales price does not correspond costs (there are recognized cases inside this commodity in which certain cylinders are clearly unprofitable whereas others may be really profitable). This is also possible scenario with new suppliers whose target to increase business with new customer (foot-in-the-door technique).	Commercial
3.	Production locations of sub-components are different and the proportion of best-cost-country (BCC) components may vary.	Location
4.	Companies have different manufacturing strategies which impact on pricing (make-or-buy). Within selected group of suppliers, there are quite remarkable differences (e.g. in manufacturing strategies of machined components like	Manufacturing

	pistons, cylinder cap- and head, end-eyes etc.). Rod- and tube production are typically in-house processes within all suppliers.	
5.	Volume has certain impact on pricing. However, it is possible that original volume (as a base for the purchasing price) does not correspond the current price. Additionally, sometimes the budgeted volume forecasts may be incorrect. Thus, usually volume should be used only for evaluation of results, not as value driver.	Quantity
6.	Cost structures of companies are different and pricing principles of suppliers may vary (e.g. the share of overheads and profit).	Commercial
7.	Different phases of annual long term agreement (LTA) reductions (e.g. -3%, -3% and -3%) may distort MLPP model with similar cylinders (e.g. especially cylinders which have been in serial production for a long time).	Commercial
8.	Quality requirements may vary among different brands (e.g. volume vs. premium platform), which impacts on pricing.	Quality
9.	Even though sales prices are taken from same quarter, market situation (recession vs. growth, supply vs. demand) may have been completely different at the time of original pricing.	Commercial
10.	Tracking principles of raw material cost are different (e.g. several indices) and raw material content may have been evaluated differently.	Commercial

Moreover, the results identified in this research are based on certain selected sample and distribution of data points in terms of cylinder type and suppliers has a certain impact on the results. In this study, one supplier has clearly the biggest portion (61 %) of the data points, which means that the impact of this particular supplier to the model is higher compared to others. Similarly, the distribution of cylinder types and purchasing prices affects the technical value. The cylinder types were pretty evenly distributed in terms of quantity of data points like shown in table 6 on “Research methods” chapter (type 3=35 %, type 1=26 %, type 4=21 % and type 2=18 %). In terms of purchasing prices, roughly 70 % of the prices were within 60 and 140 €, which means that the analysis will be weighted accordingly. In addition, one selected cylinder type is purchased from one single supplier, which means that the market benchmark for them is pretty narrow. With this kind of product groups, the importance of bottom-up calculation method increases. It should also be noted, that each time certain data point is removed from regression, the model will be recalculated and the results will change. However, typically certain amounts of data points must be removed to reach sufficient statistical quality. In case of this study, the intention was to keep the amount of removed part numbers as low as possible, and basically R^2 sensitivity of -10 % was used as a threshold for data point removal.

The external validity of the results may be evaluated by assessing the generalizability of the results for other hydraulic cylinders outside the sample and also outside the organization. In general, the model selection, validation and evaluation processes utilized in this study are

well applicable for any other product groups as well. In case of hydraulic cylinders, the identified technical values should be usable for similar group of cylinders, but it must be noted, that the technical values found in this research are highly targeted for cylinder types used in tractors. The usage of categorical value drivers partially reduces the direct utilization of the results. Inside the corporation, the findings of this research can be further utilized for example for the whole global cylinder purchasing spend including all other regions (APA, NA and SA) as well. In this case, one possibility is to create models for instance only for type 1 & 3 cylinders, since the global data should include enough parts for more detailed study of individual cylinder types.

In terms of reliability of the statistical results, the Q^2 value (Stone Geisser Criterion) indicates the model's stability (responsiveness) in case an individual data point is removed from the analysis. In both models, the Q^2 value is on the high level (0,94 and 0,91), which means that removal of individual data point does not significantly change the results of the model. However, it must be remembered that these Q^2 values are relevant only for the selected sample and selected combination of value drivers. In terms of overall reliability of the research, it is necessary to notice that with MLPP analysis the reliability of the whole analysis is based on carefully executed processes, starting from data collection and ending with the evaluation of multiple different model options. Additionally, long-term and iterative model testing plays an important role in the generation of technical value and model results.

7.2 Sensitivity analysis

Due to statistical nature of this research, the results achieved are highly dependent on the collected technical input data and selected value drivers. These elements are in the major role, when the sensitivity of this research is evaluated (see figure 58).

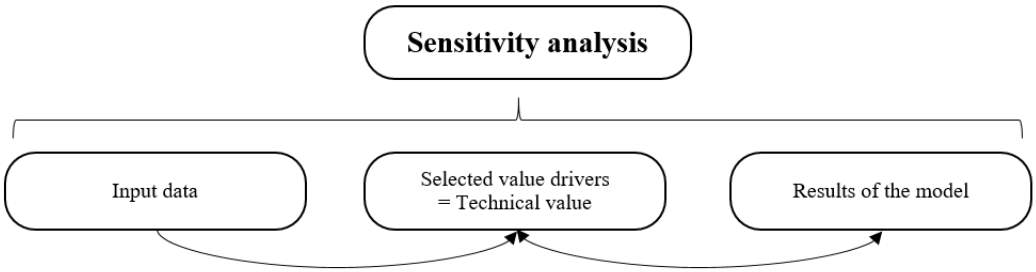


Figure 58. Sensitivity analysis of this research.

Since the iterative testing forms an important part of the model creation, the arrow between technical value and results points in both directions. In principle, there could be an additional arrow between input data and results as well, to ensure that the sensitivity of input data is on sufficient level in relation to desired results. In case of this study, the intention during data collection was to collect all technical data possibly impacting on pricing as accurately as possible, and then perform the selection of value drivers as separate process step. However, although the data collection was carried out carefully, the collected data practically always includes certain amount of errors. In this research, the errors may be related to data collection itself, errors done by suppliers during pricing or other factors impacting on pricing like agreements, discounts and volumes. For instance, one possible error element in input data could be the weight data, since the availability of reliable weight information of the cylinders was surprisingly poor. Thus, the weight information was partially collected by asking the information directly from the suppliers. The weight information was collected in kilos with a precision of one decimal, which was mainly based on the accuracy of the official weights in technical drawings. All other technical dimensions were collected with precision of 1 millimeter and price information with two decimals, which corresponds the accuracy used in the ERP system.

Additionally, due to the fact that the data collection was done manually, the risk of human error increases. The sensitivity related to value driver selection and determination of technical value is mostly based on the question, if the correct or the best combination of value drivers was actually chosen. In this regard, ensuring the right selection of value drivers, different model options were tested very extensively. As a result of value driver testing, the decision of utilization of categorical value drivers was made. Due to several statistical indicators utilized during the model development, it was relatively easy to follow up the impact for example on model results, when the combination of value drivers was changed. In the model selection phase (after the values were exported from MLPP software), the utilized precision of statistical indicators was two decimals. In practice, the software uses greater precision during the calculation, but precision of two decimals proved to be sufficient for model comparison purposes. For instance, in case of adjusted R^2, the precision of two decimals means the precision of one percent in the degree of explanation, which clearly distinguished the models from each other and enabled the comparison of different options.

7.3 Objectivity

The objectivity of this research can be evaluated by assessing whether other researchers would have ended up to the same results, and which factors may impact on the objectivity of the research. Due to the relatively limited discussion in literature especially on the practical application of MLPP, the theoretical background of this research is partially based on a few main sources (VDI 2018, Processbench user menu & e-learning material). However, interviews and benchmarking provided valuable information and strengthened the basis for model development. Furthermore, some commercial sources were utilized to expand the technical knowledge base for hydraulic cylinders. To ensure the objectivity of the research, all utilized research methods are described as accurately as possible starting from the data collection and processing and ending with the evaluation of the results. The model selection made during the study are largely based on the statistical indicators, but the perspective of the corporation and researcher was also in a key role, when the models were assessed especially in terms of usability. This could also be one area that could undermine the objectivity of the research. However, the factors influencing decision making are described in a detail level (e.g. why certain value drivers is working or not working in analysis of NPI cylinders?). Furthermore, a researcher always forms one part of the whole research process. Thus, for example the background and location of the researcher (e.g. AGCO site) or the previous knowledge regarding products in the scope may impact on results. In case of this research, the background of the researcher or the existing product knowledge could have an impact especially on the processing of data points during model testing (e.g. which parts are taken out of regression?). To overcome this issue, the threshold used for data point removal (R^2 sensitivity > 10 %) during model testing was relatively high and all data points were treated as equally as possible, despite the relatively comprehensive knowledge regarding the products in scope. Additionally, the bottom-up calculations used for model evaluations are also based on the knowledge and skills of the author, so those can be considered to be at least partially subjective.

7.4 Key findings

The most important finding of this research was that MLPP proved to be suitable statistical analysis method for exploring the pricing of different type of hydraulic cylinders used in

tractors. Within the selected scope of products, in order to guarantee sufficient statistical reliability, the usage of cylinder type as categorical value driver was necessary due to technical differences especially among type 4 and type 2 cylinders. Additionally, cylinder weight had a strong correlation with pricing and the weight-based models were statistically best, but it was possible to create a reliable model based on operational length and inner diameter of the cylinder tube as well, which definitely improves the usability of MLPP during early phases of NPI projects. In general, it was challenging to further improve the statistical quality of weight-based models by selecting an additional technical characteristic as a value driver. This can probably be explained by the fact that weight and cylinder type are already such a powerful value drivers, that an additional value driver is not bringing any further added value. On the other hand, building the model purely based on technical characteristics did not give statistically reliable results, and with most combinations, it was not possible to combine dimensions like inner- or outer diameters and lengths into the same technical value. Within selected sample, the reason is probably the diversity of cylinders in terms of length and diameter. For example the diameter of type 2 cylinders is typically very low compared to operational length, whereas with other cylinder types this ratio is completely different. Basically, the best identified combination based on technical characteristics only was based on operational length and rod diameter (adjusted R^2=0,88, Q^2=0,87 and F=219,18). However, even though these value drivers explain the pricing of selected sample pretty well, this kind of combination does not take into account the cylinder tube at all and is thus quite deficient.

During model development, R^2 sensitivity [%] value was utilized to identify clear outliers from the model. Even though it gave good guidance, in some cases the impact of other data points with significantly lower R^2 sensitivity value may be remarkable. Thus, a removal of data points from the model should be carefully considered. An interesting finding regarding categorical value drivers was that in case the traditional data input format is used, there is a higher tendency for the unexplained minimum and maximum values (confidence interval) of regression coefficients. The tricky thing with this kind of models is that the main statistical indicators (adjusted R^2 and Q^2) may indicate good results (>0,90), but in terms of validity of regression coefficients the model is useless. In these cases, value driver folding (matrix input) turned out to be useful method to overcome this issue. At the same time, the statistical quality of the model was improved. In general, the detailed study of regression coefficients

in terms of confidence intervals and t-values was something, which can be further utilized in the validation of existing MLPP models inside the corporation.

Even though the importance of the weight in relation to pricing was identified during early meters of the research, it was interesting to notice that also models based on technical characteristics and cylinder types offered respectable results, and for the most part, the results of these models were in line with the weight-based models. Additionally, although weight as individual factor is useful for rough price estimations, including the cylinder type as a categorical value driver significantly improves the accuracy of analysis, which can be directly seen by comparing the statistical results of one-dimensional and multidimensional weight-based models. In practice, the comparison between the scatter plot presented in figure 45 and scatter plots based on MLPP (see for example figure 50) reveals the added value of categorical value drivers and multilinear regression for the price analysis.

7.5 Comparison with former research

Even though more advanced regression analysis methods are increasingly used especially among companies with a lot of purchasing activities, and there are several software solutions for both linear- and nonlinear regression analysis, there is a limited amount of former research available regarding the practical applications of PP method. A lot of information can be found in regards with mathematical proofs of multilinear regression analysis, but especially in the business world, they can often be too heavy to apply into reality. The VDI 2817 Part 1 (2018) document created by Association of German Engineers discusses the implementation process of performance pricing method and this was utilized as a guideline in this research. In practice, the main procedure introduced in this document follows the principles already used inside corporation. However, particularly the detailed tips for instance on the validation of individual regression coefficients and residuals were useful and brought added value to this research as well.

There do not appear to be any previous studies examining the pricing of hydraulic cylinders by means of MLPP. Zhou, Chen & Gao 2013 discusses the price forecasting for hydraulic cylinders through case based reasoning- and back propagation neural networks methods, and the selected cost factors for forecasting consist of quantitative and qualitative factors. In

terms quantitative factors, cylinder bore diameter, stroke and producers price index were utilized. Qualitative factors were pressure, cylinder type (servo or not), installation type and supplier qualification. From previously mentioned factors, the cylinder type was considered to have the most significant weight for the price forecasting. The VDI 2817 Part 2 (2020) presents examples of possible value drivers and filters to be used for different product groups, and perhaps the nearest benchmark commodity is the automobile shock absorbers. By comparing the value drivers tested in this thesis, the portfolio of possible value drivers is pretty similar, and for instance the overall weight of shock absorber is on top of the list in this document as well. Moreover, the compressed (collapsed length), extended length (operational length), piston rod diameter and internal diameter of cylinder tube can be found from this list. In general, VDI 2817 Part 2 proposes to use country of origin, manufacturer and quantity as a filter during evaluation of the models.

7.6 Novelty value and generalization of the results

This research applied MLPP analysis method for hydraulic cylinders quite extensively including several process steps like identification of value drivers, model testing, statistical validation and evaluation of the model. At the same time, this study brought up the most important cost drivers impacting on pricing of hydraulic cylinders. Meanwhile, the amount of public studies related to MLPP, cost drivers and pricing of hydraulic cylinders is very limited, which increases the relevance of this research. Although the study was made for the selected sample including certain types of cylinders and the objectives were mainly related to client company, the similar approach may be utilized for other research purposes as well. It is relatively common that product, service, feature or phenomenon is affected by several variables, which impact on the outcome of the matter or object under investigation. In practice, the applicability of MLPP analysis is basically limitless, as long as the multi-stage nature of the model development process are kept in mind.

7.7 Future research ideas

From a corporation perspective, the further investigation will be done for the potential unfavorable cylinders identified in this research by means of more detailed bottom-up calculations. Furthermore, the application of MLPP process utilized during this study is

applicable for other commodities as well, which facilitates to increase the spend analysis coverage inside the corporation. The detailed validation process introduced in this study may be carried out for existing MLPP models. With help of recently launched APT (AGCO Purchasing Tool) spend data, it is possible to do data collection for global cylinder spend, and the higher number of parts probably enables the creation of own models for each cylinder type as well. On the other hand, based on this research, the usability of categorical value drivers probably allows to place all data points into the same model. In terms of global models, managing the differences caused by economics is probably the greatest challenge, which needs to be evaluated during model planning. Another further research idea is related to practical usability of MLPP models in the organization. This is important especially due to the upcoming data repository feature, which enables key stakeholders in purchasing organization to have an access to existing MLPP model database. Thus, regular and structured model updates due to changes in economics are necessary to ensure the validity and usability of the models. Due to differences among product categories analyzed with MLPP, the update method should meet the requirements of several commodities.

In general, the utilization of automatic machine learning applications would likely increase the accuracy and efficiency of the data collection. At the same time, it would enable the utilization of more comprehensive assortment of value drivers, since for example the accurate measuring of sub-component dimensions could be carried out automatically. This would probably also lower the threshold for MLPP model creation. However, despite the more advanced data collection possibilities, the input data (technical drawings, 3Ds and product group categorizations in ERP etc.) must be consistent, which is probably still the greatest challenge among many companies. In general, the data collection phase in the beginning of MLPP model development works also as a really efficient indicator for company’s data processing- and design level, since it reveals the weak points of ERP (e.g. deficient product categorization) and PLM (e.g. missing technical specifications) systems.

8 CONCLUSIONS

As a statistical method, MLPP offers an interesting approach for analysis of high number of parts, since it enables to form a comprehensive overview of the relationships of products, in terms of pricing and for example suppliers inside the selected commodity. The creation of MLPP model is relatively easy, but the analysis of model results as well as the validation and evaluation of the model are significantly more challenging. At this point, technical- and cost related product knowledge are required. Even though MLPP allows to explore large amount of data based on detailed statistical indicators- and methods, the output of the model must always be carefully evaluated. The risk of this type of analysis is that in some cases statistically good results may be achieved with nonsensical value drivers as well (e.g. value drivers have high correlation with each other). Thus, it is necessary to perform certain pre-study, like explore the dependencies among potential value drivers before actual model is created. In terms of purchasing, the preferred approach for model creation would be to utilize performance related parameters as value driver, since typically those describe the “things” that customer is willing to pay for. However, utilization of performance parameters is not always possible, and typically technical value includes both performance- and cost based parameters.

Even though the common goal in MLPP model creation is usually to explain the pricing of suppliers as extensively as possible and the intention is to create high-quality statistical model, the main purpose of MLPP should be kept in mind. In general, the preferable purpose of MLPP is to bring up the unfavorable products in terms of pricing, which are then further analyzed with more detailed methods like bottom-up calculation. Especially in the business world, creating a perfect mathematical model or over-analyzing the model should not be the priority, but MLPP should be used as a tool to search potential on the higher level. Due to the nature of MLPP analysis, finding a “single truth” in terms of technical value may be a challenge, since especially with more complex products there are always features and factors, which cannot be taken into account in selected value drivers. Hence, the most important thing with MLPP is to understand, what is actually considered in the technical value formula and what is not. Within this research, the results achieved from the models supported surprisingly well the existing knowledge of relationships between suppliers and pricing in the commodity of hydraulic cylinders, since many of already recognized potential

part numbers were identified as unfavorable in MLPP results as well. Although the statistical results of different test models varied due to different combinations of value drivers, the same data points were constantly found either on top or below of the regression line in both models. By keeping two models (1.4 and 3.4) up to the end of the model development process also enabled to compare the results based on different combination of value drivers, which turned out to be a useful approach for the assessment of the individual data points and model results.

LIST OF REFERENCES

AGCO. 2021. AGCO Your Agriculture Company. [Web page]. [Accessed 05.05.2021]. Available: https://www.agcocorp.com/brands.html.

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. 2018. Multivariate Analysemethoden - Eine anwendungsorientierte Einführung. 15th Ed. Berlin: Springer Gabler, Berlin, Heidelberg. pp. 58-124.

Bedre, R. 2021. Multiple linear regression (MLR). Data Science Blog. [Web page]. [Accessed 26.05.2021]. Available: https://www.reneshbedre.com/blog/multiple-linear-regression.html#coefficient-of-determination-r-squared-and-adjusted-r-squared.

Berkman, E. T. & Reise, S. P. 2012. A Conceptual Guide to Statistics Usings SPSS. Thousand Oaks, California 91320: Sage Publications Ltd. 312 p.

Berz, G. 2021. Processbench GmbH. Interview. 11.02.2021.

Bosch Rexort. Cylinders. [Web page]. [Accessed 12.03.2021]. Available: https://www.boschrexroth.com/en/xc/products/product-groups/industrial-hydraulics/cylinders.

Chapple, P. 2015. Principles of Hydraulic Systems Design. 2nd Ed. New York: Momentum Press. 287 p.

Chatterjee, S. & Simonoff, J. S. 2013. Handbook of Regression Analysis. New Jersey: John Wiley & Sons, Inc. 225 p.

Childs, P. R. 2019. Mechanical Design Engineering Handbook. 2nd ed. Oxford: Butterworth Heinemann. Elsevier Ltd. 968 p.

Chin, W. W. 1998. The partial least squares approach to structural equation modeling. Modern methods for business research. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, pp. 295-336.

Coşkun, G. & Yilmaz A. 2021. Determining the best price with linear performance pricing and checking with fuzzy logic. Computers & Industrial Engineering. Department of Industrial Engineering, Sakarya University. Volume 154. pp. 1-14

Degelman. 2016. Hydraulic Cylinder Repair Guide. [Web document]. [Accessed 12.03.2021]. Available: https://degelman.com/pub/resources/manuals/cylinders/cylinder-repair/143367-Hyd-Cylinder-Repair-Guide.pdf.

Doddannavar, R. & Barnard, A. 2005. Practical hydraulic systems operation and troubleshooting for engineers and technician. Oxford: Newnes. 240 p.

Faraway, J. J. 2002. Practical regression and ANOVA using R. University of Bath. 201 p.
Galal Rabie, M. 2009. Fluid power engineering. US: McGraw-Hill Education. 420 p.

Geisser, S. 1975. The Predictive Sample Reuse Method with Applications. The Journal of the American Statistical Association. Volume 70. pp. 320-328.

Güthenke, G., & Möller, H. 2007. Application of price performance analysis in different supplier-buyer relationships. 23rd IMP-conference in Manchester, UK. pp. 1-17.

Harrel, F. E. 2001. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. 2nd Ed. New York: Springer. 582 p.

Heinze, G. & Dunkler, D. 2017. Five myths about variable selection. Transplant International. Volume 30. pp. 6-10.

Heney, P. 2012. What are hydraulic cylinders? Mobile Hydraulic Tips - A Fluid Power Word Resource. [Web document]. [Accessed 12.03.2021]. Available: https://www.mobilehydraulictips.com/what-are-hydraulic-cylinders/.

Hidalgo, B. & Goodman, M. 2013. Multivariate or Multivariable Regression?. American Journal of Public Health, 103(1). pp. 39-40.

Hilti. 2017. Hilti Performance Pricing Training Material.

Hydroline. 2019. Hydraulic Cylinder Blog. What is hydraulic cylinder? [Web page]. [Accessed 12.03.2021]. Available at: https://hydroline.fi/blog/what-is-a-hydraulic-cylinder/.

IBM, SPSS Statistics - Linear Regression Variable Selection Methods. [Web page]. [Accessed 01.04.2021]. Available: https://www.ibm.com/support/knowledgecenter/SSLVMB_sub/statistics_mainhelp_ddita/spss/base/linear_regression_methods.html.

Jalonen, L. 2021. Design Engineer. Valtra Oy Ab. Interview. 17.03.2021.

James, G., Witten, D., Hastie, T. & Tibshirani, R. 2013. An Introduction to Statistical Learning with Applications to R. New York: Springer Science+Business Media. 418 p.

Karjalainen, M. 2021. Design Engineer. Valtra Oy Ab. Interview. 23.02.2021.

Leitol, S. 2021. Supply Chain Management, Functional Excellence, Benchmarking and Value Analytics. Siemens AG. Interview. 28.05.2021.

Liebherr. Engineering Expertise for Hydraulic Cylinders. [Web page]. [Accessed 01.04.2021]. Available: https://www.liebherr.com/en/int/products/components/hydraulics/hydraulic-cylinders/hydraulic-cylinder-technology/hydraulic-cylinder-technology.html.

Mankki, M. 2021. Design Engineer. Valtra Oy Ab. Interview. 23.02.2021.

Meierhofer, J. 2021. Design engineer, Fendt GmbH. Interview. 22.03.2021.

Mellin, I. 2006. Tilastolliset menetelmät: Lineaarinen regressioanalyysi. Helsinki University of Technology. pp. 231-433.

MEPS. 2021. The Latest Global Steel Prices, Indices & Forecasts. [Web document]. [Accessed 01.04.2021]. Available: https://www.meps.co.uk/gb/en/pages/the-latest-global-steel-prices-indices-forecasts.

Mobley, R. K. 2000. Fluid Power Dynamics. Boston: Newnes. 288 p.

Montgomery, D., Peck, E. A. & Vining, G. G. 2012. Introduction to Linear Regression Analysis. 5th Ed. New Jersey: John Wiley & Sons, Inc. 645 p.

Möller, H. 2009. Methodik zur Erzielung von Preistransparenz im Einkauf. Aachen: Apprimus-Verlag. 263 p.

Münch, R. M. 2015. Effizient mit Performance Pricing - Saphirion. All about Sourcing. pp. 8-9.

Newman, W. R. & Krehbiel, T. C. 2007. Linear Performance Pricing: A collaborative Tool for focused Supply Cost Reduction. Journal of Purchasing and Supply Management. Volume 13 (2). pp. 152–165.

Ovako. 2016. Ovako Products – Product Catalog. [Web document]. [Accessed 10.04.2021]. Available: https://www.ovako.com/en/steel-portfolio/ovako-brands/hard-chromes-surfaces/.

Parker. Heavy Duty Hydraulic Cylinders. Product Catalog. [Web document]. [Accessed 15.03.2021]. Available: https://www.parker.com/Literature/Industrial%20Cylinder/cylinder/cat/english/RDH_1320-2.pdf.

Parr, A. 2011. Hydraulics and Pneumatics - A Technician's and Engineer's Guide. 3rd Ed. Oxford: Butterworth Heinemann. Elsevier Ltd. 248 p.

Pipino, L. L. Yang, L. W. & Wang, R. Y. 2002. Data Quality Assessment. Communications of the ACM, Volume 45, pp. 211-218.

Pitkänen, J. 2021. Design Engineer. Valtra Oy Ab. Interview. 23.02.2021.

Proch, M., Krampf, P., & Schluechtermann, J. 2013. Linear Performance Pricing als Instrument zur Kostenoptimierung in der Supply Chain. Die Betriebswirtschaft, Volume 73 (6). pp. 515–534.

Processbench. 2017. E-learning material. analycess® Procurement.

Processbench. 2021. User Manual. analycess® Procurement.

Processbench & AGCO. 2017. Preisregressionsanalyse bei AGCO: Hydraulikschläuche. MLPP User Day presentation. Babenhausen.

Processbench, analycess® Procurement. [Web page]. [Accessed 09.04.2021]. Available: https://www.processbench.com/en/analycess_procurement/price_comparison_analysis.

Ruppert, D. & Matteson, D. S. 2015. Statistics and Data Analysis for Financial Engineering with R examples. 2nd Ed. New York: Springer Science+Business Media. 719 p.

Ryan, T. P. 1997. Modern Regression Methods. Wiley Series in Probability and Statistics. New York: John Wiley & Sons, Inc. 510 p.

Rydén, J., 2014. Lecture 6. Multiple linear regression. Model Validation, Model choice. Regression and Analysis of Variance. Department of Mathematics, Uppsala University.

Saphirion. Non-Linear Performance Pricing (NLPP). [Web page]. [Accessed 19.05.2021]. Available: https://www.saphirion.com/nlpp.html.

Saphirion AG. 2016. Evaluation “Performance Pricing” Solutions (Part 1). [Web page]. [Accessed 12.06.2021]. Available: https://medium.com/performance-pricing/evaluating-performance-pricing-solutions-16d0004b3b48.

SCA. 2021. AGCO Supplier Cost Analysis Team MLPP Data Repository.

Schuh, C., Kromoser, R., Pearce, J., Raudabaugh, J. L., Strohmer, M. L. & Triplat, A. 2017. The Purchasing Chessboard. 3rd Ed. New York: Springer Science+Business Media. 245 p.

Schmidt, F. 2019. Application of cost management methodologies for sustainable cost reductions of purchased parts - Comparison and evaluation of different methods in the industrial application. Journal of Applied Leadership and Management, Volume 7, pp. 66-80.

Sherefkin, R. 2003. Consultant vexes vendors. Automotive News. Detroit. Crain Communications, Incorporated. Volume 78(6063). pp. 1,46.

Shin, L., 2018. Graphs and ML: Multiple Linear Regression. Web Article on “Towards data science” web page. [Accessed 16.04.2021]. Available: https://towardsdatascience.com/graphs-and-ml-multiple-linear-regression-c6920a1f2e70.

Shrestha, N. 2020. Detecting Multicollinearity in Regression Analysis. American Journal of Applied Mathematics and Statistics, Volume 8(2). pp. 39-42.

Sieben, B. 2021. Head of Sourcing Excellence. Hilti Corporation. Interview. 28.05.2021.

SKF. Products/ Hydraulic Cylinders. [Web page]. [Accessed 16.03.2021]. Available: https://www.skf.com/group/products/industrial-seals/hydraulic-seals/general-technical-information/introduction-to-fluid-power/hydraulic-cylinders.

Soltau, O. 2018. Performance Pricing – Was ist das eigentlich? Herausforderungen und Entwicklungen der Preis-Leistungs-Analyse. [Web article]. [Accessed 26.03.2021]. Available: https://blog.vdi.de/2018/11/performance-pricing/

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society, Series B, Volume 36(2). pp. 111-133.

Uranie, 2018. Our Products. [Web page]. [Accessed 10.04.2021]. Available: https://www.uranie-inter.com/en/chromed-bar-chromed-tubes/.

Valveny, N. & Gilliver, S., 2016. How to interpret and report the results from multivariable analyses. Medical Writing, Volume 25(3). pp. 37-42.

VDI 2817 – Part 1. 2018. Verein Deutscher Ingenieure. Performance pricing (PP) – Fundamentals and application. Düsseldorf. Verein Deutscher Ingenieure e.V. 44 p.

VDI 2817 – Part 2. 2020. Verein Deutscher Ingenieure. Performance pricing (PP) – Commodity Library. Düsseldorf. Verein Deutscher Ingenieure e.V. 94 p.

Wakkee, M., Hollestein, M. & Nijsten, T. 2014. Multivariable Analysis. Journal of Investigative Dermatology. Volume 134(5). pp. 1-5.

Watton, J. 2009. Fundamentals of Fluid Power Control. New York: Cambridge University Press. 509 p.

Welc, J. & Rodriquez, P. J. 2018. Applied Regression Analysis for Business: Tools, Traps and Applications. Springer International Publishing. 286 p.

Wold, H. 1982. Soft Modeling: The Basic Design and Some Extensions. In K. G. Jöreskog and H. Wold (Eds.), Systems under indirect observation, Volume 2. pp. 1-54.

Wold, S., Sjöström, M. & Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, Volume 58. pp. 109-130.

Yan, X. & Gang Su, X., 2009. Linear Regression Analysis - Theory and Computing. Singapore: World Scientific Publishing Co. Pte. Ltd. 348 p.

Zhou, M., Chen, Y. & Gao, T., 2013. A price forecast model of hydraulic cylinders through integration of case based reasoning and back-propagation neural networks. International Conference on Natural Computation (ICNC). pp. 111-115.

Summarized results of interviews.

Interview with Dr. Berz:

- In principle, performance parameters to be used as value drivers instead of general technical characteristics (e.g. weight). In case of hydraulic cylinder, performance parameters could be for instance cylinder force, lifetime, stroke or operational length.
- In terms of performance pricing, the weight of the product should not be the value driver, since it is not the property customer is willing to pay for. However, it is still pretty obvious that weight is one of the main cost drivers due to its impact on raw material consumption and manufacturing costs.
- Performance parameter approach would bring out the cost efficiency of suppliers.
- In general, the values of adjusted R^2 and Q^2 should serve as a good guide during model development. F and F critical to be used as supportive indicators.
- In case model includes pairs (left & right side cylinder) with different part numbers, either of these should be removed to avoid unnecessary strengthening of the regression.
- The statistical contribution of possible phase-out- items should be evaluated carefully, if those are included as a part of model.

Interviews with engineering:

Cylinder force

- The generally agreed hydraulic pressure used in tractors is typically around 200-210 bar. Everything is built around this (Pitkänen 2021). A possible pressure increase would require massive changes to the entire hydraulic system and additional pressure reducers for other implements, which is not feasible (Mankki 2021).
- The required cylinder force is determined by using efficiency factor 0,85, which takes into account pressure losses and mechanical friction losses (Pitkänen 2021).

Specifications and dimensioning

- Hydraulic cylinder is part of kinematic system, which partially set requirements for cylinder. Product management sets certain demands (e.g. lifting force) for the system, and the system will be built based on this. (Pitkänen 2021).

- Lifting force calculators are utilized in dimensioning for selected cylinder types (Pitkänen 2021, Karjalainen 2021).
- Cylinder specification done based on performance parameters and main dimensions (e.g. no specifications for wall thickness or seal kit) (Pitkänen 2021). With new products, the intention is to specify as little as possible. In general, design is mostly based on know-how of the suppliers (e.g. piston/rod assembly, snap-ring vs. screwed cylinder head). This allows suppliers to utilize cost efficient solutions (e.g. standard components). (Mankki 2021, Karjalainen 2021).

Cylinder types

- Design and dimensioning of selected cylinder type is done purely by supplier (Karjalainen 2021).
- Cylinder type 4 is the most different in terms of functionality and dynamics (Meierhofer 2021).
- ABC1234582 cylinder probably comparable to other type 4 cylinders in the model (Jalonen 2021)
- Type 1- and 3 cylinders have different assembly position (pulling vs. pushing cylinders). This causes probably slight differences on sealing, but there is not any major differences. (Mankki 2021, Meierhofer 2021).

Performance parameters

- From engineering point of view, the most important performance parameter is that cylinder is leak-free. However, this is really challenging to measure based on numbers (Mankki 2021; Pitkänen 2021).
- Lifetime to be used as value driver? However, there is no way to measure this accurately. The durability is validated during laboratory & field tests (Mankki 2021; Pitkänen 2021).
- Removing grease nipples would offer small cost benefit, but in terms of quality impression and usability those are important (Mankki 2021).
- One value driver could be bearing (Y/N) or the quality of the bearing (Meierhofer 2021).

APPENDIX 2.

Correlation matrix based on R^2 (all cylinders 80 pcs).

	P	CL	ID	OD	WT	S	RD	W	NOP	NOB	NOG	OL	APS	APR	TESA	RESA	APSx S	APSx CL	APSx OL	TESAx S	TESAx CL	TESAx OL	RESAx S	RESAx CL	RESAx OL	IDx OL	RDx OL
P	1,00	0,77	0,04	0,06	0,14	0,60	0,09	0,83	0,00	0,41	0,15	0,76	0,03	0,01	0,11	0,04	0,50	0,69	0,66	0,60	0,80	0,78	0,49	0,70	0,66	0,76	0,79
CL	0,77	1,00	0,02	0,02	0,03	0,73	0,00	0,86	0,00	0,34	0,24	0,97	0,01	0,01	0,03	0,00	0,48	0,76	0,70	0,59	0,88	0,83	0,45	0,69	0,63	0,88	0,87
ID	0,04	0,02	1,00	0,99	0,11	0,05	0,04	0,18	0,02	0,01	0,00	0,03	0,99	0,86	0,72	0,02	0,44	0,33	0,38	0,22	0,15	0,18	0,06	0,03	0,05	0,18	0,04
OD	0,06	0,02	0,99	1,00	0,19	0,06	0,05	0,20	0,02	0,01	0,00	0,04	0,97	0,83	0,81	0,03	0,46	0,34	0,39	0,26	0,17	0,21	0,08	0,04	0,06	0,19	0,05
WT	0,14	0,03	0,11	0,19	1,00	0,07	0,08	0,12	0,00	0,00	0,05	0,05	0,11	0,06	0,59	0,06	0,13	0,09	0,11	0,25	0,18	0,22	0,10	0,08	0,09	0,08	0,07
S	0,60	0,73	0,05	0,06	0,07	1,00	0,01	0,74	0,00	0,18	0,29	0,86	0,04	0,03	0,08	0,00	0,72	0,60	0,67	0,87	0,69	0,79	0,75	0,60	0,68	0,82	0,85
RD	0,09	0,00	0,04	0,05	0,08	0,01	1,00	0,06	0,30	0,04	0,00	0,00	0,03	0,02	0,07	0,95	0,02	0,01	0,02	0,02	0,01	0,02	0,26	0,32	0,32	0,01	0,10
W	0,83	0,86	0,18	0,20	0,12	0,74	0,06	1,00	0,01	0,27	0,19	0,88	0,15	0,10	0,21	0,03	0,74	0,92	0,91	0,78	0,95	0,94	0,60	0,76	0,74	0,96	0,89
NOP	0,00	0,00	0,02	0,02	0,00	0,00	0,30	0,01	1,00	0,09	0,01	0,00	0,02	0,01	0,01	0,50	0,00	0,00	0,00	0,00	0,00	0,00	0,09	0,12	0,11	0,00	0,01
NOB	0,41	0,34	0,01	0,01	0,00	0,18	0,04	0,27	0,09	1,00	0,00	0,30	0,01	0,01	0,00	0,00	0,09	0,20	0,17	0,12	0,25	0,21	0,12	0,26	0,22	0,24	0,31
NOG	0,15	0,24	0,00	0,00	0,05	0,29	0,00	0,19	0,01	0,00	1,00	0,27	0,00	0,00	0,01	0,00	0,13	0,12	0,13	0,22	0,20	0,22	0,21	0,18	0,20	0,21	0,24
OL	0,76	0,97	0,03	0,04	0,05	0,86	0,00	0,88	0,00	0,30	0,27	1,00	0,02	0,02	0,05	0,00	0,59	0,75	0,74	0,72	0,87	0,87	0,57	0,70	0,69	0,92	0,92
APS	0,03	0,01	0,99	0,97	0,11	0,04	0,03	0,15	0,02	0,01	0,00	0,02	1,00	0,88	0,72	0,02	0,41	0,30	0,35	0,19	0,12	0,15	0,05	0,02	0,03	0,15	0,03
APR	0,01	0,01	0,86	0,83	0,06	0,03	0,02	0,10	0,01	0,01	0,00	0,02	0,88	1,00	0,57	0,05	0,37	0,27	0,32	0,17	0,10	0,13	0,00	0,00	0,00	0,13	0,01
TESA	0,11	0,03	0,72	0,81	0,59	0,08	0,07	0,21	0,01	0,00	0,01	0,05	0,72	0,57	1,00	0,05	0,41	0,29	0,35	0,34	0,23	0,28	0,10	0,06	0,08	0,18	0,07
RESA	0,04	0,00	0,02	0,03	0,06	0,00	0,95	0,03	0,50	0,00	0,00	0,00	0,02	0,05	0,05	1,00	0,00	0,00	0,00	0,01	0,00	0,00	0,22	0,29	0,28	0,00	0,06
APSxS	0,50	0,48	0,44	0,46	0,13	0,72	0,02	0,74	0,00	0,09	0,13	0,59	0,41	0,37	0,41	0,00	1,00	0,79	0,90	0,89	0,68	0,79	0,59	0,44	0,52	0,80	0,62
APSxCL	0,69	0,76	0,33	0,34	0,09	0,60	0,01	0,92	0,00	0,20	0,12	0,75	0,30	0,27	0,29	0,00	0,79	1,00	0,98	0,71	0,91	0,89	0,40	0,55	0,52	0,93	0,70
APSxOL	0,66	0,70	0,38	0,39	0,11	0,67	0,02	0,91	0,00	0,17	0,13	0,74	0,35	0,32	0,35	0,00	0,90	0,98	1,00	0,81	0,88	0,90	0,49	0,54	0,55	0,94	0,71
TESAxS	0,60	0,59	0,22	0,26	0,25	0,87	0,02	0,78	0,00	0,12	0,22	0,72	0,19	0,17	0,34	0,01	0,89	0,71	0,81	1,00	0,77	0,89	0,70	0,53	0,62	0,82	0,75
TESAxCL	0,80	0,88	0,15	0,17	0,18	0,69	0,01	0,95	0,00	0,25	0,20	0,87	0,12	0,10	0,23	0,00	0,68	0,91	0,88	0,77	1,00	0,97	0,46	0,65	0,61	0,94	0,81
TESAxOL	0,78	0,83	0,18	0,21	0,22	0,79	0,02	0,94	0,00	0,21	0,22	0,87	0,15	0,13	0,28	0,00	0,79	0,89	0,90	0,89	0,97	1,00	0,57	0,64	0,65	0,95	0,84
RESAxS	0,49	0,45	0,06	0,08	0,10	0,75	0,26	0,60	0,09	0,12	0,21	0,57	0,05	0,00	0,10	0,22	0,59	0,40	0,49	0,70	0,46	0,57	1,00	0,79	0,91	0,57	0,79
RESAxCL	0,70	0,69	0,03	0,04	0,08	0,60	0,32	0,76	0,12	0,26	0,18	0,70	0,02	0,00	0,06	0,29	0,44	0,55	0,54	0,53	0,65	0,64	0,79	1,00	0,97	0,67	0,89
RESAxOL	0,66	0,63	0,05	0,06	0,09	0,68	0,32	0,74	0,11	0,22	0,20	0,69	0,03	0,00	0,08	0,28	0,52	0,52	0,55	0,62	0,61	0,65	0,91	0,97	1,00	0,66	0,90
IDxOL	0,76	0,88	0,18	0,19	0,08	0,82	0,01	0,96	0,00	0,24	0,21	0,92	0,15	0,13	0,18	0,00	0,80	0,93	0,94	0,82	0,94	0,95	0,57	0,67	0,66	1,00	0,87
RDxOL	0,79	0,87	0,04	0,05	0,07	0,85	0,10	0,89	0,01	0,31	0,24	0,92	0,03	0,01	0,07	0,06	0,62	0,70	0,71	0,75	0,81	0,84	0,79	0,89	0,90	0,87	1,00

Correlation matrix based on R^2 (active and NPI cylinders 62 pcs)

	P	CL	ID	OD	WT	S	RD	W	NOP	NOB	NOG	OL	APS	APR	TESA	RESA	APSx S	APSx CL	APSx OL	TESAx S	TESAx CL	TESAx OL	RESAx S	RESAx CL	RESAx OL	IDx OL	RDx OL
P	1,00	0,72	0,05	0,07	0,18	0,58	0,16	0,79	0,00	0,34	0,12	0,72	0,04	0,01	0,15	0,09	0,51	0,61	0,60	0,60	0,75	0,73	0,55	0,69	0,67	0,71	0,79
CL	0,72	1,00	0,04	0,04	0,05	0,72	0,01	0,83	0,00	0,23	0,18	0,96	0,02	0,02	0,06	0,00	0,51	0,73	0,68	0,61	0,87	0,81	0,46	0,62	0,58	0,87	0,84
ID	0,05	0,04	1,00	0,99	0,08	0,06	0,09	0,25	0,02	0,00	0,00	0,05	0,98	0,83	0,71	0,06	0,44	0,41	0,44	0,21	0,19	0,21	0,09	0,08	0,09	0,23	0,08
OD	0,07	0,04	0,99	1,00	0,15	0,08	0,11	0,27	0,03	0,00	0,00	0,06	0,97	0,80	0,80	0,07	0,47	0,43	0,46	0,26	0,23	0,25	0,12	0,10	0,11	0,25	0,10
WT	0,18	0,05	0,08	0,15	1,00	0,10	0,13	0,17	0,01	0,00	0,10	0,07	0,08	0,02	0,55	0,10	0,14	0,10	0,12	0,28	0,22	0,25	0,16	0,13	0,15	0,10	0,12
S	0,58	0,72	0,06	0,08	0,10	1,00	0,02	0,73	0,00	0,12	0,23	0,87	0,05	0,04	0,12	0,00	0,74	0,56	0,65	0,88	0,69	0,80	0,75	0,56	0,66	0,81	0,85
RD	0,16	0,01	0,09	0,11	0,13	0,02	1,00	0,13	0,34	0,04	0,00	0,01	0,08	0,01	0,16	0,94	0,06	0,05	0,05	0,06	0,05	0,05	0,32	0,44	0,41	0,04	0,16
W	0,79	0,83	0,25	0,27	0,17	0,73	0,13	1,00	0,01	0,17	0,14	0,85	0,21	0,13	0,32	0,08	0,78	0,90	0,90	0,80	0,94	0,94	0,66	0,77	0,76	0,95	0,89
NOP	0,00	0,00	0,02	0,03	0,01	0,00	0,34	0,01	1,00	0,11	0,02	0,00	0,03	0,02	0,02	0,56	0,00	0,00	0,00	0,00	0,00	0,00	0,10	0,16	0,15	0,00	0,01
NOB	0,34	0,23	0,00	0,00	0,00	0,12	0,04	0,17	0,11	1,00	0,03	0,20	0,01	0,01	0,00	0,00	0,06	0,12	0,10	0,09	0,16	0,14	0,08	0,16	0,13	0,16	0,21
NOG	0,12	0,18	0,00	0,00	0,10	0,23	0,00	0,14	0,02	0,03	1,00	0,21	0,00	0,01	0,03	0,00	0,10	0,08	0,09	0,18	0,16	0,18	0,17	0,14	0,16	0,15	0,19
OL	0,72	0,96	0,05	0,06	0,07	0,87	0,01	0,85	0,00	0,20	0,21	1,00	0,03	0,03	0,09	0,00	0,63	0,72	0,72	0,75	0,86	0,87	0,59	0,64	0,65	0,91	0,91
APS	0,04	0,02	0,98	0,97	0,08	0,05	0,08	0,21	0,03	0,01	0,00	0,03	1,00	0,86	0,71	0,05	0,42	0,38	0,41	0,19	0,16	0,18	0,07	0,06	0,07	0,20	0,06
APR	0,01	0,02	0,83	0,80	0,02	0,04	0,01	0,13	0,02	0,01	0,01	0,03	0,86	1,00	0,52	0,02	0,35	0,33	0,36	0,14	0,13	0,14	0,01	0,00	0,00	0,17	0,01
TESA	0,15	0,06	0,71	0,80	0,55	0,12	0,16	0,32	0,02	0,00	0,03	0,09	0,71	0,52	1,00	0,11	0,45	0,38	0,42	0,37	0,31	0,35	0,18	0,15	0,17	0,25	0,14
RESA	0,09	0,00	0,06	0,07	0,10	0,00	0,94	0,08	0,56	0,00	0,00	0,00	0,05	0,02	0,11	1,00	0,02	0,02	0,02	0,02	0,02	0,02	0,28	0,40	0,37	0,01	0,11
APSxS	0,51	0,51	0,44	0,47	0,14	0,74	0,06	0,78	0,00	0,06	0,10	0,63	0,42	0,35	0,45	0,02	1,00	0,80	0,91	0,89	0,71	0,81	0,64	0,48	0,56	0,84	0,68
APSxCL	0,61	0,73	0,41	0,43	0,10	0,56	0,05	0,90	0,00	0,12	0,08	0,72	0,38	0,33	0,38	0,02	0,80	1,00	0,97	0,68	0,89	0,86	0,42	0,53	0,51	0,92	0,68
APSxOL	0,60	0,68	0,44	0,46	0,12	0,65	0,05	0,90	0,00	0,10	0,09	0,72	0,41	0,36	0,42	0,02	0,91	0,97	1,00	0,79	0,87	0,89	0,52	0,54	0,56	0,93	0,71
TESAxS	0,60	0,61	0,21	0,26	0,28	0,88	0,06	0,80	0,00	0,09	0,18	0,75	0,19	0,14	0,37	0,02	0,89	0,68	0,79	1,00	0,78	0,90	0,74	0,55	0,65	0,83	0,79
TESAxCL	0,75	0,87	0,19	0,23	0,22	0,69	0,05	0,94	0,00	0,16	0,16	0,86	0,16	0,13	0,31	0,02	0,71	0,89	0,87	0,78	1,00	0,97	0,50	0,63	0,61	0,93	0,81
TESAxOL	0,73	0,81	0,21	0,25	0,25	0,80	0,05	0,94	0,00	0,14	0,18	0,87	0,18	0,14	0,35	0,02	0,81	0,86	0,89	0,90	0,97	1,00	0,62	0,64	0,66	0,95	0,85
RESAxS	0,55	0,46	0,09	0,12	0,16	0,75	0,32	0,66	0,10	0,08	0,17	0,59	0,07	0,01	0,18	0,28	0,64	0,42	0,52	0,74	0,50	0,62	1,00	0,82	0,93	0,60	0,83
RESAxCL	0,69	0,62	0,08	0,10	0,13	0,56	0,44	0,77	0,16	0,16	0,14	0,64	0,06	0,00	0,15	0,40	0,48	0,53	0,54	0,55	0,63	0,64	0,82	1,00	0,97	0,64	0,87
RESAxOL	0,67	0,58	0,09	0,11	0,15	0,66	0,41	0,76	0,15	0,13	0,16	0,65	0,07	0,00	0,17	0,37	0,56	0,51	0,56	0,65	0,61	0,66	0,93	0,97	1,00	0,65	0,89
IDxOL	0,71	0,87	0,23	0,25	0,10	0,81	0,04	0,95	0,00	0,16	0,15	0,91	0,20	0,17	0,25	0,01	0,84	0,92	0,93	0,83	0,93	0,95	0,60	0,64	0,65	1,00	0,87
RDxOL	0,79	0,84	0,08	0,10	0,12	0,85	0,16	0,89	0,01	0,21	0,19	0,91	0,06	0,01	0,14	0,11	0,68	0,68	0,71	0,79	0,81	0,85	0,83	0,87	0,89	0,87	1,00

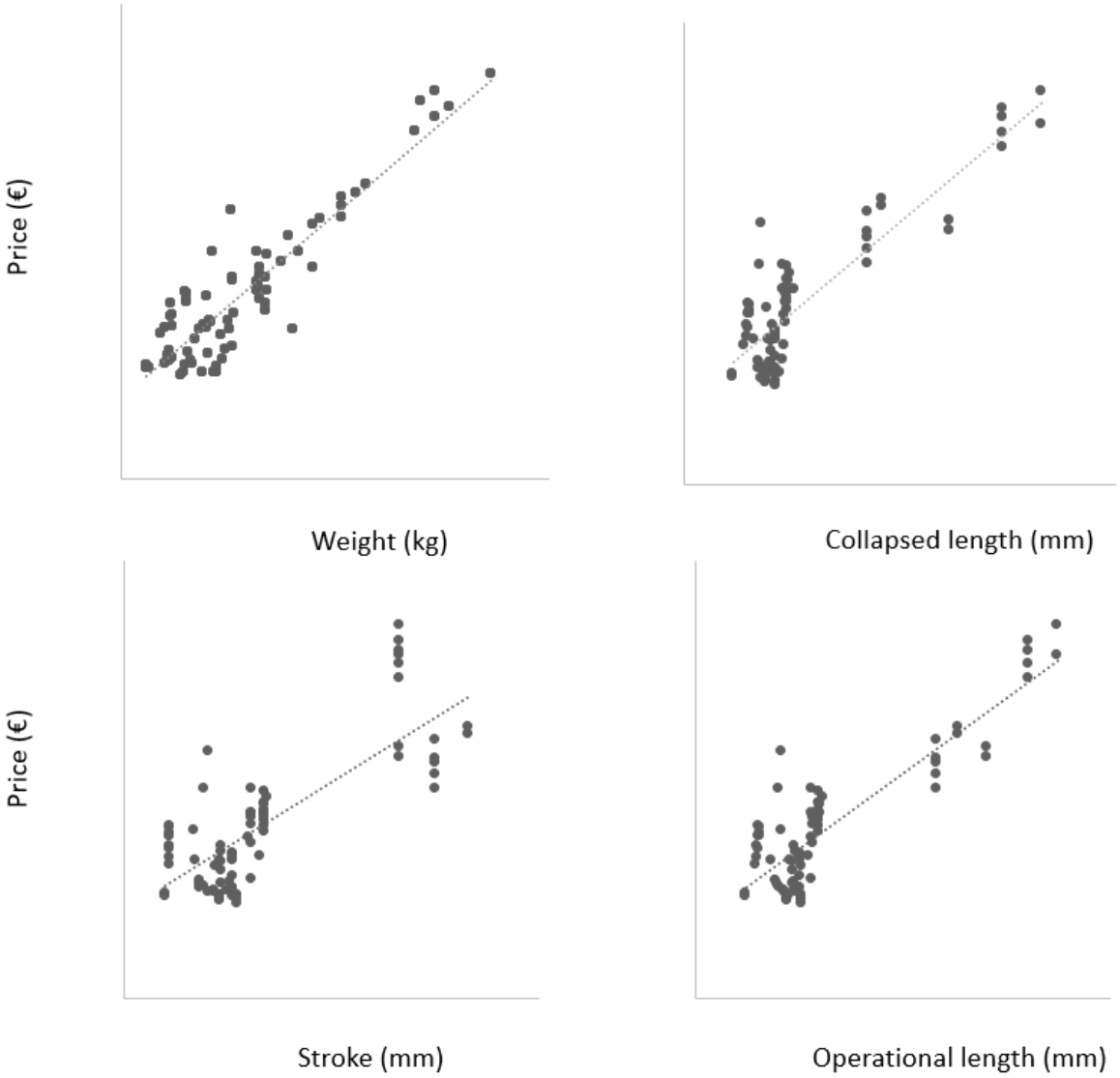
APPENDIX 3.

Change in the degree of explanation of the correlation (%) with two different samples.

	P	CL	ID	OD	WT	S	RD	W	NOP	NOB	NOG	OL	APS	APR	TESA	RESA	APSxS	APSxCL	APSxOL	TESAxS	TESAxCL	TESAxOL	RESAxS	RESAxCL	RESAxOL	IDxOL	RDxOL
P	0%	5%	-1%	-2%	-4%	1%	-7%	4%	0%	8%	3%	4%	-1%	0%	-4%	-5%	-1%	8%	5%	0%	5%	4%	-6%	1%	-1%	5%	0%
CL	5%	0%	-2%	-2%	-2%	0%	-1%	4%	0%	11%	6%	1%	-1%	-1%	-3%	0%	-3%	3%	2%	-1%	1%	1%	-1%	7%	5%	1%	3%
ID	-1%	-2%	0%	0%	3%	-1%	-5%	-7%	-1%	0%	0%	-2%	0%	2%	1%	-4%	0%	-9%	-6%	1%	-4%	-3%	-3%	-5%	-4%	-5%	-4%
OD	-2%	-2%	0%	0%	4%	-2%	-6%	-8%	-1%	0%	0%	-2%	0%	3%	1%	-4%	-1%	-9%	-7%	0%	-5%	-4%	-4%	-6%	-5%	-6%	-4%
WT	-4%	-2%	3%	4%	0%	-3%	-5%	-4%	0%	0%	-5%	-3%	3%	3%	4%	-4%	-1%	-1%	-1%	-2%	-4%	-4%	-6%	-6%	-6%	-2%	-5%
S	1%	0%	-1%	-2%	-3%	0%	-1%	1%	0%	6%	6%	-1%	-1%	0%	-4%	0%	-1%	4%	2%	-2%	1%	-1%	0%	4%	3%	1%	0%
RD	-7%	-1%	-5%	-6%	-5%	-1%	0%	-7%	-4%	0%	0%	-1%	-5%	2%	-8%	0%	-4%	-3%	-4%	-3%	-3%	-3%	-6%	-11%	-9%	-3%	-6%
W	4%	4%	-7%	-8%	-4%	1%	-7%	0%	-1%	10%	5%	3%	-6%	-3%	-10%	-5%	-4%	2%	0%	-2%	1%	1%	-6%	-1%	-2%	1%	-1%
NOP	0%	0%	-1%	-1%	0%	0%	-4%	-1%	0%	-3%	0%	0%	-1%	0%	-1%	-6%	0%	0%	0%	0%	0%	0%	-1%	-5%	-3%	0%	-1%
NOB	8%	11%	0%	0%	0%	6%	0%	10%	-3%	0%	-3%	10%	0%	0%	0%	0%	3%	8%	6%	3%	8%	7%	4%	11%	9%	9%	10%
NOG	3%	6%	0%	0%	-5%	6%	0%	5%	0%	-3%	0%	6%	0%	0%	-2%	0%	3%	5%	4%	3%	4%	4%	4%	4%	4%	5%	5%
OL	4%	1%	-2%	-2%	-3%	-1%	-1%	3%	0%	10%	6%	0%	-2%	-1%	-4%	0%	-4%	4%	2%	-3%	1%	0%	-2%	6%	4%	1%	2%
APS	-1%	-1%	0%	0%	3%	-1%	-5%	-6%	-1%	0%	0%	-2%	0%	2%	1%	-3%	0%	-9%	-7%	1%	-4%	-3%	-3%	-4%	-4%	-5%	-3%
APR	0%	-1%	2%	3%	3%	0%	2%	-3%	0%	0%	0%	-1%	2%	0%	5%	2%	2%	-6%	-4%	2%	-2%	-1%	0%	0%	0%	-3%	-1%
TESA	-4%	-3%	1%	1%	4%	-4%	-8%	-10%	-1%	0%	-2%	-4%	1%	5%	0%	-7%	-4%	-9%	-8%	-3%	-8%	-7%	-8%	-9%	-9%	-7%	-7%
RESA	-5%	0%	-4%	-4%	-4%	0%	0%	-5%	-6%	0%	0%	0%	-3%	2%	-7%	0%	-2%	-2%	-2%	-2%	-2%	-2%	-6%	-11%	-9%	-1%	-5%
APSxS	-1%	-3%	0%	-1%	-1%	-1%	-4%	-4%	0%	3%	3%	-4%	0%	2%	-4%	-2%	0%	-2%	-1%	0%	-3%	-3%	-5%	-4%	-5%	-4%	-6%
APSxCL	8%	3%	-9%	-9%	-1%	4%	-3%	2%	0%	8%	5%	4%	-9%	-6%	-9%	-2%	-2%	0%	0%	3%	2%	3%	-2%	2%	1%	1%	2%
APSxOL	5%	2%	-6%	-7%	-1%	2%	-4%	0%	0%	6%	4%	2%	-7%	-4%	-8%	-2%	-1%	0%	0%	2%	1%	1%	-3%	0%	-1%	0%	0%
TESAxS	0%	-1%	1%	0%	-2%	-2%	-3%	-2%	0%	3%	3%	-3%	1%	2%	-3%	-2%	0%	3%	2%	0%	-1%	-1%	-4%	-2%	-3%	-1%	-4%
TESAxCL	5%	1%	-4%	-5%	-4%	1%	-3%	1%	0%	8%	4%	1%	-4%	-2%	-8%	-2%	-3%	2%	1%	-1%	0%	0%	-4%	2%	0%	1%	0%
TESAxOL	4%	1%	-3%	-4%	-4%	-1%	-3%	1%	0%	7%	4%	0%	-3%	-1%	-7%	-2%	-3%	3%	1%	-1%	0%	0%	-5%	1%	-1%	1%	-1%
RESAxS	-6%	-1%	-3%	-4%	-6%	0%	-6%	-6%	-1%	4%	4%	-2%	-3%	0%	-8%	-6%	-5%	-2%	-3%	-4%	-4%	-5%	0%	-3%	-2%	-3%	-4%
RESAxCL	1%	7%	-5%	-6%	-6%	4%	-11%	-1%	-5%	11%	4%	6%	-4%	0%	-9%	-11%	-4%	2%	0%	-2%	2%	1%	-3%	0%	0%	3%	2%
RESAxOL	-1%	5%	-4%	-5%	-6%	3%	-9%	-2%	-3%	9%	4%	4%	-4%	0%	-9%	-9%	-5%	1%	-1%	-3%	0%	-1%	-2%	0%	0%	1%	1%
IDxOL	5%	1%	-5%	-6%	-2%	1%	-3%	1%	0%	9%	5%	1%	-5%	-3%	-7%	-1%	-4%	1%	0%	-1%	1%	1%	-3%	3%	1%	0%	0%
RDxOL	0%	3%	-4%	-4%	-5%	0%	-6%	-1%	-1%	10%	5%	2%	-3%	-1%	-7%	-5%	-6%	2%	0%	-4%	0%	-1%	-4%	2%	1%	0%	0%

APPENDIX 4.

Scatter plots for potential value drivers.



APPENDIX 5.

Value driver group 1 (Weight + Cylinder types).

Group	1. Weight + Cylinder types			
Number	1.1	1.2 (matrix input for weight)	1.3	1.4 (matrix input for weight)
Status	ACTIVE + NPI + PHASED-OUT		ACTIVE + NPI	
Data points	78	78	62	62
Value drivers	Weight + All cylinder types as separate			
Number of value drivers	5	4	5	4
Latent variables	4	3	3	4
Technical value	35.65 + 4.22 * Weight / [kg] + 0 * Type 3 + 34.69 * Type 4 + 4.68 * Type 1 + 9.23 * Type 2	44.48 + 3.73 * Type 1 / [kg] + 4.23 * Type 2 / [kg] + 3.73 * Type 3 / [kg] + 7.53 * Type 4 / [kg]	38.19 + 3.85 * Weight / [kg] + 0 * Type 3 + 34.25 * Type 4 + 5.97 * Type 1 + 17.99 * Type 2	48.1 + 3.41 * Type 1 / [kg] + 4.09 * Type 2 / [kg] + 3.27 * Type 3 / [kg] + 7.03 * Type 4 / [kg]
Adjusted R2	0,94	0,94	0,94	0,95
Q2	0,94	0,93	0,94	0,94
F	230,9	289,67	197,94	257,27
F critical	4,06	4,49	4,21	4,64
Sign F	0	0	0	0
Comments	Type 3--> Type 1 --> Type 2 --> Type 4	Type 3--> Type 1 --> Type 2 --> Type 4	Type 3--> Type 1 --> Type 2 --> Type 4	Type 3--> Type 1 --> Type 2 --> Type 4
Next action	No actions	No actions	No actions	Validation
Removed data points	3	3	3	3

Group	1. Weight + Cylinder types			
Number	1.5	1.6 (matrix input for weight)	1.7	1.8 (matrix input for weight)
Status	ACTIVE + NPI + PHASED-OUT		ACTIVE + NPI	
Data points	78	78	62	62
Value drivers	Weight + (Type 1 and 3) + Type 4 + Type 2			
Number of value drivers	4	3	4	3
Latent variables	3	2	2	3
Technical value	38.87 + 4.10 * Weight / [kg] + 0 * (Type 1 and 3) + 32.35 * Type 4 + 10.19 * Type 2	44.46 + 4.23 * Type 2 / [kg] + 3.73 * (Type 1 and 3)) / [kg] + 7.53 * Type 4 / [kg]	44.2 + 3.58 * Weight / [kg] + 0 * (Type 1 and 3) + 29.72 * Type 4 + 21.37 * Type 2	48,63 + 4,07 * Type 2 / [kg] + 3,27 * (Type 1 and 3)) / [kg] + 7,04 * Type 4 / [kg]
Adjusted R2	0,94	0,94	0,94	0,95
Q2	0,94	0,94	0,92	0,94
F	286,32	391,59	237,47	356,71
F critical	4,49	5,17	4,64	5,32
Sign F	0	0	0	0
Comments	(Type 1 and 3) --> Type 2 --> Type 4	(Type 1 and 3) --> Type 2 --> Type 4	(Type 1 and 3) --> Type 2 --> Type 4	(Type 1 and 3) --> Type 2 --> Type 4
Next action	No actions	No actions	No actions	Validation
Removed data points	3	3	3	2

Group	1. Weight + Cylinder types			
Number	1.9	1.10 (matrix input for weight)	1.11	1.12 (matrix input for weight)
Status	ACTIVE + NPI + PHASED-OUT		ACTIVE + NPI	
Data points	78	78	62	62
Value drivers	Weight + (Type 1, 2 and 3) + Type 4			
Number of value drivers	3	2	3	2
Latent variables	2	2	2	2
Technical value	35.05 + 4.43 * Weight / [kg] + 0 * (Type 1, 2 and 3) + 33.73 * Type 4	37.30 + 4.34 * (Type 1, 2 and 3) / [kg] + 8.37 * Type 4 / [kg]	38.24 + 0.07 * (Type 1, 2 and 3) / [kg] + 4.08 * Type 4 / [kg] + 4.15 * Weight / [kg]	38.24 + 4.21 * (Type 1, 2 and 3) / [kg] + 8.22 * Type 4 / [kg]
Adjusted R2	0,94	0,93	0,93	0,93
Q2	0,93	0,93	0,93	0,93
F	371,74	554,77	267,11	407,81
F critical	5,17	6,41	5,32	6,57
Sign F	0	0	0	0
Comments	(Type 1, 2 and 3) --> Type 4	(Type 1, 2 and 3) --> Type 4	(Type 1, 2 and 3) --> Type 4. Matrix input leads into slightly better statistical results	(Type 1, 2 and 3) --> Type 4
Next action	No actions	No actions	No actions	Validation
Removed data points	3	2	2	2

APPENDIX 6.

Value driver group 2 (Weight + Technical characteristics).

Group	2. Weight + Technical characteristics			
Number	2.1	2.2	2.3	2.4
Status	ACTIVE + NPI + PHASED-OUT	ACTIVE + NPI	ACTIVE + NPI + PHASED-OUT	ACTIVE + NPI
Data points	78	62	78	62
Value drivers	Weight + Rod diameter + Operational length		Weight + Inner diameter + Operational length + Rod diameter	
Number of value drivers	3	3	4	4
Latent variables	2	3	4	4
Technical value	11.49 + 0.03 * Operational length / [mm] + 0.97 * Rod Diameter / [mm] + 2.51 * Weight / [kg]	5.06 + 0.04 * Operational length / [mm] + 1.19 * Rod Diameter / [mm] + 1.60 * Weight / [kg]	90.84 - 0.92 * Inner Diameter / [mm] - 0.05 * Operational length / [mm] + 0.71 * Rod Diameter / [mm] + 6.44 * Weight / [kg]	84.52 - 0.93 * Inner Diameter / [mm] - 0.04 * Operational length / [mm] + 0.87 * Rod Diameter / [mm] + 5.82 * Weight / [kg]
Adjusted R2	0,89	0,89	0,93	0,94
Q2	0,89	0,89	0,93	0,93
F	214,82	166,35	263,89	223,38
F critical	5,17	5,32	4,48	4,63
Sign F	0	0	0	0
Comments	-	-	(-) sign for ID.	(-) sign for ID and OL
Next action	No actions	Validation	Rejected	Rejected
Removed data points	3	3	2	2
General comment				

APPENDIX 7.

Value driver group 3 (Cylinder types + Technical characteristics).

Group	3. Cylinder types + Technical characteristics			
Number	3.1	3.2 (matrix input for OL)	3.3	3.4 (matrix input for OL)
Status	ACTIVE + NPI + PHASED-OUT		ACTIVE + NPI	
Data points	78	78	62	62
Value drivers	All cylinder types as separate + Inner diameter + Operational length			
Number of value drivers	6	5	6	5
Latent variables	4	5	4	5
Technical value	-34.14 + 0.85 * Inner Diameter / [mm] + 0.09 * Operational length / [mm] + 0 * Type 3 + 52.03 * Type 4 + 5.03 * Type 1 + 14.82 * Type 2	-41.97 + 0.15 * Type 1 Ol / [mm] + 0.11 * Type 2 OL / [mm] + 0.52 * Inner Diameter / [mm] + 0.15 * Type 3 OL / [mm] + 0.29 * Type 4 OL / [mm]	-39.53 + 0.67 * Inner Diameter / [mm] + 0.10 * Operational length / [mm] + 8.19 * Type 3 + 61.67 * Type 4 + 13.32 * Type 1 + 0 * Type 2	-24.01 + 0.11 * Type 1 Ol / [mm] + 0.10 * Type 2 OL / [mm] + 0.59 * Inner Diameter / [mm] + 0.10 * Type 3 OL / [mm] + 0.23 * Type 4 OL / [mm]
Adjusted R2	0,91	0,93	0,93	0,92
Q2	0,88	0,91	0,92	0,91
F	129,19	173,55	121,86	136,66
F critical	3,75	4,05	3,91	4,20
Sign F	0	0	0	0
Comments	Negative y-intercept	Negative y-intecept	Negative y-intercept.	Negative y-intercept
Next action	No actions	No actions	Validation	Validation
Removed data points	3	2	3	2
General comment	Negative y-intercept with all test models. Matrix input for RD (instead of OL) didn't give realistic results			

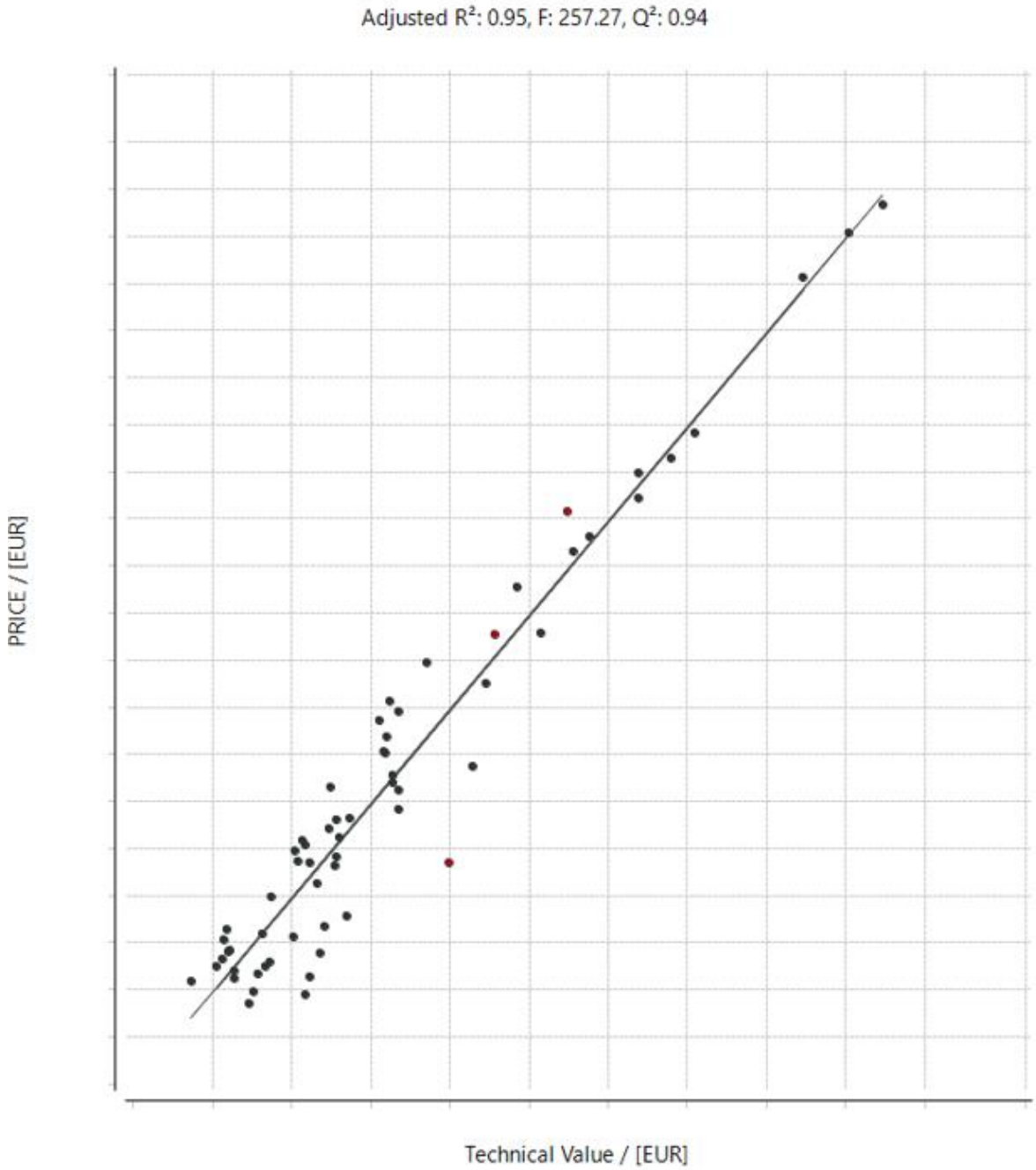
Value driver group 4 (Technical characteristics only).

Group	4. Technical characteristics only			
Number	4.1	4.2	4.3	4.4
	ACTIVE + NPI + PHASED-OUT	ACTIVE + NPI	ACTIVE + NPI + PHASED-OUT	ACTIVE + NPI
Data points	78	62	78	62
Value drivers	Operational length + Rod end section area + Tube end section area		Operational length + Rod diameter	
Number of value drivers	3	3	2	2
Latent variables	3	3	2	2
Technical value	11.37 + 0.08 * Operational length / [mm] + 0.02 * Rod end section area / [mm2] + 0.01 * Tube end section area / [mm2]	17.86 + 0.07 * Operational length / [mm] + 0.02 * Rod end section area / [mm2] + 0,00 * Tube end section area / [mm2]	-5.78 + 0.08 * Operational length / [mm] + 1.42 * Rod Diameter / [mm]	-8.83 + 0.07 * Operational length / [mm] + 1.56 * Rod Diameter / [mm]
Adjusted R2	0,87	0,88	0,86	0,88
Q2	0,86	0,87	0,86	0,87
F	167,54	143,50	240,28	219,18
F critical	5,17	5,33	6,41	6,59
Sign F	0	0	0	0
Comments	The value of coefficient for TESA is really low	The value of coefficient for TESA is really low (0,00228..)	Negative y-intercept	Negative y-intercept. If tube ID is chosen instead of RD, both adj. R^2 and Q^2 are below 0,8
Next action	Rejected	Rejected	No actions	Validation
Removed data points	3	3	3	3
Overall comment	Challenging to reach same statistical quality without type or weight. Combinations of tube ID/area and RD/area are not working (non-sensical technical values)			

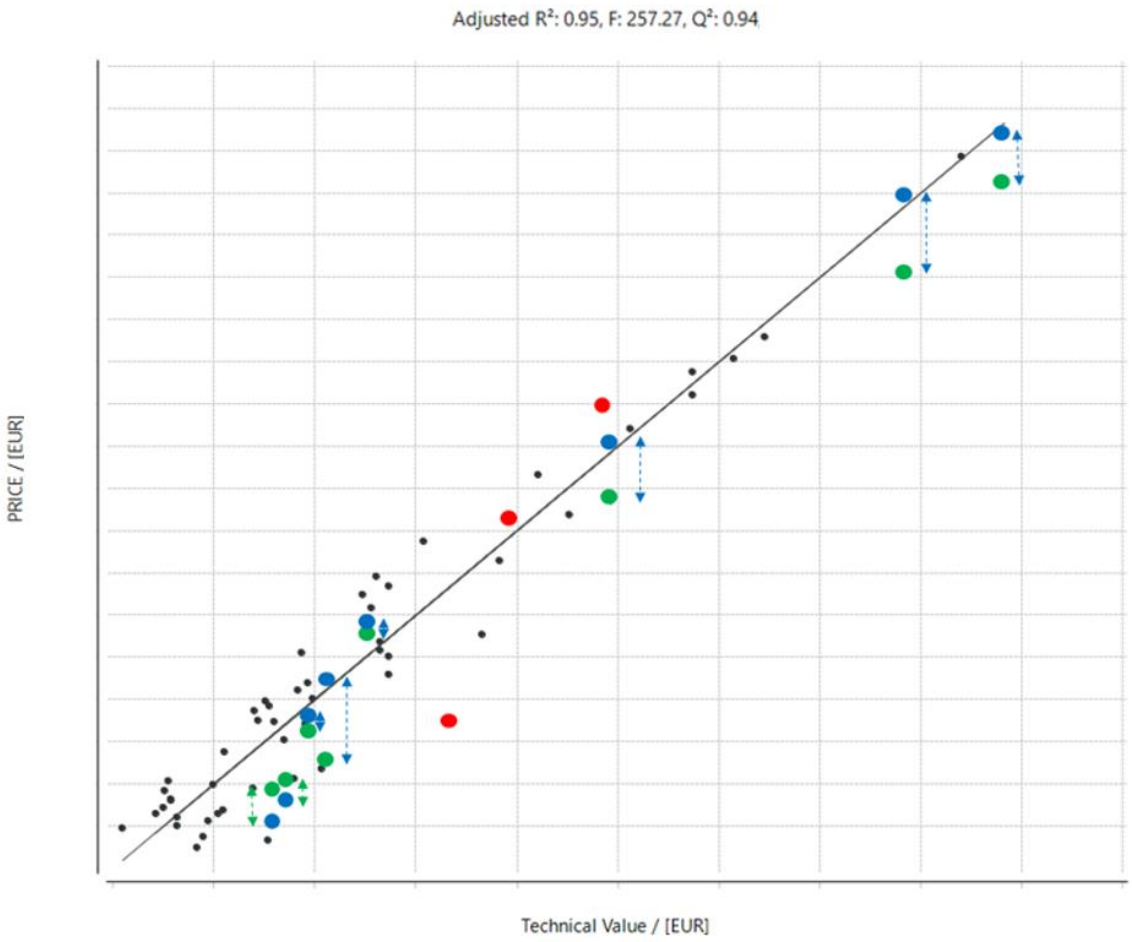
Scatter plots for different model options (including bottom-up calculations).

- Blue data points=Analyzed data points (supplier price)
- Green=Result of bottom-up calculation (shown in scatter plot, but not part of the regression)
- Red=Outliers (not part of the regression)

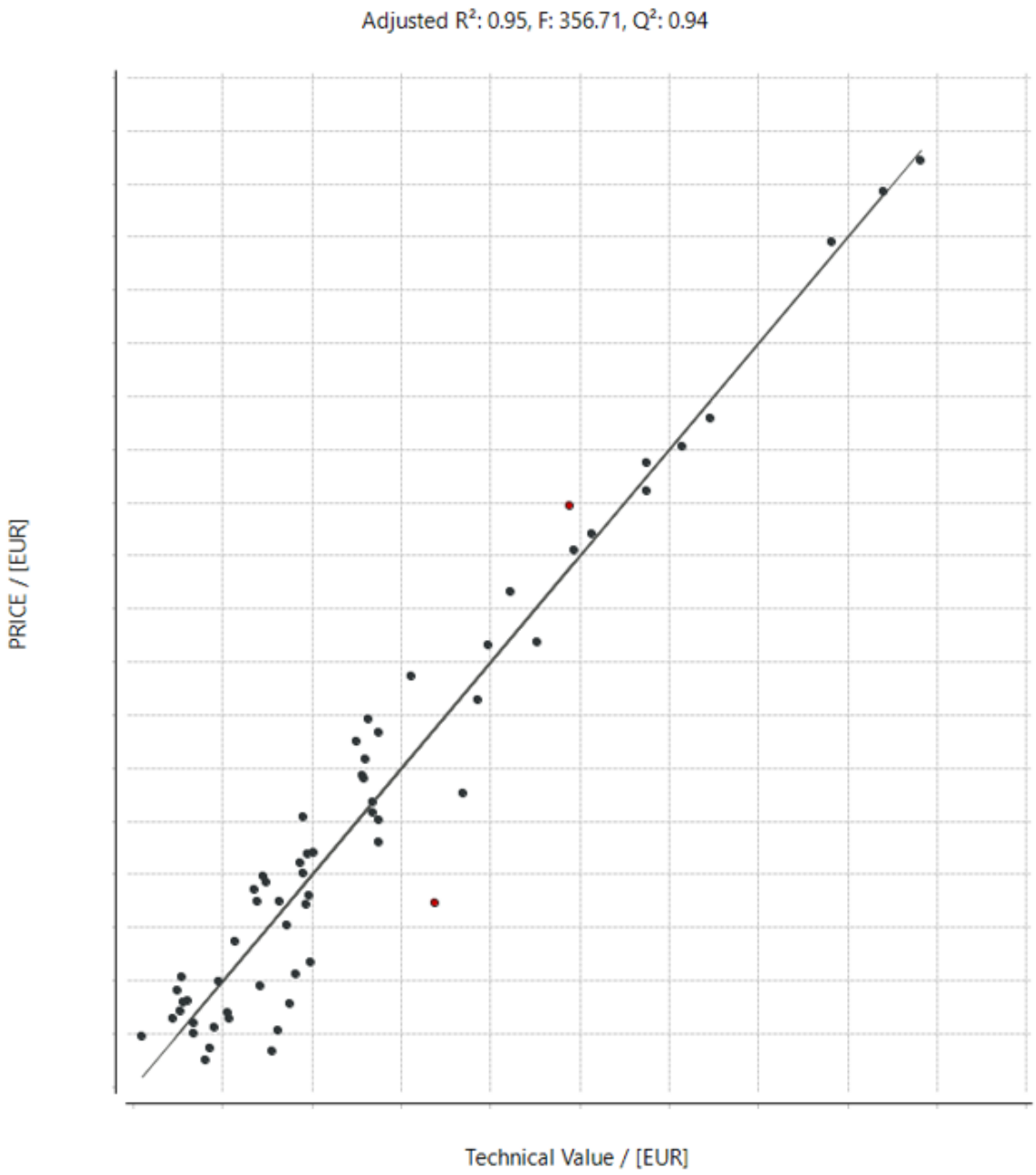
Model 1.4 (Weight as matrix input + All cylinder types as separate)



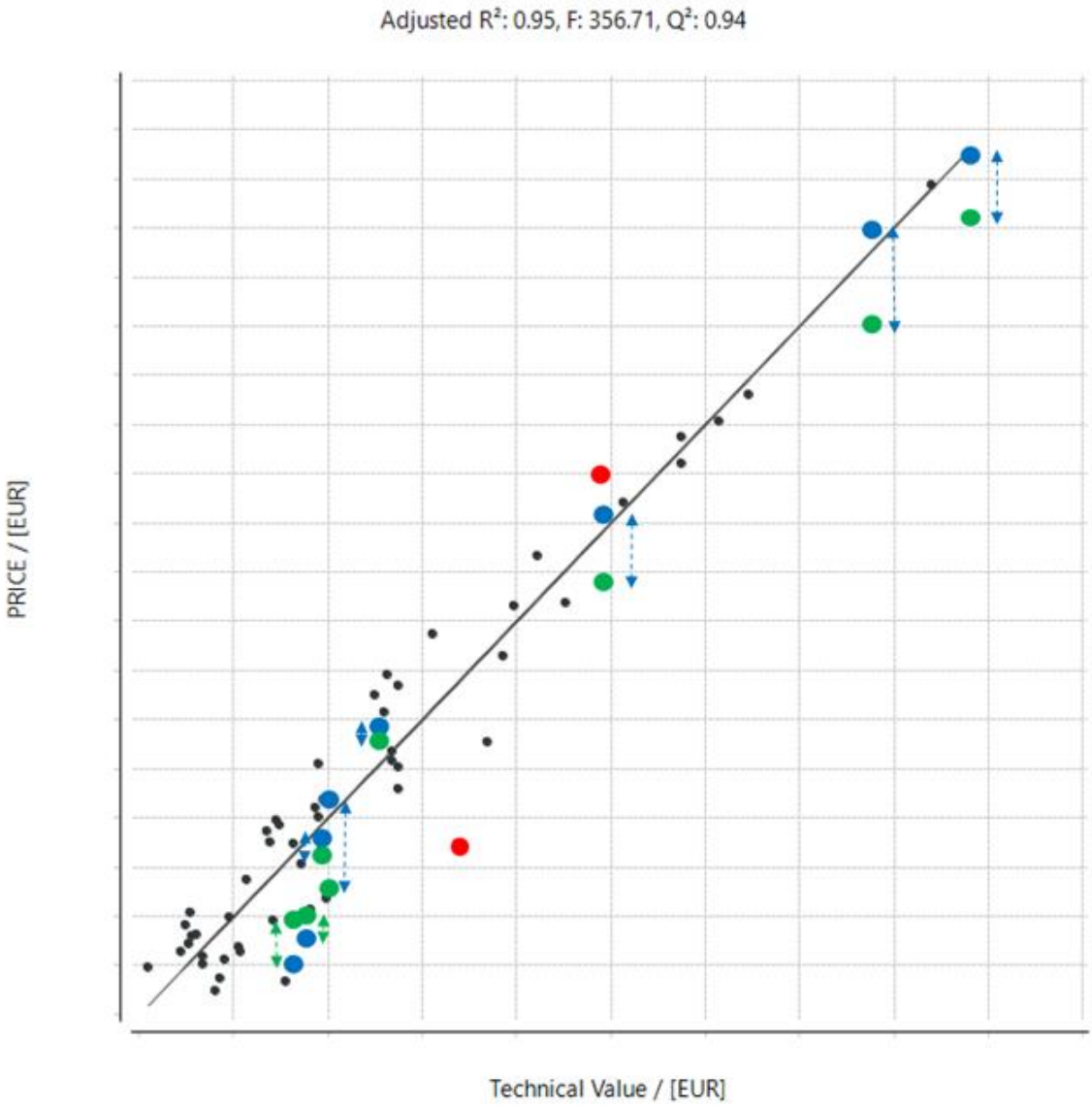
Model 1.4 with bottom-up calculations.



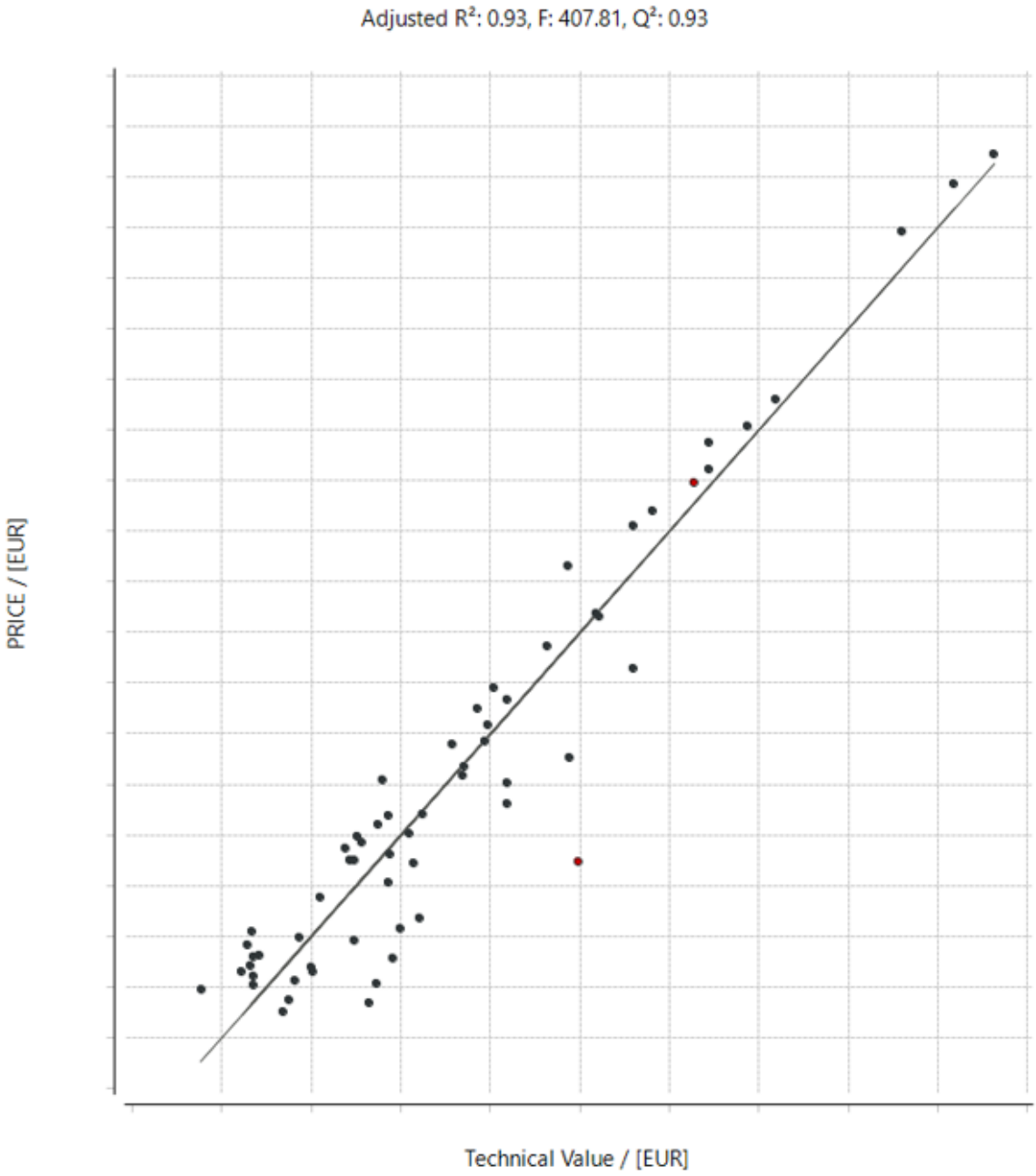
Model 1.8 (Weight as matrix input + (Type 1 and 3) + Type 4 + Type 2)



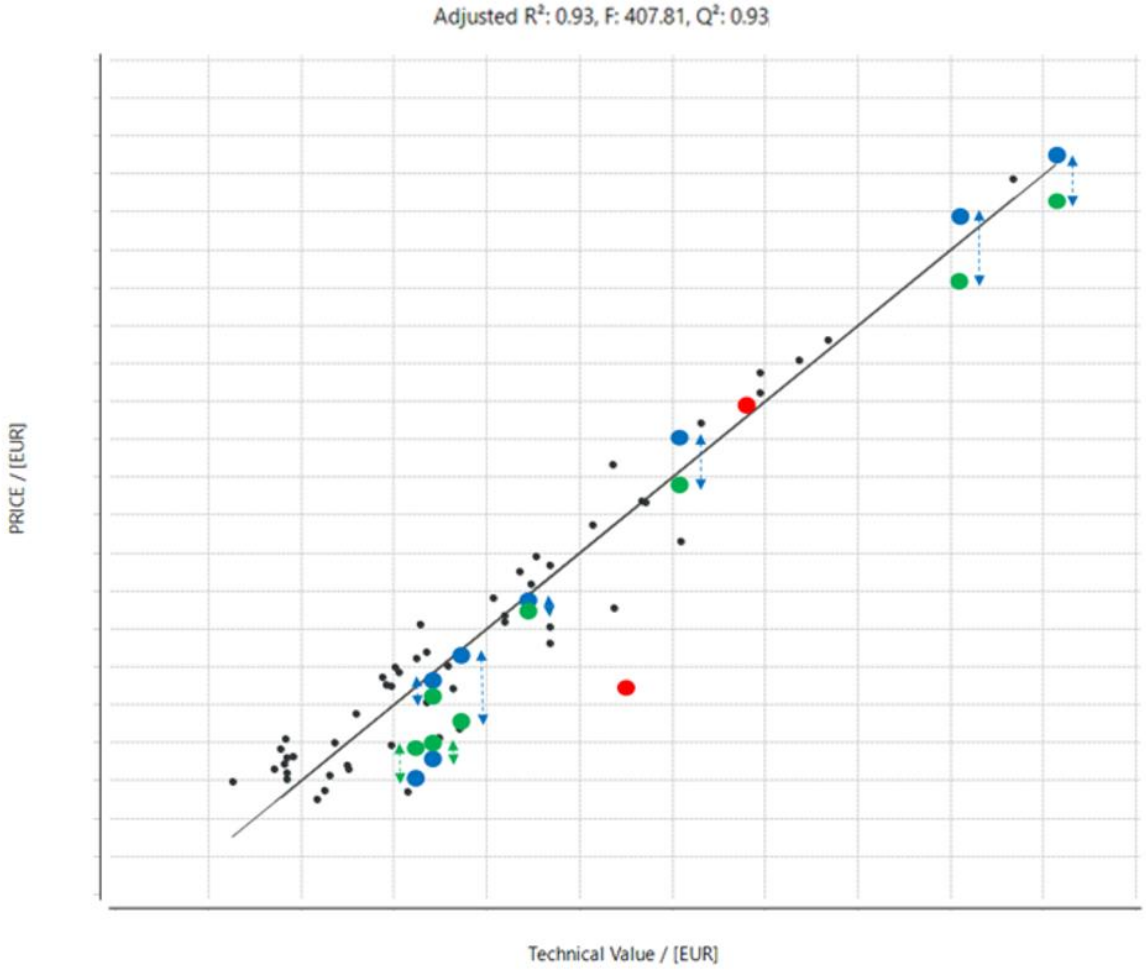
Model 1.8 with bottom-up calculations.



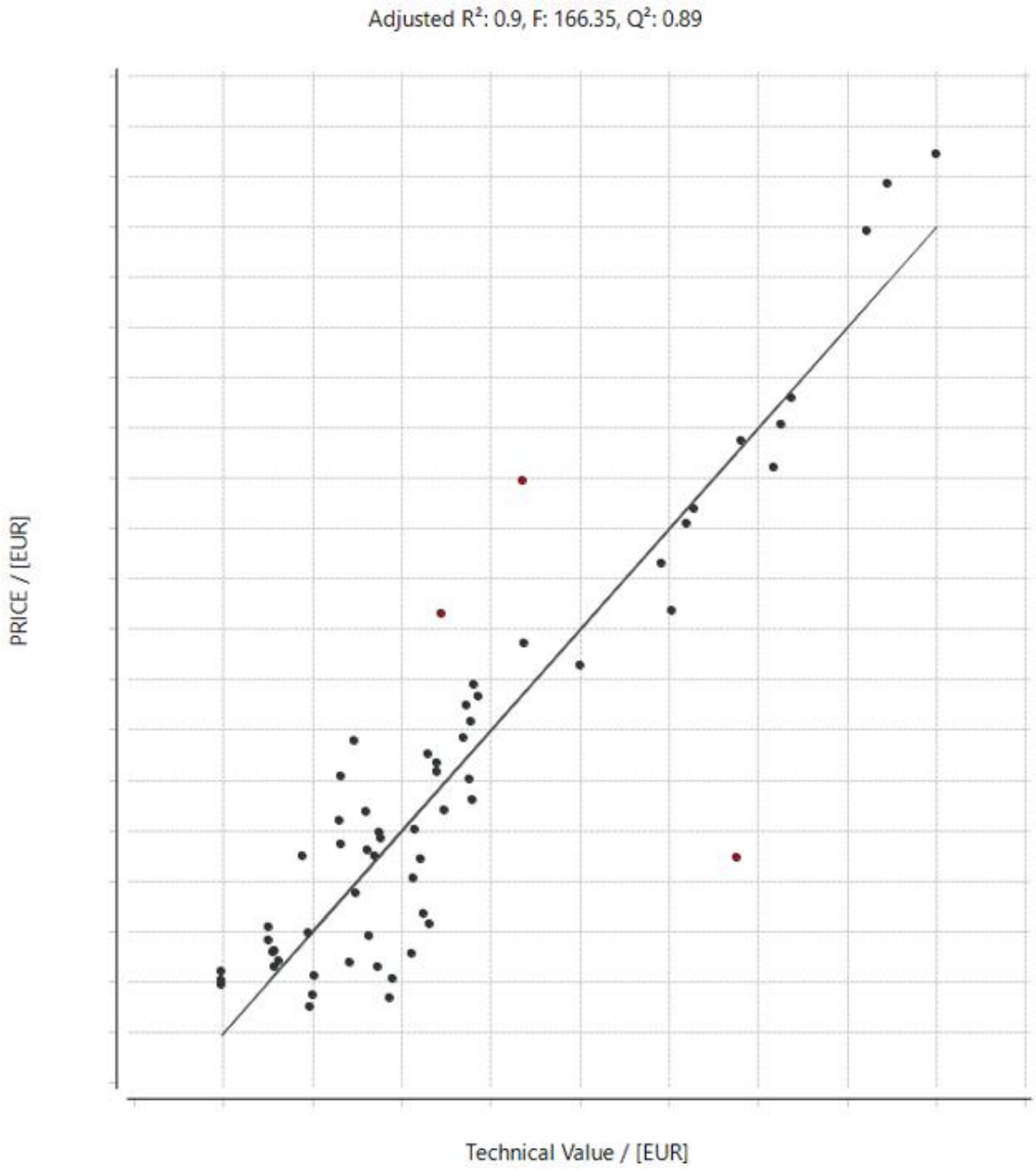
Model 1.12 (Weight as matrix input + (Type 1, 2 and 3) + Type 4)



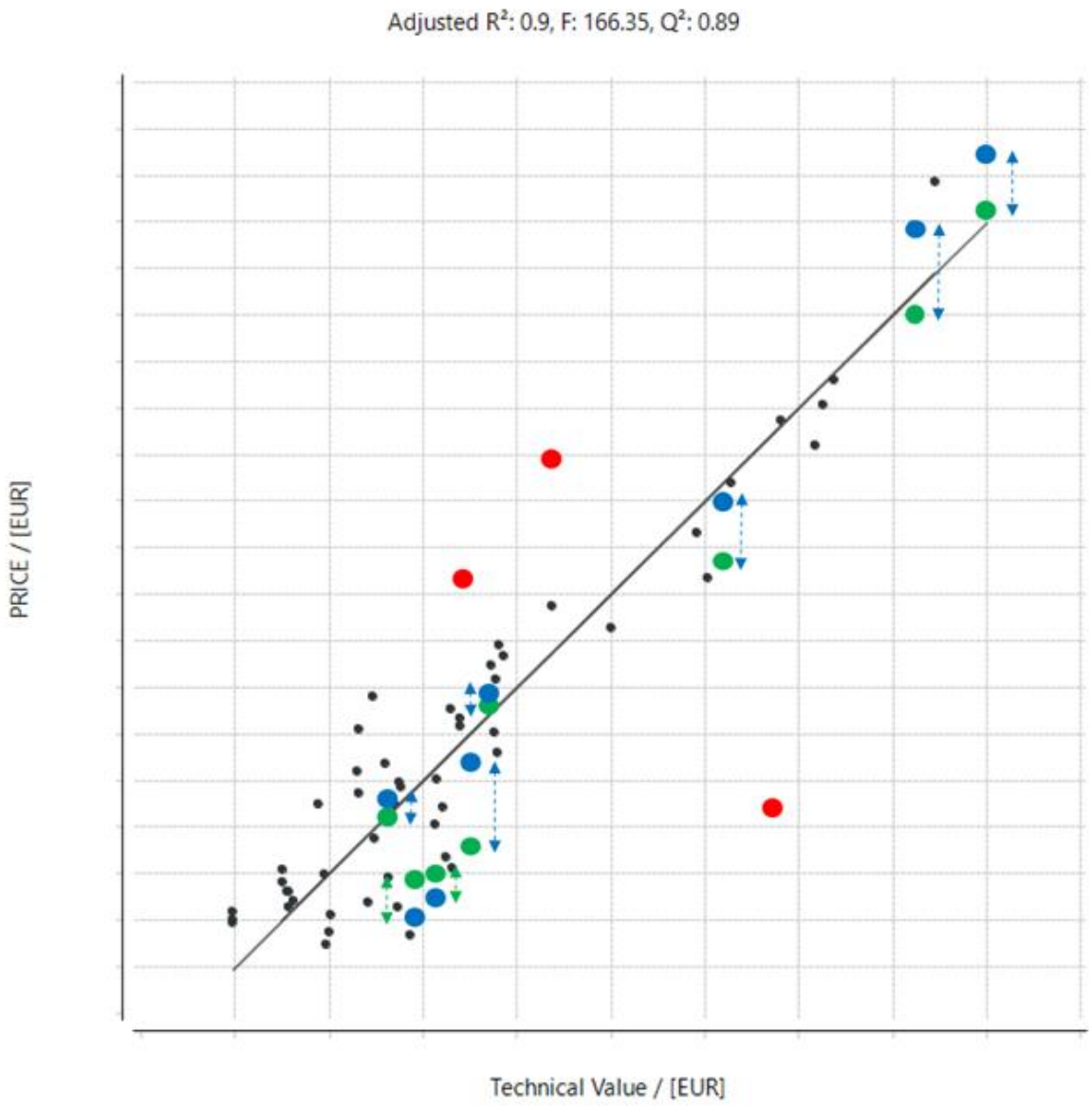
Model 1.12 with bottom-up calculations



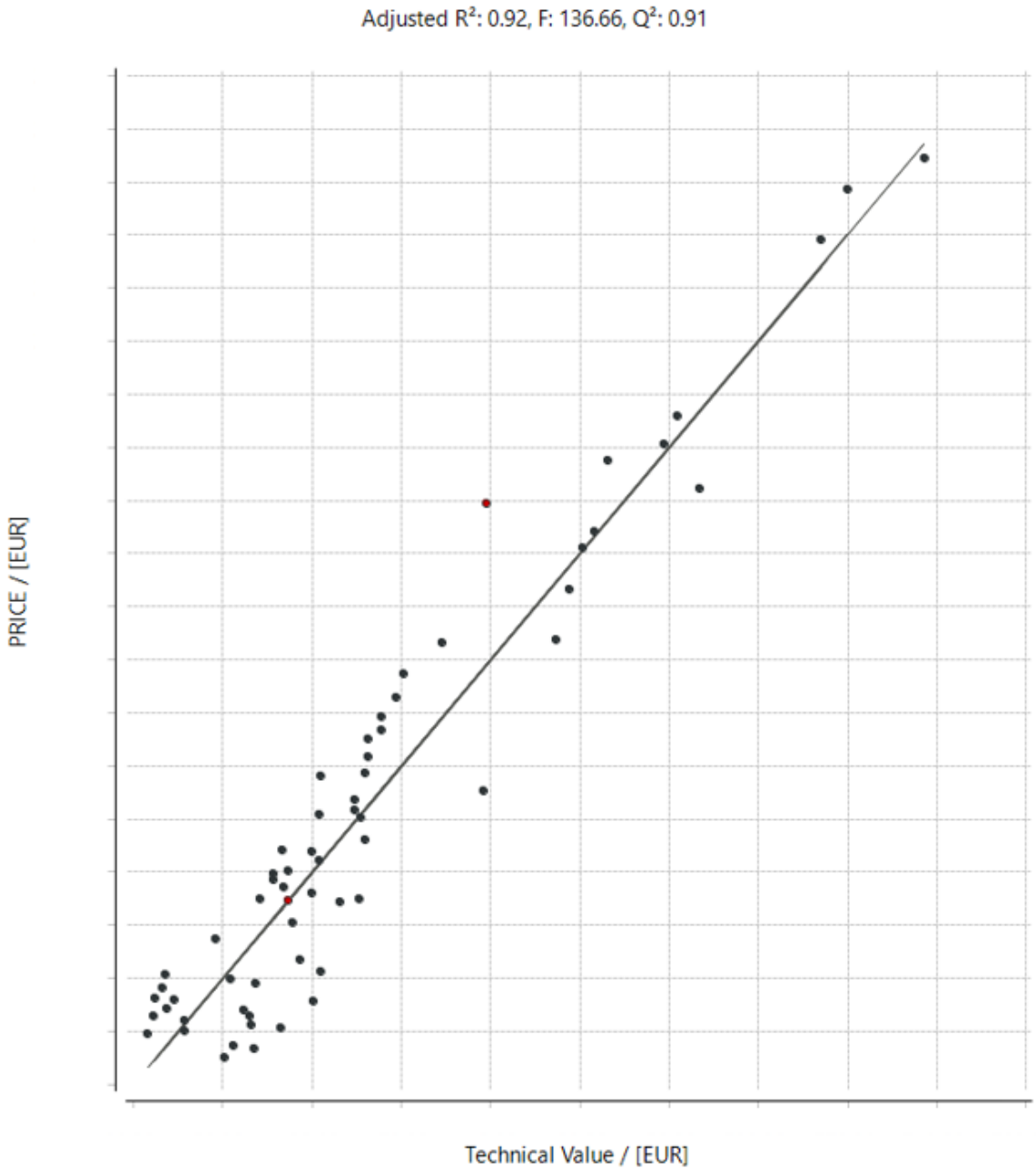
Model 2.2 (Weight + Rod diameter + Operational length)



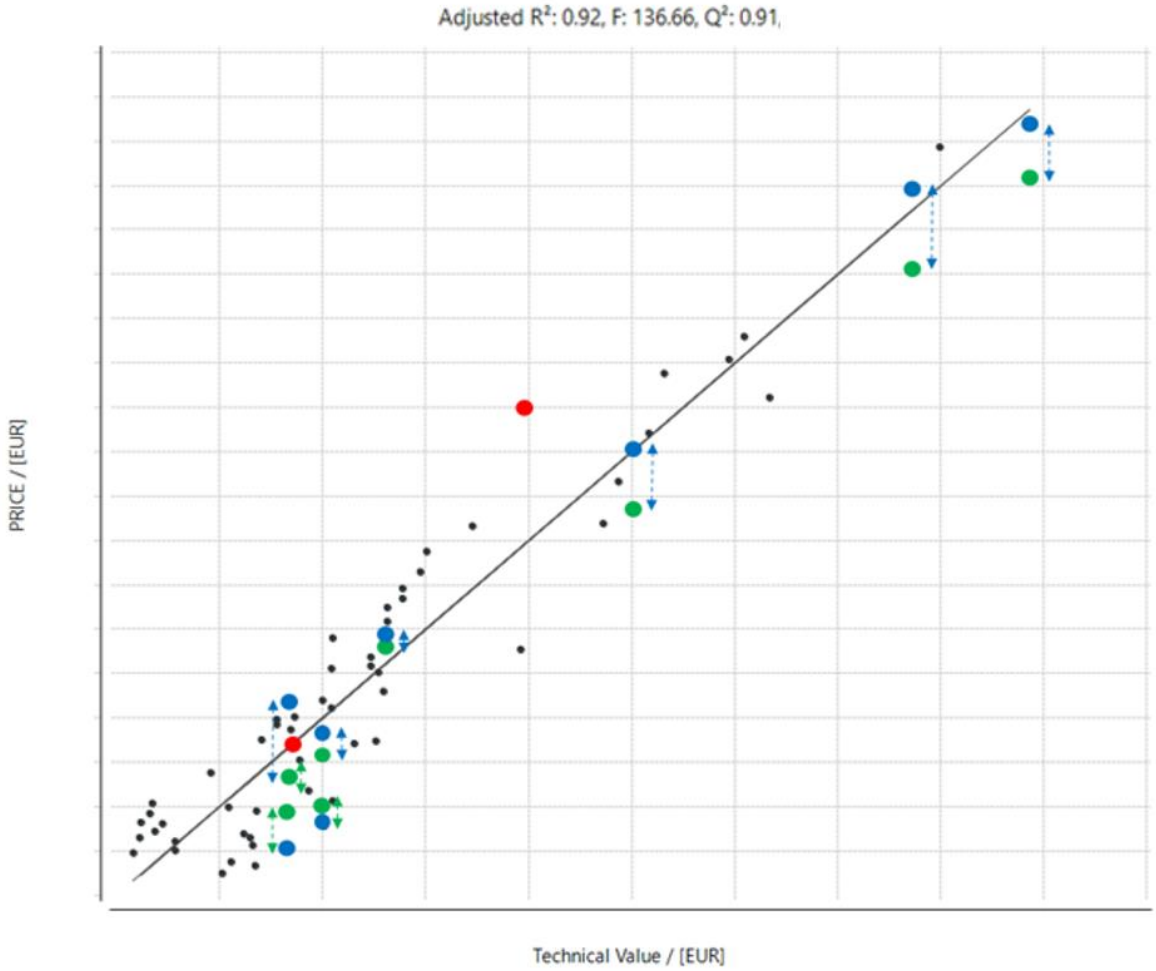
Model 2.2 with bottom-up calculations



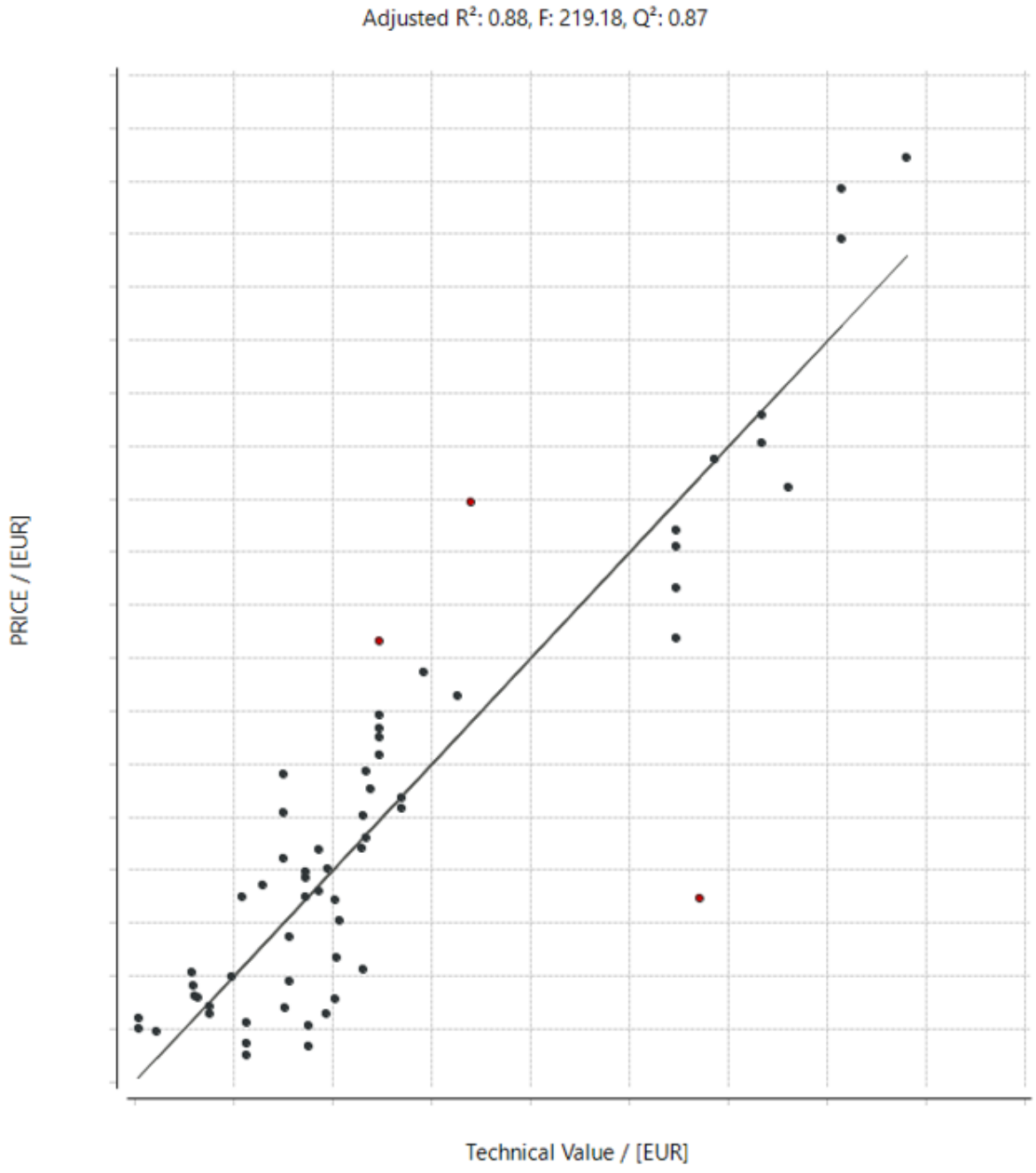
Model 3.4 (All cylinder types as separate + Inner diameter + Operational length)



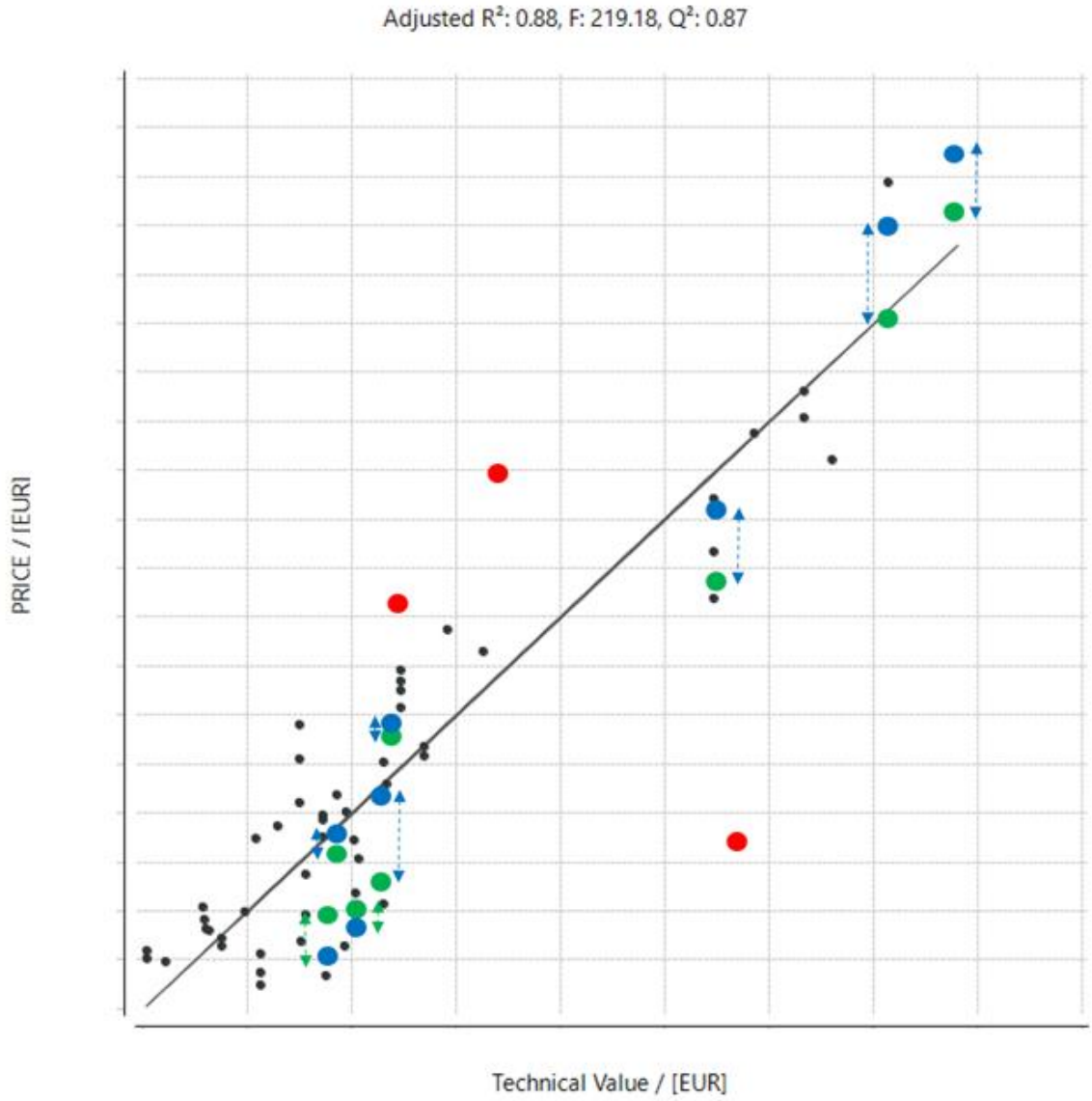
Model 3.4 with bottom-up calculations



Model 4.4 (Operational length + Rod diameter)



Model 4.4 with bottom-up calculations



APPENDIX 10.

Change in statistical indicators & technical values after bottom-up calculations included as part of the model.

Model	1.4		1.8		1.12	
	Orig.	PCM incl.	Orig.	PCM incl.	Orig.	PCM incl.
Adjusted R^2	0,95	0,95	0,95	0,95	0,93	0,94
Q^2	0,94	0,94	0,94	0,95	0,93	0,93
F	257,27	303,98	356,71	419,90	407,81	500,39
F critical	4,64	4,56	5,32	5,24	6,57	6,49
Sign F	0,00	0,00	0,00	0,00	0,00	0,00
Technical value	48.10 + 3.41 * Type 1 / [kg] + 4.09 * Type 2 / [kg] + 3.27 * Type 3 / [kg] + 7.03 * Type 4 / [kg]	48.33 + 3.30 * Type 1 / [kg] + 3,99 * Type 2 / [kg] + 3.22 * Type 3 / [kg] + 6,88 * Type 4 / [kg]	48,63 + 4,07 * Type 2 / [kg] + 3,27 * (Type 1 and 3) / [kg] + 7,04 * Type 4 / [kg]	48,44 + 3,99* T ype 2 / [kg] + 3,22 * (Type 1 and 3) / [kg] + 7,00 * Type 4 / [kg]	38.24 + 4.21 * (Type 1, 2 and 3) / [kg] + 8.22 * Type 4 / [kg]	37.75 + 4.16 * (Type 1, 2 and 3) / [kg] + 8.23 * Type 4 / [kg]
Comment						

Model	2.2		3.4		4.4	
	Orig.	PCM incl.	Orig.	PCM incl.	Orig.	PCM incl.
Adjusted R^2	0,89	0,90	0,92	0,93	0,87	0,89
Q^2	0,89	0,90	0,91	0,92	0,86	0,89
F	166,35	195,21	136,66	171,20	219,18	276,86
F critical	5,32	5,25	4,20	4,12	6,59	6,50
Sign F	0,00	0,00	0,00	0,00	0,00	0,00
Technical value	5.06 + 0.04 * Operational length / [mm] + 1.19 * Rod Diameter / [mm] + 1.60 * Weight / [kg]	-4.89 + 0.06 * Operational length / [mm] + 1.42 * Rod Diameter / [mm] + 0.54 * Weight / [kg]	-24.01 + 0.11 * Type 1 Ol / [mm] + 0.10 * Type 2 OL / [mm] + 0.59 * Inner Diameter / [mm] + 0.10 * Type 3 OL / [mm] + 0.23 * Type 4 OL / [mm]	-23.52 + 0.11 * Type 1 Ol / [mm] + 0.10 * Type 2 OL / [mm] + 0.55 * Inner Diameter / [mm] + 0.11 * Type 3 OL / [mm] + 0.24 * Type 4 OL / [mm]	-8.83 + 0.07 * Operational length / [mm] + 1.56 * Rod Diameter / [mm]	-7.38 + 0.08 * Operational length / [mm] + 1.48 * Rod Diameter / [mm]
Comment		Negative y-intercept after PCM data points added				

APPENDIX 11.

Observations for model 1.4 data points within selected technical value ranges.

Favorable data points in green, unfavorable in red.

Model	**1.4**	
Technical value range	85-105	105-125
Type 3	• Weight range 12,1-15,3 kg, AVG=13,8 kg • Favorable data points consist mainly on type 3 cylinders (7 pcs from 3 different suppliers) • All type 3 cylinders within this value range are below reg. line • Includes type 3 cylinders (ABC1234591 & ABC1234589) in which pricing currently unprofitable (compensated with other products)	• Weight range 18,7 kg – 22,5 kg, AVG=20,1 kg • The most of type 3 cylinders on top of the reg. line • If e.g. ABC1234579 is compared to ABC1234622 (both having exactly the same tech. value), the major explanatory factor is probably supplier (welded vs. forged end-eye/piston rod?). ABC1234574 (the most expensive within this range) is slightly larger in terms of ID (+5 mm) and OL (+19 mm) compared to ABC1234579. • Other clearly unfavorable PNs: ABC1234604, ABC1234623, ABC1234606, ABC1234567
Type 1	• Weight range 11,7-15,8 kg, AVG=13,6 kg. • Almost all type 1 cylinders are unfavorable. Partially due to "overpriced" ABC1234588 and ABC1234576 (compensation for type 3 cylinders), which for sure impacts on position of type 3 cylinders in whole model. • Only ABC1234609 clearly below reg. line (competitiveness was noted also during RFQ process). There is not clear technical explanation, why this individual cylinder is so cheap	• No type 1 cylinders within this value range
Type 4	• Weight range 6,25-7,2 kg, AVG=6,9 kg • Type 4 cylinders mostly on top of the reg. line • ABC1234570 only part number slightly below regression line, volume is clearly the highest • In case of type 4 cylinders, mounting arrangements, oil port design and piston rod design possible explanatory factors for variation (for example the difference in oil port design in ABC1234616 & ABC1234617)	• Type 4 cylinders close to reg. line • Based selected value drivers, ABC1234582 is clearly expensive. With more detailed review, this type 4 cylinder includes a special tube/block arrangement in vicinity of oil ports, which may partially explain the higher price. On the other hand, ABC1234582 is smaller in terms of RD (-15 mm), ID (-60 mm) and OL (-6 mm)
General comments	• If type 3 and type 1 are compared, type 3 cylinders are bigger in terms of ID (type 3 avg. 95 mm vs. type 1 83 mm). • No major differences on weight/RD/CL/WT etc. type 1 cylinders have slightly longer OL (+20 mm) • With type 1 cylinders, lower ID but similar weight in relation to type 3 cylinders partially explained by heavier cylinder head structure	• Some pricing differences on this area may be supplier related, involving different technical solutions (e.g. piston rod design/material), that are not specified in drawings/3D-models

APPENDIX 12.

Observations for model 3.4 data points within selected technical value ranges.

Model	3.4	
Technical value range	85-105	105-125
Type 3	• Weight range 12,1-24,1 kg, AVG=14,5 kg • Favorable data points consist mainly on type 3 cylinders (7 pcs from 3 different suppliers) • Includes ABC1234568 (which is taken out of the reg. and is heavier due to different design (plunger). This is interesting data point, since it's located at reg. line. Although R^2 sensitivity is low (-0,34 %), this individual data point has a strong impact on whole model in terms of R^2, Q^2 and F. • Includes type 3 cylinders (ABC1234591 & ABC1234589) in which pricing currently unprofitable (compensated with other products)	• Weight range 15,3 kg – 27 kg, AVG=20,2 kg • The most of type 3 cylinders on top of the reg. line. • Really similar data point locations compared to 1.4 • In addition to highlighted unfavorable data points (ABC1234604, ABC1234623, ABC1234606, ABC1234567) in model 1.4, model 3.4 includes also ABC1234605 as unfavorable data point, which was favorable in 1.4 (this makes sense, since this cylinder has wall thickness of 10mm (clearly thicker compared to others on this range) and it is thus heavier. In terms of ID and OD, this is really comparable to others). This is a good example, how the selection of value drivers impacts on results
Type 1	• Weight range 9,9-15,8 kg, AVG=12,7 kg. • Almost all type 1 cylinders are unfavorable. Partially due to "overpriced"ABC1234588 and ABC1234576 (compensation for type 3 cylinders), which for sure impacts on position of type 1 in the whole model. • Other unfavorable PNs: ABC1234625, ABC1234573, ABC1234626, ABC1234608 (like in model 1.4) • ABC1234609 seem to be good value for money in this model as well	• No type 1 cylinders within this value range
Type 4	• Weight range 6,9-8,9 kg, AVG=7,4 kg • Type 4 cylinders mostly on top of the reg. line • ABC1234570 only part number below regression line, volume is clearly the highest • ABC1234616 and ABC1234582 clearly unfavorable data points (like in model 1.4)	• Weight range 6,2-12,9 kg, AVG=9,4 kg • Type 4 cylinders mostly on top of the reg. line • ABC1234616 and ABC1234582 clearly unfavorable data points (like in model 1.4) • ABC1234597 seems to be good value for money (perhaps due to relatively long OL in relation to weight (in model 1.4 it was slightly on top of the regression line) • ABC1234583 and ABC1234590 located very close to reg. line like in 1.4
General comments	• Even though model 3.4 has different value drivers compared to 1.4, the locations of data points seem to be relatively same. Especially "the most" favorable and unfavorable data points are located similarly in relation to regression line	• Some pricing differences on this area may be supplier related, involving different technical solutions (e.g. piston rod design/material), that are not specified in drawings/3D-models

APPENDIX 13.

Data points (32 pcs) on top of the regression line (Model 1.4)

Grey color = Common part number found from both models 1.4 and 3.4.

Model 1.4		
Part number	**Technical Value**	**Potential to AVG line**
ABC1234579	112,15	13,16 %
ABC1234574	121,57	11,49 %
ABC1234623	109,35	12,47 %
ABC1234616	97,34	12,26 %
ABC1234628	157,12	7,30 %
ABC1234604	114,53	9,67 %
ABC1234606	111,26	8,52 %
ABC1234611	71,06	12,00 %
ABC1234625	90,22	9,48 %
ABC1234626	88,03	9,43 %
ABC1234620	144,18	5,89 %
ABC1234627	70,25	10,28 %
ABC1234567	110,68	6,62 %
ABC1234601	61,88	11,00 %
ABC1234573	90,97	7,71 %
ABC1234582	110,88	6,06 %
ABC1234608	88,79	6,57 %
ABC1234617	96,63	5,37 %
ABC1234594	82,09	6,20 %
ABC1234581	98,52	5,08 %
ABC1234592	71,54	6,23 %
ABC1234602	71,35	6,16 %
ABC1234614	68,39	6,14 %
ABC1234580	138,84	3,04 %
ABC1234613	70,02	5,72 %
ABC1234597	92,05	2,94 %
ABC1234619	174,84	1,49 %
ABC1234584	158,49	1,54 %
ABC1234588	101,94	2,07 %
ABC1234600	162,58	0,87 %
ABC1234576	99,56	0,59 %
ABC1234583	112,97	0,48 %

APPENDIX 14.

Data points (38 pcs) on top of the regression line (Model 3.4).

Model 3.4		
Part number	**Technical Value**	**Potential to AVG line**
ABC1234628	139,24	17,85 %
ABC1234574	120,41	12,33 %
ABC1234582	102,04	13,55 %
ABC1234580	129,12	9,83 %
ABC1234605	118,96	10,47 %
ABC1234579	115,49	10,58 %
ABC1234611	67,13	16,87 %
ABC1234623	112,54	9,92 %
ABC1234627	66,48	15,10 %
ABC1234592	64,92	14,90 %
ABC1234604	115,49	8,91 %
ABC1234619	166,40	6,25 %
ABC1234588	93,28	10,39 %
ABC1234616	101,69	8,34 %
ABC1234606	112,54	7,47 %
ABC1234594	78,47	10,34 %
ABC1234618	219,98	3,80 %
ABC1234625	91,30	8,40 %
ABC1234614	64,55	11,41 %
ABC1234573	91,30	7,38 %
ABC1234602	69,06	9,17 %
ABC1234613	67,50	9,12 %
ABC1234608	88,34	7,04 %
ABC1234567	111,97	5,53 %
ABC1234601	63,27	9,00 %
ABC1234576	94,68	5,46 %
ABC1234572	214,07	2,28 %
ABC1234583	109,47	3,56 %
ABC1234599	181,93	2,16 %
ABC1234581	100,05	3,60 %
ABC1234626	93,82	3,48 %
ABC1234590	109,47	1,96 %
ABC1234607	178,97	0,94 %
ABC1234593	71,35	0,85 %
ABC1234600	163,45	0,34 %
ABC1234584	160,50	0,29 %
ABC1234617	101,69	0,41 %
ABC1234568	94,53	0,17 %

APPENDIX 15.

Total saving potential for active serial parts in relation to average line (quantity considered and the order of part numbers in the list according to total saving potential) (Model 1.4).

Part number	Potential to average (%)	Comparable potential in 3.4 (%)
ABC1234567	7,09 %	5,93 %
ABC1234574	12,98 %	13,94 %
ABC1234579	15,16 %	12,18 %
ABC1234573	8,35 %	7,99 %
ABC1234582	6,45 %	14,42 %
ABC1234601	12,37 %	10,12 %
ABC1234581	5,35 %	3,80 %
ABC1234592	6,64 %	15,89 %
ABC1234594	6,62 %	11,02 %
ABC1234604	10,71 %	9,87 %
ABC1234606	9,32 %	8,16 %
ABC1234580	3,14 %	10,14 %
ABC1234602	6,56 %	9,77 %
ABC1234611	13,63 %	19,17 %
ABC1234608	7,03 %	7,53 %
ABC1234597	3,03 %	-
ABC1234588	2,11 %	10,60 %
ABC1234584	1,57 %	0,30 %
ABC1234616	13,98 %	9,51 %
ABC1234614	6,54 %	12,16 %
ABC1234613	6,06 %	9,67 %
ABC1234576	0,60 %	5,50 %
ABC1234600	0,87 %	0,34 %
ABC1234583	0,48 %	3,58 %
ABC1234617	5,67 %	0,44 %
ABC1234620	6,26 %	-
ABC1234619	1,51 %	6,34 %
Total potential in relation to purchasing spend (%)	**2,81 %**	

APPENDIX 16.

Total saving potential for active serial parts in relation to average line (quantity considered and the order of part numbers in the list according to total saving potential) (Model 3.4).

Part number	**Potential to average (%)**
ABC1234567	5,53 %
ABC1234574	12,34 %
ABC1234582	13,55 %
ABC1234592	14,90 %
ABC1234579	10,58 %
ABC1234580	9,83 %
ABC1234573	7,38 %
ABC1234588	10,39 %
ABC1234594	10,34 %
ABC1234576	5,47 %
ABC1234601	9,00 %
ABC1234605	10,47 %
ABC1234572	2,28 %
ABC1234602	9,16 %
ABC1234604	8,91 %
ABC1234611	16,87 %
ABC1234581	3,61 %
ABC1234583	3,56 %
ABC1234606	7,47 %
ABC1234608	7,03 %
ABC1234590	1,96 %
ABC1234614	11,41 %
ABC1234613	9,12 %
ABC1234599	2,16 %
ABC1234616	8,34 %
ABC1234593	0,84 %
ABC1234568	0,17 %
ABC1234607	0,94 %
ABC1234584	0,29 %
ABC1234600	0,34 %
ABC1234619	6,25 %
ABC1234618	3,81 %
ABC1234617	0,41 %
Total potential in relation to purchasing spend (%)	**3,68 %**