SCENTIA INTERNATIONAL ECONOMIC REVIEW Volume 1 - Issue 2 - 2022

Parametric estimation of supplier's plant construction costs

Remo Rossi, Mendel University Brno, Czech Republic

Abstract

Cost engineers of buying enterprises perform detailed product cost calculations of externally manufactured components. The aim of these calculations is to determine what a product should cost and to support purchasing functions in fact-based negotiations. While product cost engineers have deep knowledge in the calculation of direct cost, they need support in the calculation of supplier's indirect cost categories. The calculation of industrial rent, which is expressed in annual cost per m² of occupied plant building floor space can be improved by providing accurate construction cost estimates. Construction costs are strongly impacting the calculation of supplier's annual building depreciation, which is a crucial cost driver for the determination of the industrial rent. Academic literature is actually not providing an accurate and suitable cost model for product cost engineers, which is estimating construction cost per m² depending on different industrial building categories and alternative supplier plant locations. The paper aims to close this gap by applying linear regression analysis on a set of European construction cost data considering two industrial building categories: "warehouses/basic factory units" and "high-tech factories". By regressing construction cost against construction labor rates within different supplier plant locations it was possible to form suitable and accurate parametric regression functions with R² values between 0.74 and 0.88. Next to high R² values acceptable mean average percentage errors between 7.45% and 11.77% could be realized by comparing estimated with observed construction cost.

The estimation of industrial construction costs based on the paper's results can be used to improve the calculation of industrial rent, which is one cost element, that has to be covered within product cost engineer's Should Cost Calculations.

Keywords: cost estimation, construction cost, industrial plant buildings

JEL codes: C02, C51

1. Introduction

In highly competitive markets such as the automotive industry product cost calculations have become a practical standard in target cost setting for externally purchased and manufactured components (Hülsbömer, 2015; Möller, 2020, 105-106; Möller, 2020, 118). These Should Cost Calculations are contributing to higher profitability through increased bargaining power in fact-based negotiations due to improved knowledge of supplier's product cost structures (Ask/Laseter, 1998; Parsa, 2019, 61-63). Although they are directly affecting company's profitability only few authors describe the process and the information needed in order to carry out these calculations. Roy et al. (2011) present a rich description of data and information requirements needed in order to perform Should Cost Calculations within the automotive industry. They cluster this information within different cost categories, so that cost engineers are able to build their product cost calculations on a solid basis. Regarding the calculation of supplier's plant building costs, which are to be covered within machine hourly rates or overhead ratios Roy et al. (2011) give only few insights since they are focussing stronger on cost elements and information that contribute to direct manufacturing costs. Referring to costs of manufacturing buildings they name annual depreciation costs, maintenance and other cost elements while they state these costs are to be allocated within the machine hourly rates according the floor space occupied by different machine cost centres. (Roy et al., 2011, 7). In this context it is common cost engineering practice to aggregate all annual cost related to the plant building within a cost factor, which is also known under the term industrial rent, that is expressed in currency units per m² (Rossi, 2021). Next to the manufacturing building Roy et al. (2011) are indirectly referring to cost of other plant building categories such as warehouses for incoming goods, that have to be covered within material overheads (Roy et al., 2011, 6). Construction cost is one crucial parameter, which needs to be taken in consideration, in order to provide sound estimates for fully accounted industrial rents of supplier's manufacturing facilities (Rossi, 2021).

In this context following **research question** is raised:

"How can product cost engineers provide accurate construction cost estimates of supplier's plant buildings"?

This paper aims to answer this question by providing a parametric cost model which is based on regression analysis. The model enables cost engineers to provide fast and accurate estimates of construction costs per m² for different plant building categories, while country specific requirements can be taken in consideration.

Generally, cost estimation methods can be classified in qualitative and quantitative approaches (Cavallieri et al., 2004; Niazi et al., 2006). Qualitative methods select a cost optimal solution among alternative designs, rather than to predict a precise absolute quantitative value. (Layer et al., 2002, 502; Layer, 2003, 16; Salmi et al., 2016, 249). According Niazi et al. (2006) quantitative cost estimation can be grouped in two major subcategories: analytic and parametric cost estimation. Analytical approaches decompose the to be estimated object into the basic elements and/or manufacturing operations and are based on a bill of materials/bill of processes. They attempt to consider all resource spending for material, labor and overheads that are finally aggregated to total cost. Due to their transparent decomposition of cost analytic cost estimation methods are considered to provide the most accurate results. Their downside is, that they require intensive cost knowledge and huge time efforts in order to carry out the calculations. Consequently, these approaches fail in situations, in which cost engineers cannot rely on

deep cost knowledge or in case that the to be estimated object is not well defined (Cavalieri et al., 2004, 168; Ganorkar et al., 2017, 317-319; Niazi et al., 2006, 568); Farineau et al., (2001), 80).

Parametric cost estimation techniques instead utilize statistical methods in order to create cost estimation functions based on a limited set of product describing technical or physical variables, which are supposed to be the crucial cost drivers (Layer et al., 2002; Cavalieri et al., 2004; Duverlie/Castelain, 1999, 896). Hence, they can be applied even in situations, where the design of the to be estimated object can be only described in a rudimental manner (Cavallieri, 2004, 168). In this context parametric approaches might be also a suitable cost estimation technique for indirect cost categories, in which cost engineers might not have detailed knowledge or in case that a detailed analytic calculation doesn't stand in relation to the required time efforts.

Regression analysis is a frequently applied method that is used in construction cost estimation. Kim et al. (2004) for instance investigated the performance of cost estimation based on multiple regression analysis compared to cost estimation based on case-based reasoning and artificial neural networks. They included within their analysis a dataset of 530 construction cost records focussing on residential buildings. Within their comparison of different cost estimation techniques multiple regression analysis achieved an acceptable accuracy of a mean absolute estimation error of 6,95% compared to 2,97%, that has been achieved through neural networks or 4,81% by utilizing case-based reasoning. Even though multiple regression analysis was resulting in less accurate results compared to other cost estimation methods, the authors point out advantages such as time effectiveness to carry out the calculations and the transparent explanation of cost drivers.

Lowe et. al (2006) applied linear regression analysis on a dataset of 286 buildings in the United Kingdom and generated six cost estimation functions. 41 potential independent variables based on technical building parameters and to be fulfilled requirements were taken in consideration for model forming. Five independent variables were commonly used within all developed regression equations and the minimum of independent variables was 8. The achieved R² value was within the range of 0.661 to 0.668, while calculated mean absolute percentage errors ranged from 19,3 to 21,7%.

Sommez (2008) presents a cost estimation approach for building costs, which combines parametric cost estimation based on multiple regression analysis with probabilistic cost estimation. His study is based on 20 US-building projects, while 20 different parameters have been selected. His model was able to conduct calculations, that can be applied within the conceptual phase of building constructions. The average imputed error was deviating only 12 percent compared to detailed cost estimations.

Kim and Hong (2012) demonstrated a cost model, which combined a revised case-based reasoning technique with regression analysis. Their mixed cost modelling approach was enabling construction cost estimations for railroad bridges within the early planning phase. The model is verified based on five case studies and was able to reduce the error of cost estimation by 16.2% compared to previous applied cost model.

By reviewing the academic literature, it can be summarized, that regression analysis is a frequently applied method within in the field of construction cost estimation but was not focusing on the prediction of construction costs of industrial plant buildings based on limited information. Available models are not tailored to the needs of product cost engineers, who need assistance in determining supplier's industrial rents, that are to be included in their Should Cost Calculations. This paper closes this gap and provides accurate and effective parametric cost estimation functions of construction cost per m²

for multiple plant building categories and alternative plant locations based on regression analysis.

2. Materials and Methods

Within the section 2.1 "Materials" the underlying dataset will be described in detail on which the regression analysis will be based on. Section 2.2 "Methods" instead, will focus on the methodological approach of this paper. Next to a brief description of regression analysis and to be fulfilled requirements it includes an explanation of the validation technique as well that will be used, in order to rate the accuracy of the generated cost estimation functions. The creation of the cost estimation functions will be finally presented in chapter 3 "Results", before the paper will end with the chapters 4: "Discussion" and 5 "Conclusions".

2.1. Materials

The dataset which will be used, in order to apply linear regression analysis, is based on a global construction cost survey, which was conducted by Turner and Townsend in 2019. This study provides construction cost data of real estate projects and covers 27 commercial and industrial building categories for 64 country records. 20 out of these 64 records are related to Europe. Each record is referring to country specific data related to construction cost, construction-labor-rates and standard construction materials. The underlying cost data was collected within the last quarter of 2018, while building costs were referring to average values of construction cost based on typical regional standards of building construction (Turner and Townsend, 2019, 114). This paper will focus on the European dataset, in order to create cost estimation functions for construction cost of different categories of supplier plant buildings. In this context two types of industrial buildings will be taken in consideration. The first category, warehouses/basic factory units" covers industrial buildings with two possible utilitzations: Warehousing or the implementation of basic manufacturing and assembly operations, that require only a low level of automatization. The second building category "high tech factories" covers manufacturing buildings, that need a higher level of automatization. Since innovative and modern production technologies, which are exemplarily known under terms like industry 4.0, have not been implemented so far in larger scales (Krzywdzinski, 2017, 248), it can be concluded, that the term "high tech factory" is referring to manufacturing buildings, that are suitable to enable the implementation of highly automatized conventional production lines such as CNC production-, press-, or welding-lines (Krzywdzinski, 2017, 253-254).

Next to the explanation of to be covered building categories it is important to highlight all costs, which are covered within the term construction costs. The term construction costs refers to direct construction costs. Direct construction costs cover all costs related to plumbing, heating, ventilation, air conditioning, fire protection and all systems referring to electricity, communication and transportation. Furthermore it includes costs for the substructure, columns, upper floors, staircases, roof, internal and external walls and doors, finishes, ceiling, and fitments (Turner and Townsend, 2019, 114). Next to the cost items mentioned above contractor's profit margins and cost related to preliminaries are also included (Turner and Townsend, 2019, 20). Conversely developer's internal cost,

local authority fees, headwork charges, cost of land and costs referring to legal and financial aspects are excluded. Same counts for external works, landscaping, professional fees, demolition, loose furniture, fittings, equipment, sales taxes, test bores, site investigations, removal of significant ground constructions, abnormal footings and cost for onsite or underground parking (Turner and Townsend, 2019, 114).

All cost data within this paper is expressed in EUR, while currency exchange rates from the study of Turner and Townsend (2019) were taken in consideration (Turner and Townsend, 2019,114). The conversion of all values in one base currency was needed in order to apply regression analysis and to enable a comparison of calculated results.

The following three tables summarize the data used for each building category within the European dataset. The model forming data points were coded with the letters "**MF**" and an ascending number (*MF 1, MF 2, ...*). Data points which were not used to form the cost estimation functions were coded by "**OUT**" within the same logic (*OUT 1, OUT 2, ...*). Each data table refers to one of the previously mentioned building categories and includes cost data of construction costs per m² representing the dependent variable and one independent variable, which is referring to country-region specific construction labor rates. The reason why some data points were excluded from regression analysis and why only one single independent variable was considered for model forming will be pointed out in detail within the result section of this paper. For the category "**warehouses/basic factory units**" the total European dataset was spitted within two subsets "**United Kingdom and Ireland**" and "**Continental Europe**". The explanation for this split will be also given within the result section of this paper.

Table 1: Dataset United Kingdom and Ireland - Warehouses/ basic factory units

#	Country	Region	y Construction cost Warehouses/ basic factory units [Eur/m²]	x ₁ General construction labor rate [Eur/hour]
MF 1	Ireland	Dublin	1025	27
MF 2	UK	London	1029	26,84
MF 3	UK	Scotland	802	24,61
MF 4	UK	North	829	23,49
MF 5	UK	Central	847	25,72
MF 6	UK	South	884	24,61
OUT 1	UK	Northern Ireland	783	17,89

Table 2: Dataset Continental Europe – Warehouses/ basic factory units basic

#	Country	Region	y Construction cost Warehouses/ basic factory units [Eur/m²]	X ₁ General construction labor rate [Eur/hour]
MF 1	Austria	Vienna	687	32
MF 2	France	Paris	751	26
MF 3	Germany	Berlin	645	30
MF 4	Germany	Frankfurt	660	31
MF 5	Germany	Munich	713	33
MF 6	Netherlands	Amsterdam	820	35
MF 7	Poland	Warsaw	442,37	8,34
MF 8	Russia	Moscow	446,9	9,86
MF 9	Spain	Barcelona	492	19
MF 10	Spain	Madrid	484	19
MF 11	Sweden	Stockholm	871,44	36,7
MF 12	Turkey	Istanbul	283,04	4,47
OUT 1	Switzerland	Zürich	1270,57	84,11

Table 3: Dataset complete Europe- high tech factories

#	Country	Region	y Construction cost Warehouses/ factory units basic	X ₁ General construction labor rate [Eur/hour]
			[Eur/m ²]	
MF 1	Austria	Vienna	687	32
MF 2	France	Paris	751	26
MF 3	Germany	Berlin	645	30
MF 4	Germany	Frankfurt	660	31
MF 5	Germany	Munich	713	33
MF 6	Ireland	Dublin	1025	27
MF 7	Netherlands	Amsterdam	820	35
MF 8	Poland	Warsaw	442,37	8,34
MF 9	Russia	Moscow	446,9	9,86
MF 10	Spain	Barcelona	492	19
MF 11	Spain	Madrid	484	19
MF 12	Turkey	Istanbul	283,04	4,47
MF 13	UK	London	1028,95	36,84
MF 14	UK	Northern Ireland	782,89	17,89
MF 15	UK	Scotland	801,91	24,61
MF 16	UK	North	828,75	23,49
MF 17	UK	Central	846,64	25,72
MF 18	UK	South	883,55	24,61
OUT 1	Switzerland	Zürich	1270,57	84,11
OUT 2	Sweden	Stockholm	871,44	36,7

2.2. Methods

To generate cost estimation functions that are suitable to predict European construction cost of different plant building categories within a certain spread range the method of linear regression analysis was utilized. This method is based on the statistical correlation of an available dataset, which includes values for a single dependent and at least one independent variable. The approach is based on the condition of minimizing the sum of squares of deviations between existing data (*observed values*) and calculated values (*fitted values*) within a regression function. (Frost, 2019). In case of cost estimation, the independent variables refer typically to technical characteristics, while the depending variable is representing the cost of the to be estimated object (Pahl et al., 2007, 722-723; Ehrlenspiel et al., 2005, 458-459).

Linear regression models can be analytically expressed by equation no. 1 (Daniels/Minot, 2020, 196):

$$\hat{y} = a + b_1 x_1 + b_2 x_2 + \dots + b_n x_n \tag{1}$$

with:

i ε N; i=1-n;

ŷ: fitted or predicted value of the independent variable

x_i independent variable I

a, b_i: estimated constant a and coefficients b_i: of the independent variables

In this context \hat{y} denotes to the predicted or fitted value of the independent variable of the to be estimated object, $\mathbf{x_i}$ to the independent variables, which are rated with the estimated coefficients $\mathbf{b_i}$, while \mathbf{a} is referring to an estimated constant (Daniels/Minot, 2020, 196).

Next to the \mathbb{R}^2 -value and the ρ -values for the independent variables x_i the constant and coefficients (a, b_i) are the key outputs regression analysis (Frost, 2019, 34). The R^2 -value is indicating the variability in percentage of estimated costs, which can be explained based on the independent variables x_i , while the ρ -values are indicating whether the correlation between independent variables $\mathbf{x_i}$ and dependent variable \mathbf{y} are statistically significant (Frost, 2019, 29-31). A correlation coefficient R²=1 represents a perfect fit or correlation between actual and estimated values. R²-value close to 1 are indicators of a good correlation and a good performance of the cost estimation functions (Algahtani and Whyte, 2016, 35). The ρ -values for the constant a and the coefficients b_i indicate the probability to identify values equalling to \mathbf{a} and $\mathbf{b_i}$ Only by chance, with the probabilities ρ_i in case of non-existing statistical relationships between y and x_i (Daniels/Minot, 2020, 200-201). Within this paper a confidence interval of 95% was taken in consideration, which means that the maximum acceptable level of all **ρ-values** was set to 0.05. The cost values of the previously explained dataset and variables were transformed into EUR, while Microsoft Excel and the add in "Xrealstat" were used, in order to perform regression analysis. As the result section will highlight, the construction labor rate was fully suited to form significant and accurate regression functions, so that there was no need to consider other independent variables. In this regard it is important to mention, that the construction labor rates were fully fringed, which means, that they cover total cost from employers' perspective. Hence, they include next to the salary additional

contributions for pensions, health insurance and other benefits (Turner and Townsend, 2019, 18).

After having identified appropriate linear regression functions for all to be covered building categories a final validation was performed, in order to determine the achieved cost accuracies. This cross validation was needed, since high R^2 values indicating a good model fit and low ρ -values representing low probabilities, that model forming parameters (a, b_i) were determined just by chance, are not a guarantee for accurate predictions (Sommez, 2008, 1012). One method which can be utilized for validation purposes is the "leave one out" technique. By employing this technique a single datapoint is removed from the model forming dataset, while the remaining ones are used to form a regression function, which is used within a next step to estimate the value of the removed record. By repeating that procedure for all data points the mean absolute percentage error (MAPE) can be derived according to equation no. 2. This value is indicating the average deviation of predicted compared to observed values (Sommez, 2008, 1014).

MAPE =
$$\sum_{i} abs \left(\frac{c_e - c_o}{c_o} \right) / n$$
 (2)

with:

 c_e : estimated value of construction cost per m² c_o: observed value of construction cost per m² n: no of records within the model forming dataset

In context of the to be achieved accuracy of the cost estimation functions **MAPE-values** at a level of +/-15 % were set as a maximum to be accepted deviation. The decision for accepting deviations up to 15% is based on the fact, that occupied manufacturing floor space is typically not the most important cost driver among multiple parameters, that need to be considered in product cost calculations. This target is close to the allowed range in accuracy of a class two estimate, which is used for bid/tender controls according the definition of the Advancement of Cost Engineering (AACE, 2005, 2).

Model-forming and validation within this paper can be described as an iterative or stepwise process, which included as well data cleansing/removal of outliers in combination with the check of crucial theoretical requirements of regression analysis. Academic literature is pointing out following seven classical requirements, which must be fulfilled, so that regression analysis gurantees statistically unbiased and efficient predictions (Frost, 2019, 221-231).

1 Linearity

This assumption is referring to the linearity of the regression model, so that an independent variable y can be expressed and predicted by a linear combination of independent variables x_i , the constant a and it's coefficients b_i (Thomas/Thomas, 2017, 2016). A correctly specified regression function must therefore fulfill the requirement to model the statistical relationships and effects between the variables $(x_i;y)$ in a linear manner, so that the investigated subject is specified appropriately (Frost, 2019, 221-223).

2 The error term has a population mean of zero

This assumption is basically focusing on the unpredictability or randomness of the error term, which forces the average value of residuals to equal zero. In this case the model is not biased, which means that it is not sytematically under- or overestimating the observed

values. Whenever a costant \mathbf{a} is added to the model this requirement is automatically fulfilled, since it is ensured, that the mean of residuals are equal to zero (Frost, 2019, 224). Due to the reason that all to be determined regression equations will include a constant \mathbf{a} , there is no need to check this assumption individually within the result section of this paper.

3 No correlation among independent variables x_i and the error term u_i

In case that correlation between an independent variable and the error term would exist, this assumption would be violated. The error term would not be a randomly derived unpredictable term and the estimate of the coefficients (a, b_i) would be biased.

In order to verify this assumption scatter plots with residuals u_i on the one and the values of the independent variables x_i on the other axis can be used, while the aim is to determine a randomly destributed pattern arround the value of zero (Frost, 2019, 224-225).

4 No autocorrelation within the error term

This assumption is stating, that the observations of the error term don't have to correlated with each other, which means, that a single observation should not enable to predict the following observation within time series (e.g. GDP or inflation data). An appropriate method to reject potential autocorrelation is to graph the residuals in sequence and to identify randomness within the scatterplot. Whenever a cyclical pattern is recognized instead of randomness the assumption of autocorrellation is hurt, while adding further independent variables to the model may remove the effect (Frost, 2019,226).

Due to the reason that the regression analysis and underlying data is not referring to time series, the analysis of this pre requirement will not be part of the to be checked regression diagnostics within the results section of this paper.

5 No heteroscedasticity within the error term

No heteroscedasticity can be observed, whenever there is a constant variance within the error term for all observations $\mathbf{u_i}$. A simple way to check this assumption is to create a scatterplot with residual values $\mathbf{u_i}$ on the one axis and fitted or predicted values $\mathbf{y_i}$ on the other axis. Existing heteroscedasticity would be recognized through the increase of residuals in a cone shape within one direction of the x-axis (Frost, 2019, 227-229).

6 No Multicolinearity between independent variables x_i

The term multicolinearity is used in case that the independent variables x_i are correlated with each other. This is a problem, since the key goal of applying regression analysis is to determine the individual contribution of each model forming variable, in order to predict the dependent variable y. In this case the coefficients b_i are representing the mean change in y for each incremental change in a single independent variable x_i , while holding all remaining factors constant. Hence the verification of no or only minor effects of colinearity among the independent variables x_i is crucial for the quality of the overall model. Multicolinearity can decrease the precision of estimated regression coefficients b_i , which impacts again the statistical power of the regression function. In such cases low p-values do not necessarily guarantee, that the selected independent variables are statistically significant, which means, that an accurate and causal attribution of the variables x_i on the explained variable y is not guaranteed.

Allthough this might not affect the overall precision of the total predicted values or the goodness to fit statistics (Frost, 2019, 241-245), the understanding of independent effects

may be important to explain the calculated results and to conduct fact based negotiations. In order to identify multi-colinearity and it's strenght the calculation of variation inflation factors (VIFs) is a suitable method. In this regard a VIF of 1 indicates non existing multicolinearity. However there is no consensus within academic literature on a threshold of VIF values causing problems related to multi-colinearity. While a rule of thump is indicating that VIF values greater than 10 are problematic, some authors set a threshold value of 5 or even 4 (Daniels/Minot, 2020, 235; Frost, 2019, 245-246). As will be highlighted within the result section of this paper, it will be possible to form

As will be highlighted within the result section of this paper, it will be possible to form accurate and significant regression functions based on one single independent variable. Hence there will be no need to analyse this pre requirement.

7 Normally destributed of residuals

In order to achieve unbiased total estimated values with a high accuracy and a minimum of variance, the fulfillment of this assumption is often described as optional. The advantage of normally distributed error terms is the possibility to conduct statistical hypothesis testing, which allows to determine the statistical significance of each single independent variable within the overall model. Normally distributed residuals are enabeling the creation of reliable predictions of the constant \mathbf{a}_i , the coefficients \mathbf{b}_i and their underlying confidence intervalls. A simple way to check this requirement is the generation of a normal prohability plot, while normal distributed residuals can be assumed, in case that they are following a straight line (Frost, 2019,229-231). Next to this graphical approach two statistical tests -"Shapiro Wilk" and "d´ Agostino-Pearson" were performed, in order to check the normality assumption of the residuals. The "Shapiro-Wilk-test" determines weather the null hypothesis, which is stating, that a data-sample is normaly distributed has to be rejected or may be retained. The null hypothesis of normality is rejected and the the outcome of the test is stating that the sample records are not normally distributed, whenever the significance probability ρ of the underlying test statistic is below or equal the significance level of α =0.05. Conversly the failure of rejection is suggesting that the sample is normally destributed (Martin/Bridgmon, 2012, 114). The "d'Agostino-Pearson test" instead checks normality by determining individual skewness and kurtosis of the to be determined sample and is afterwards testing a joint null hypothesis referring to normality of skewness and kurtosis of the underlying dataset (Daniels/Minot, 2020, 247). Whenever the significance level of the corresponding test statistic ρ is $<\alpha$ =0.05 the null hypothesis is rejected, so that the normality assumption is fulfilled in case that the significance level ρ is $>\alpha=0.05$ (Zaionts, 2021).

3. Results

This Chapter presents the creation of regression equations, the analysis of regression requirements and their validation in terms of achieved accuracy.

Regression Analysis -Warehouses/basic factory units-Complete Europe

Within a first step regression analysis was performed based on the complete European Dataset including country records of the United Kingdom and Ireland, while $\mathbf{x_1}$: general construction labor rate [EUR/hour] was considered as single independent

variable. Although all ρ -values were significantly below 0.05 the investigation was leading only to a R^2 value of 0.58. Hence it was tried to identify a potential second independent variable. After several failures a second independent variable x_2 :cost for 1000 concrete blocks (400x200mm; >10,000 block job) [EUR] was identified, that was leading to ρ -values below 0.05, while a R^2 -value of 0.76 could be identified. Within a next step the regression requirements were checked. First of all the linearity of the model was analysed by a scatter plot, in order to investigate the correlation between observed and predicted values, as well as the correlation between the independent variables and the observed values (See Fig. 1-3).

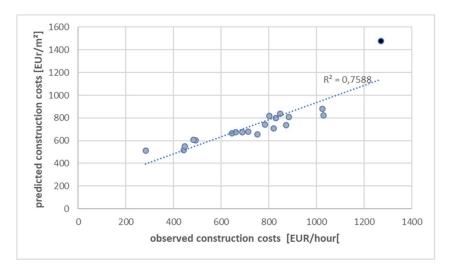


Figure 1: observed against predicted construction costs warehouses/basic factory units complete Europe

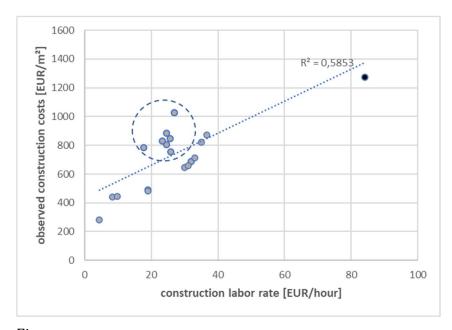


Figure 2: construction labor rate against construction cost warehouses/basic factory units Europe

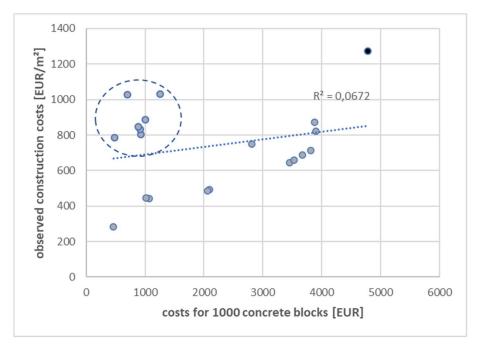


Figure 3: costs of concrete blocks against observed construction costs warehouses/basic factory units Europe

This investigation has resulted in three crucial findings:

First: The country record for Switzerland/Zurich was identified as an outlier (marked in black) in all three scatterplots. **Second:** Only a very low correlation between the cost of concrete blocks \mathbf{x}_2 and the construction costs of warehouses/basic factory units could be identified. **Third:** A second review on the scatterplot indicated, that it makes sense to split the dataset into one which refers to Continental Europe (*grey data points*) and another one, which is reflecting all records of the United Kingdom and Ireland (*grey data points within the circle*). Based on these findings it has been decided, to split the dataset in two independent subsets and to conduct regression analysis a second time by considering construction labor rate as single independent variable and to remove the outlier Zurich/Switzerland from the model forming dataset.

Regression Analysis -Warehouses/basic factory units-Continental Europe:

The second regression analysis, which was considering 12 data points (see Tab. 1), has lead to a $\mathbf{R^2}$ -value of 0.88, which means that only 12% of the variation of the model could not to be explained by the model forming variable. The ρ -values for the intercept \mathbf{a} and the coefficient $\mathbf{b_1}$ were both close to zero and significantly below 0.05, which is indicating a low probability, that the values for the coefficients \mathbf{a} , $\mathbf{b_1}$ were caused randomly, without any existing statistical causation.

Based on this initial analysis regression requirements were checked:

The **linearity assumption** was checked by analysing two scatter diagrams. The first is plotting observed against predicted construction costs for basic factory untis/

warehouses (Fig. 4). The second diagram instead is plotting construction labor rate against the observed construction costs within this building category (see Fig. 5).

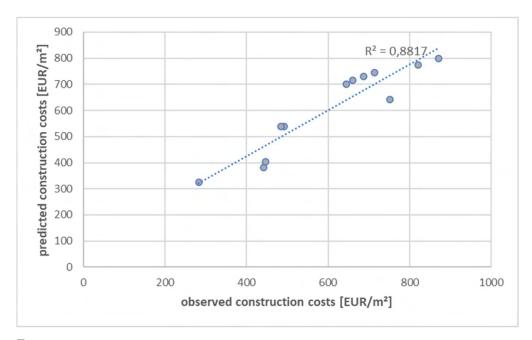


Figure 4: observed against predicted construction costs for warehouses/basic factory units Continental Europe

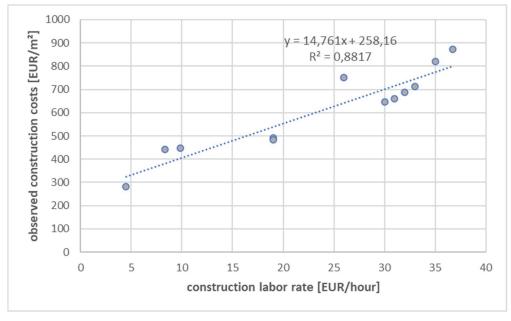


Figure 5: construction labor rate against observed construction costs warehouses/basic factory units Europe

The analysis of both scatter diagrams was leading to the conclusion that the linearity assumption was fulfilled.

This assumption of non existing correlations among independent variables x_i and the error term u_i was checked by analysing a scatter diagram, which plotted the

independent variable x_1 (constuction labor rate of a general laborer) against the residual values (see Fig 6).

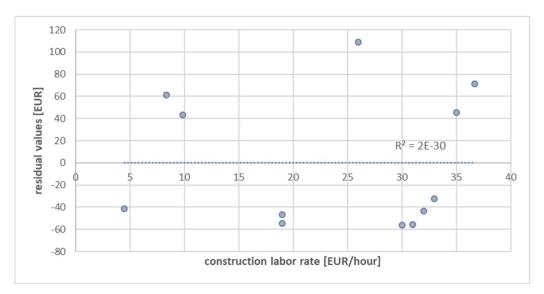


Figure 6: construction labor rate against residual values warehouses/basic factory units Continental Europe

The associated scatter-diagram was showing no significant correlation, so that it can be concluded, that this assumption was fulfilled as well.

The **requirement of heteroscedasticity within the error term** was analysed based on the scatter diagramm in Fig. 6. By reviewing the scatterplot it can be also concluded, that there is no heteroscedasticity within the sample of residuals, since there is a constant scattering arround the mean value of zero.

The **assumption of normality** was analysed by reviewing a scatterplot which compares **z-values** with the values of the studentized residuals. In addition to that a Shapiro Wilk and a d'Agostino-Peason test was taken in consideration.

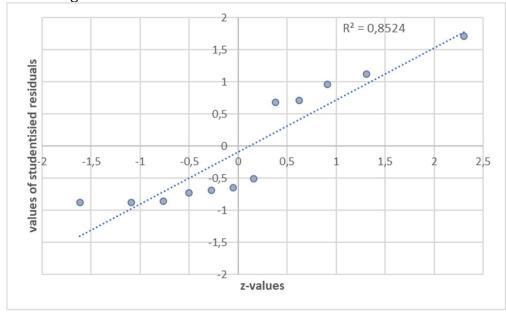


Figure 7: z-values against studentized residuals values warehouses/basic factory units Continental Europe

Since the studentized residuals are following a straight line, the scatter plot is indicating a normal distribution of the error term. This indication is as well confirmed by the results of the d'Agostino-Pearson test (ρ =0.21 > α =0.05), while the Shapiro Wilk Test (ρ =0.016 < α =0.05) failed to support the assumption of normality. Based on the graphical results and the support of at least one of both statistical tests, it can be concluded that the normality assumption is fulfilled. Based on the previous investigations it may be concluded that all crucial regression requirements were fulfilled. In addition to that a **MAPE-value** of 11.77% could be achieved by applying previously explained leave one out technique, which is indicating a very good accuracy of the cost estimation function. Further interpretation of the regression function can be given by reviewing equation no. 3:

$$\hat{y} = 258.16 + 14.71x_1 \tag{3}$$

In this context **y-hat** is indicating the estimated value of construction cost [EUR/m²] for warehouses/basic factory units in Continental Europe, while x_1 is refering to a construction labor rate [EUR/hour] of a general constuction laborer. The value of the constant a=258.16 EUR can be considered as the theoretical minimum of construction cost to be spent per m² for warehouses and basic factory units. The increase in construction cost per m² based on the change in cost of the independent parameter x_1 is impacted by additional 44.5 EUR/m² for an increase of 1 EUR within the labor hourly rate. Finally it can be summerized, that the regression equation is based on a very good fit of data ($R^2=0.88$), that the underlying formula is based on real statistical relations (ρ -values <<0.05), while the accuracy of the cost estimation formula can be rated as high (MAPE=11.77%). Since the regression function was meeting the targeted objectives, it was decided not to check for further potential independent variables, but to stop the investigation, in order to avoid an increase in model's complexity.

Regression Analysis -Warehouses/basic factory units-UK and Ireland

The regression analysis which was conducted on the European sub-dataset contained originally 7 data points (see Tab. 2). The original analysi which was considering $\mathbf{x_1}$ (construction labor rate) as single independent variable, has lead to a medium $\mathbf{R^2}$ -value of 0.51 while the $\boldsymbol{\rho}$ -values were not significant ($\boldsymbol{\rho_0}$ =0.269; $\boldsymbol{\rho_1}$ =0.068). Based on that imperfect result two scatterplots were analysed, in order to identify potential outliers.

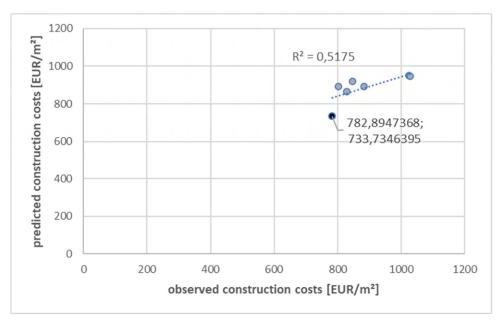


Figure 8: observed against predicted construction costs warehouses/basic factory units – UK and Ireland

By reviewing Fig. 8 and Fig. 9 it was possible to identify the record United Kingdom/Norhern Ireland (*marked in black*) as an outlier, that is potentially distorting the quality of the regression function.

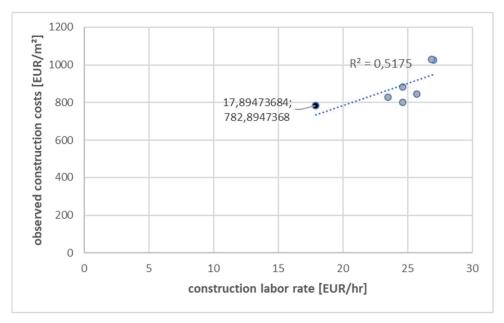
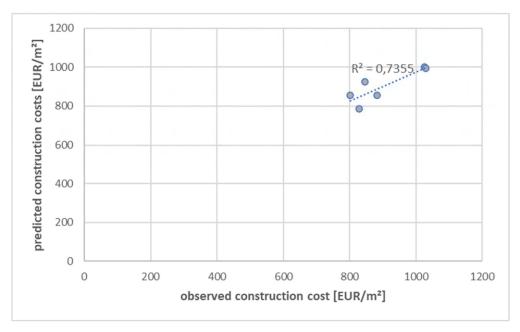


Figure 9: cconstruction labor rate against observed construction costs warehouses/basic factory units – UK and Ireland

Based on that finding it had been decided to conduct a second regression analysis based on 6 data points (see Tab. 2) and to remove Northern Ireland from the model forming dataset. The second regression analysis has lead to a high R^2 -value of 0.74. The ρ -value for the intercept a was 0.23, while coefficient b_1 was significant by reaching a value of 0.03. This is indicating at least for the varable x_1 , that the coefficient b_1 was estimated based on a real existing statistical relationship.

Within a next step crucial requirements for regression analysis are analysed in detail:

In order to check the **linearity assumption** two scatter diagramms were analysed. The first was comparing observed with predicted values of construction cost for warehouses/basic factory units. The second one was plotting the construction labor rate against the observed value of construction costs (see Fig. 10-11).



 $Figure \ 10: oobserved \ against \ predicted \ construction \ costs - warehouses/basic \ factory \ units - UK \ and \ Ireland$

The review of both diagramms indicate, that the linearity assumption of the regression model is fullfilled. Both diagramms show a clear linear relationship.

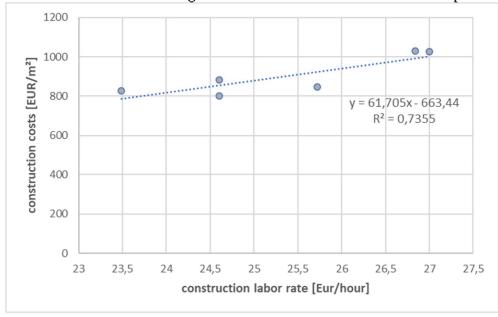
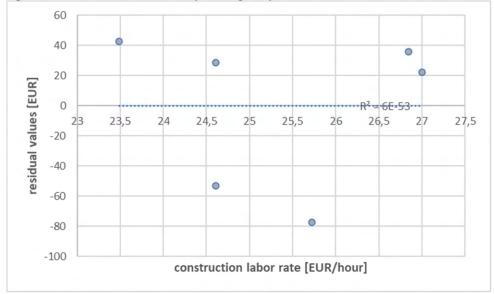


Figure 11: construction labor rate against construction costs – warehouses/basic factory units – UK and Ireland

The assumption of non existing correlation among among independent variables x_i and the error term u_i was checked by analysing a scatter diagram, while the independent variable $x_1(constuction\ labor\ rate\ of\ a\ general\ laborer)$ was plotted against the residual values (see Fig. 12).



 $Figure\ 12:\ construction\ labor\ rate\ against\ residual\ values\ -\ warehouses/basic\ factory\ units\ -\ UK\ and\ Ireland$

Since the scatter plot is indicating no significant correlation, it can be stated that this assumption is fullfilled.

The assumption of non existing heteroscedasticity within the error term was analysed based on the scatter diagramm of Fig. 12. Based on the graphical analysis it can be concluded, that there is no heteroscedasticity, since there is a constant scattering of the residuals arround the value of zero, while no bottle neck/funnel-like pattern of residuals can be observed by moving along the x-axis.

The assumption of normal destributed residuals was analysed with a scatterplot, which compares z-values with the values of the studentized residuals (see Fig. 13). In addition to that a Shapiro Wilk and d'Agostino-Peason test was performed.

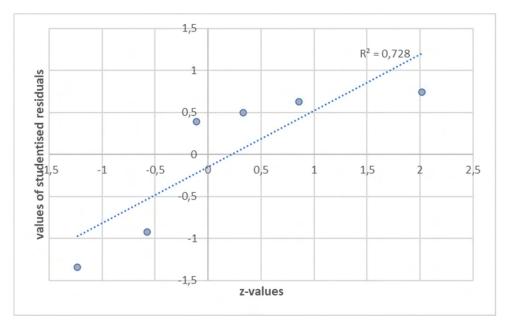


Figure 13: z-values against studentized residuals - warehouses/basic factory units - UK and Ireland

Since the studentized residuals are following a straight line the scatter plot is indicating a normal distribution of the error term. This indication is as well supported by the results of a Shapiro Wilk Test (ρ =0,057 > α =0.05). Due to the limited amount of data it was not possible to conduct a d' Agostino-Pearson test. Based on the graphical results and the result of the Shapiro Wilk test it can be concluded, that the residuals are normally distributed. All previous investigations indicate that all crucial regression requirements were met. In addition to that the accuracy of the cost estimation function was evaluated by a calculated **MAPE-value** of 7.45%, which was indicating a very good accuracy of the created cost estimation function. Futher interpretations can be given based on equation no. 4:

$$\hat{y} = -663.44 + 61.71x_1 \quad (4)$$

In this context **y-hat** is indicating the estimated value of construction cost [EUR/m²] for warehouses/basic factory units within the UK and Ireland, while x_1 is referring to a construction labor rate [EUR/hour] of a general constuction laborer. The constant a =-664.44 EUR can be considered as the theoretical minimum of construction cost to be spent per m² for warehouses and basic factory units. The increase in construction cost per m^2 based on the change in cost of the independent parameter x_1 is impacted by additional 61.71 EUR/m² for an increase of 1 EUR within the labor hourly rate. Finally it can be concluded, that the regression equation is based on a good fit of data due to \mathbb{R}^2 =0.74. Even though the ρ -value of the constant **a** is in contrary to the coefficient \mathbf{b}_1 not significant the cost estimation function is able to provide a high quality in cost prediction, which was verified by a **MAPE-value** of 7.45%. Hence it can be stated that regression function is fully suitable to predict construction costs for warehouses and basic factory units depending on their location. This argumentation is supported with a growing consensus in academic literature, that researchers should not focus on reaching low **ρ-values** below 0.05 stand alone, but also recognize the importance of the size of confidence intervalls, in which the values (a; b_i) might fall (Daniels/Minot, 2020, 201). Since the regression function is meeting the requirements to achieve the objective of this paper, it was not checked weather additional potential independent variables could be considered. Instead of that, it has been decided to keep the complexity of the model at a reasonable level.

Regression Analysis -high factory units - complete Europe

The regression analysis focussing on high tech factories was performed on the complete European Dataset by considering again the construction labor rate x_1 as single independent variable. Already in the first step the datapoint Switzerland/Zurich was removed as an outlier, due of the extremely high construction labor rate compared to the other records. By taking all remaining 19 data points in account (see Tab. 3) a medium correlation with \mathbf{R}^2 = 0.65 was identified. In addition to that, the ρ -values for the intercept \mathbf{a} and the coefficient \mathbf{b}_1 reached significant values close to zero. In order to improve the quality of the model a scatter plot was analysed, which compared the construction labor rate x_1 with the observed values \mathbf{y} (see Fig. 14).

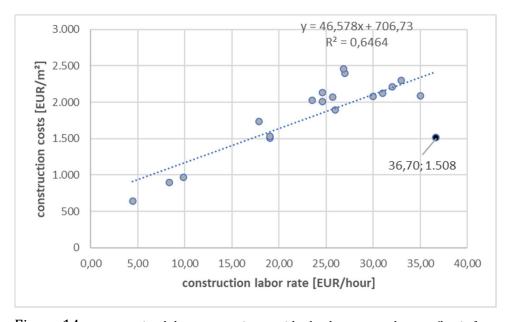


Figure 14: construction labor rate against residual values - warehouses/basic factory units - Europe

Through graphical analysis the record Stockholm/Sweden (*marked in black*) was identified as outlier and removed from the dataset. Another finding from this investigation was, that a split of the dataset in two independent samples was not suitable to improve the quality of the model intensively, so that it was decided to form one regression function for complete Europe. By removing the Swedish outlier from the dataset and performing the regression analysis a second time the ${\bf R}^2$ -value could be improved to 0.85 while the ${\bf p}$ -values for ${\bf a}$ and ${\bf b}_1$ remained significant with values close to zero. Based on these results which were indicating a very good fit of the regression function and a good quality of the model in terms of ${\bf p}$ -values <<0.05 the regression requirements were checked in detail within a next step:

The **linearity assumption** of the model was verified by two scatter plots. The first diagram was comparing observed with predicted values of construction cost for high tech

factories, while the second was plotting the construction labor rate against the observed value of construction costs within this building category (see Fig. 15-16).

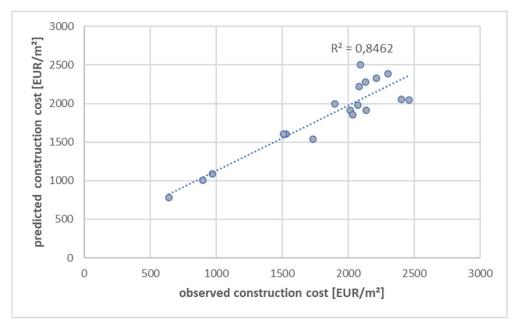


Figure 15: observed against predicted construction costs – high tech factories – complete Europe

A strong correlation between observed and predicted construction cost for high tech factories can be recognized in Fig. 15, while Fig. 16 is illustrating also a strong linear correlation between the construction labor rate x_1 and the observed values of y.



Figure 16: construction labor rate against construction costs – high tech factories – complete Europe

The assumption of non existing correlation among independent variables x_i and the error term u_i was verified by a scatter diagram by plotting the independent variable $\mathbf{x_1}(constuction\ labor\ rate)$ against the residual values (see Fig. 17). Since no correlation could be identified, the assumption is fullfilled.

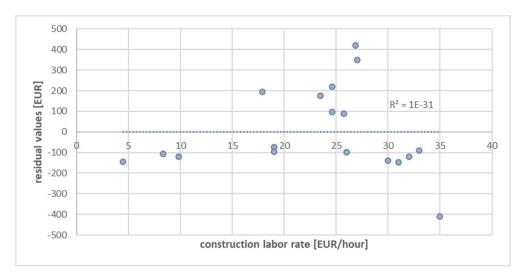


Figure 17: construction labor rate against construction costs – high tech factories – complete Europe

The **requirement of heteroscedasticity within the error** term was checked based on Fig. 17. The diagram is vizualising that the residuals are scattering relatively constantly arround the x-axis, while no bottle neck/funnel shape pattern can be observed by moving along the x-axis. This is an indicator for non existing heteroscedasticity.

The **assumption of normally destributed residuals** was analysed by using a scatter diagram, which is plotting z values on the x-axis against the values of the studentized residuals on the y-axis (see Fig. 18).

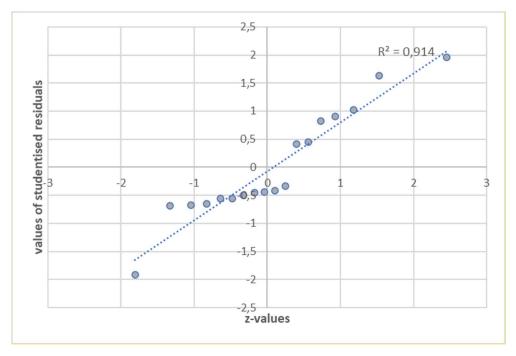


Figure 18: z-values against studentized residuals – high tech factories – Europe

Since the studentized residuals are following a straight line the scatter plot is indicating a normal distribution of the error term. This finding is also confirmed by the results of the d'Agostino-Pearson test (ρ =0.706 > α =0.05) and the Shapiro Wilk test (ρ =0,092 >

 α =0.05). Based on the graphical analysis and the results of the statistical tests it can be concluded, that the residuals are normally distributed. By applying the previously explained leave one out technique a **MAPE-value** of 11.48% could be realized, which is indicating a high accuracy in addition to the fullfillment of all regression requirements.

Futher interpretations can given by reviewing equation no. 5:

$$\hat{y} = 534.68 + 56.16x_1 \tag{5}$$

The variable **y-hat** is indicating the estimated value of construction cost [EUR/m²] for high tech factories within Europe, while x_1 is refering to a construction labor rate [EUR/hour] of a general constuction laborer. The value of the constant \mathbf{a} =534.68 EUR can be interpreted as the theoretical minimum of construction cost spending, which is achieved in case of x_1 = 0. The increase in construction cost per m² based on a one EUR change in construction labor rate \mathbf{x}_1 is 56.16 EUR/m². Finally it can be summerized that the regression equation is based on a very good fit of data (\mathbf{R}^2 =0.85). In addition to that the underlying formula is based on significant statistical relations which are expressed by the identified coefficients \mathbf{a} and \mathbf{b}_1 , since their \mathbf{p} -values are close to zero and far below $\mathbf{\alpha}$ =0.05. Also the accuracy of the cost estimation function can be rated as very high, due to a calculated MAPE value of 11,48%. Thus there was no need to add further additional independent variables in order to improve the quality of the model.

Tab. 4 is finally summerizing the results of this paper. The ${\bf R^2}$ values of the regression functions were within a range between 0.74 up to 0.88 while MAPE values between 7.45% up to 11.77% were achieved.

|--|

Region	Category	Regression Equation – all values in EUR	Quality	independent variables
Continental Europe	Warehouses/ Basic factory units	$\hat{y} = 258.16 + 14.71x_1$	$\begin{array}{l} R^2 = 0.88 \\ \rho_o = 0.00017265 \\ \rho_1 = 5.998*10^{(-6)} \\ MAPE = 11.77\% \end{array}$	\hat{y} : construction cost $\left[\frac{EUR}{m^2}\right]$ x_1 : construction labor rate general laborer $\left[\frac{EUR}{hour}\right]$
United Kingdom and Ireland	Warehouses/ basic factory units	$\hat{y} = -663.44 + 61.71x_1$	$R^2 = 0.74$ $\rho_o = 0.23$ $\rho_1 = 0.029$ MAPE=7.45%	\hat{y} : construction cost $[\frac{EUR}{m^2}]$ x_1 : construction labor rate general laborer $[\frac{EUR}{hour}]$
Europe	High-tech factories	$\hat{y} = 534.68 + 56.16x_1$	$R^{2} = 0.85$ $\rho_{o} = 0.002$ $\rho_{1} = 6.6285*10^{(-8)}$ MAPE=11.48%	$ \hat{y} : construction cost \left[\frac{EUR}{m^2} \right] $

While the next chapter "Discussion" is finally discussing the results of regression analysis the chapter "Conclusion" focusses on potential future research.

4. Discussion

By comparing the results of regression analysis, it can be confirmed, that the determined regression equations achieve a good quality in terms of cost prediction. Since only one independent variable was needed in order to form the cost estimation functions, calculations can be conducted fast, without major time efforts and without deeper

construction cost knowledge. A big advantage in this context is, that national and region-specific construction labor rates are regularly published by national or European statistical institutions (e.g., www.destatis.de; ec.europa.eu). Another advantage is, that labor rates are not fluctuating strongly compared to other cost factors such as standard construction materials based on concrete or steel. The cost estimation functions can be considered to be stable for several years, since the impact of inflation is considered within the yearly increase of labor costs.

Although all the three regression equations show a good to very good data fit in terms of regression analysis ($0.78 < R^2 < 0.88$) some limitations of the study are worth to mention. A certain level of variability in the data (12% < Var < 22%) could not be explained by the regression models. This variability can change over time. It can be stated that a certain level of variation in construction costs is driven by effects, that are not covered by changes in construction labor rates. One of these effects could be based on tendentially higher construction costs in overheating markets whenever contractors are able to enforce higher profit margins (Turner and Townsend, 2019, 20). Another limitation of the study is based on the relevance of the model forming parameters. In case of rapid changes in construction technology and/or building requirements, the original regression functions might not be appropriate anymore. In case of substitution of construction labor due to higher automatization in building construction within the future, it might be necessary to take a completely new model forming dataset and more than one model forming variable in consideration.

5. Conclusion

Finally, it can be stated, that the developed cost estimation formulae are fully suited to determine construction cost per m² for different plant building categories and locations. However, each supplier plant covers next to industrial also office floor spaces, whereat building requirements and hence construction costs are different. Hence future research could extent the results of this paper and create cost estimation functions, that are suited to calculate construction costs of office areas/buildings within a plant. Commercial building categories and underlying datasets within the study from Turner and Townsend (2019) may be unitized. Additionally, cost estimation functions could be created, that are focussing on regions outside Europe. Since building requirements and market conditions and therewith construction costs for different building categories are varying globally additional regression functions could be formed based on North American and Asian datasets. Finally, it needs to be mentioned that the developed cost estimation functions have to be integrated within a bigger cost model, which converts construction costs into annual cost for depreciation per m². This spending needs to be summed up to with additional cost elements such as insurance, maintanance, energy consumption and others, in order to result in a fully accounted annual industrial rent per m². These industrial rents could be then finally used to improve cost engineers Should Cost Calculations.

Within a next step industrial rents could be either used as direct input within the calculation of machine houlry rates or recognized as a cost element within a complexer plant cost models. Such a model could finally aim to determine supplier's material or manufacturing overheads (Rossi, 2021).

References

AACE 2005. Association for the Advancement of Cost Engineering. Cost estimate classification system – as applied in engineering, procurement and construction for the process industries. TCM-Framework: 7.3 - Cost Estimating and Budgeting. In: AACE international, S. 1–9 [Online]. Available at: http://alganature.com/article/AACE.pdf [Accessed: 02 August 2021].

ALQAHTANI, A., WHYTE, A. 2016. Estimation of life-cycle costs of buildings: regression vs artificial neural network. In: *Built Environment Project and Asset Management* 6 (1), S. 30–43. DOI: 10.1108/BEPAM-08-2014-0035.

ASK, J. A.; LASETER, T. M. 1998. Cost Modeling: A Foundation Purchasing Skill. In: *strategy+business* (issue 10) [Online]. Available at: https://www.strategy-business.com/article/9625 [Accessed 02 August 2021].

CAVALIERI, S., MACCARRONE, P., PINTO, R. 2004. Parametric vs. neural network models for the estimation of production costs: A case study in the automotive industry. In: *International Journal of Production Economics* 91 (2), S. 165–177. DOI: 10.1016/j.ijpe.2003.08.005.

DANIELS, L., MINOT, N.W. 2020. An introduction to statistics and data analysis using Stata. From research design to final report. Los Angeles: SAGE.

DUVERLIE, P., CASTELAIN, J.M. 1999. Cost Estimation During Design Step: Parametric Method versus Case Based Reasoning Method. In: *The International Journal of Advanced Manufacturing Technology* 15 (12), S. 895–906. DOI: 10.1007/s001700050147.

EHRLENSPIEL, K., KIEWERT A., LINDEMANN, U. 2005. Kostengünstig Entwickeln und Konstruieren. Kostenmanagement bei der integrierten Produktentwicklung. 5. Aufl. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg (VDI-Buch).

FARINEAU, T., RABENASOLO, B., CASTELAIN, J.M., MEYER, Y., DUVERLIE, P. 2001. Use of Parametric Models in an Economic Evaluation Step During the Design Phase. In: *The International Journal of Advanced Manufacturing Technology* 17 (2), S. 79–86. DOI: 10.1007/s001700170195.

FROST, J. 2019. Regression Analysis. An Intuitive Guide for Using and Interpreting Linear Models. 1. Edition, Pennsylvania: Statistics by Jim Publishing.

GANORKAR, A.B., LAKHE, R.R., AGRAWAL, K.N. 2017. Cost estimation techniques in manufacturing industry: concept, evolution and prospects. In: *IJEA* 8 (3/4), S. 303. DOI: 10.1504/IJEA.2017.092279.

HÜLSBÖMER, D. 2015. So kauft die Automobilbranche ein. In: Markt und Mittelstand [Online]. Available at: https://www.marktundmittelstand.de/einkauf/so-kauft-die-automobilbranche-ein-1225011/. [Accessed: 02 August 2021].

KIM, B-S., HONG, T. 2012. Revised Case-Based Reasoning Model Development Based on Multiple Regression Analysis for Railroad Bridge Construction. In: *J. Constr. Eng. Manage.* 138 (1), S. 154–162. DOI: 10.1061/(ASCE)CO.1943-7862.0000393.

KIM, G.-H., AN, S.-H., KANG, K.-I. 2004. Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. In: *Building and Environment* 39 (10), S. 1235–1242. DOI: 10.1016/j.buildenv.2004.02.013.

- KIM, G.-H., YOON, J.-E., AN, S.-H., CHO, H.-H., KANG, K.-I. 2004. Neural network model incorporating a genetic algorithm in estimating construction costs. In: *Building and Environment* 39 (11), S. 1333–1340. DOI: 10.1016/j.buildenv.2004.03.009.
- KRZYWDZINSKI, M. 2017. Automation, skill requirements and labour-use strategies: high-wage and low-wage approaches to high-tech manufacturing in the automotive industry. In: *New Technology, Work and Employment* 32 (3), S. 247–267. DOI: 10.1111/ntwe.12100.
- LAYER, A. 2003. Case based cost estimation. A building block for product cost management and design-for-X. Dissertation. Universiteit Twente, Twente [Online]. Available at: https://ris.utwente.nl/ws/portalfiles/portal/6119138/thesis_Layer.pdf/. [Accessed: 02 August 2021].
- LAYER, A., BRINKE, E., VAN HOUTEN, F., KALS, H., HAASIS, S. 2002. Recent and future trends in cost estimation. In: *International Journal of Computer Integrated Manufacturing* 15 (6), S. 499–510. DOI: 10.1080/09511920210143372.
- LOWE, D.J., EMSLEY, M.W., HARDING, A. 2006. Predicting Construction Cost Using Multiple Regression Techniques. In: *Journal of Construction. Engineering and. Management* 132 (7), S. 750–758. DOI: 10.1061/(ASCE)0733-9364(2006)132:7(750).
- MARTIN, W.E., BRIDGMON, K.D. 2012 Quantitative and statistical research methods. From hypothesis to results. First edition. San Francisco, CA: Jossey-Bass (Research methods for the social sciences, 42).
- MÖLLER, S. 2020. Cost Engineering. In: SCHNURRENBERGER, B.J. (Editor). Praxisbeiträge zum KMU-Management. Wachstumsfinanzierung, Consulter-Auswahl, Change-Management, Cost Engineering, Online-Marketing & E-Commerce, Verkaufstraining: KMU-Info.de Verlagsservices.
- NIAZI, A., DAI, J.S., BALABANI, S., SENEVIRATNE, L. 2006 Product Cost Estimation: Technique Classification and Methodology Review. In: *Journal of Manufacturing Science and Engineering* 128 (2), S. 563–575. DOI: 10.1115/1.2137750.
- PAHL, G., BEITZ, W., FELDHUSEN, J., GROTE, K.-H. 2007. Konstruktionslehre. Grundlagen erfolgreicher Produktentwicklung; Methoden und Anwendung. 7. Aufl. Berlin, Heidelberg: Springer.
- PARSA, D. 2019 Should-Cost Analysis as an Alternative to Open Book Accounting. Master of Science Thesis. Tampere University, Tampere. Industrial Engineering and Management [Online]. Available at: https://trepo.tuni.fi/handle/123456789/27082 [Accessed 31.07.2021].
- ROY, R., SOUCHOROUKOV, P., SHEHAB, E. 2011. Detailed cost estimating in the automotive industry: Data and information requirements. In: *International Journal of Production Economics* 133 (2), S. 694–707. DOI: 10.1016/j.ijpe.2011.05.018.
- ROSSI, R. 2021. Cost estimation of supplier's industrial rent for plant buildings from cost engineer's perspective. In: *IfM Institut für Management (.) (2021): IfM Impulse. Das Wissensmagazin des Instituts für Management.* IfM Institut für Management. Hallwang/Salzburg: IfM (19).
- SALMI, A., DAVID, P., BLANCO, E., SUMMERS, J.D: 2016 A review of cost estimation models for determining assembly automation level. In: *Computers & Industrial Engineering* 98, S. 246–259. DOI: 10.1016/j.cie.2016.06.007.

SOMMEZ, R. 2008. Parametric Range Estimating of Building Costs Using Regression Models and Bootstrap. In: *J. Constr. Eng. Manage.* 134 (12), S. 1011–1016. DOI: 10.1061/(ASCE)0733-9364(2008)134:12(1011).

THOMAS, N., Thomas, A.V. 2016 Regression Modelling for Prediction of Construction Cost and Duration. In: *Applied Mechanics and Materials* 857, S. 195–199. DOI: 10.4028/www.scientific.net/AMM.857.195.

 $TURNER\ AND\ TOWNSEND\ 2021.\ International\ construction\ market\ survey\ [Online].$ $Available\ at:\ http://www.infrastructure-$

intelligence.com/sites/default/files/article_uploads/Turner%20Townsend%20Internat ional%20Construction%20Market%20Survey%202019.pdf. [Accessed: 02 August 2021].

ZAIONTS, C. 2021 D'Agostino-Pearson Test [Online]. Available at:

https://www.real-statistics.com/tests-normality-and-symmetry/statistical-tests-normality-symmetry/dagostino-pearson-test/ [Accessed:09 August 2021].