Oldenbourg Verlag

Onur Dursun

Early Estimation of Project Determinants

Predictions through Establishing the Basis of New Building Projects in Germany

Schriftenreihe Bauökonomie –
herausgegeben von Prof. Dr. Christian Stoy

Early Estimation of Project Determinants

Predictions through Establishing the Basis of New Building Projects in Germany

von

Dr.-Ing. Onur Dursun

2

Oldenbourg Verlag München

Editor: Dr. Stefan Giesen
Production editor: Tina Bonertz
Cover design: hauser lacour

Dissertation, Universität Stuttgart (D 93), 2013

Library of Congress Cataloging-in-Publication Data
A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the German National Library
The German National Library lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Rosenheimer Straße 143, 81671 München, Germany
www.degruyter.com/oldenbourg
Part of De Gruyter

Printed in Germany

This paper is resistant to aging (DIN/ISO 9706).

ISBN 978-3-11-034638-1
eISBN 978-3-11-034787-6

Early Estimation of Project Determinants

Predictions through Establishing the Basis of New Building Projects in Germany

Faculty of Architecture and Urban Planning, University of Stuttgart
A thesis submitted in fulfillment of the requirement for the award of the Degree of Doctor of Engineering (Dr. Eng.)

by
Onur Dursun
born in Izmir, Turkey

Supervisor: Prof. Dr. Christian Stoy
Co-Supervisor: Prof. Dr. Arnold Tautschnig
Co-Supervisor: Prof. Dr. Michael Oberguggenberger

Date of Defense: 22.August.2013

Institute for Construction Economics, University of Stuttgart
2013

List of Abbreviations

AFS	Average floor size
ANNs	Artificial neural networks
ANOVA	Analysis of Variance
APE	Absolute percentage error
ASH	Average storey height
BKI	Cost Information Centre of German Chamber of Architects GmbH
BTC	Bromilow's time-cost model
CBR	Case-based reasoning
CG	Cost group according to hierarchical structure of DIN 276-1
CT	Client type
DPA	Developed plot area
FAC	Facility type
GBV	Gross building volume
GC	Ground conditions
GEFA	Gross external floor area
GRN	General regression neural net
HOAI	Regulation on fees for architects and engineers
iid	independently and identically distributed
k300	cost of structure - construction works
k300+k400	cost of structure
k310	cost of excavation
k320	cost of foundations
k330	cost of external walls
k340	cost of internal walls
k350	cost of floors and ceilings
k360	cost of roofs
k370	cost of structural fitments
k390	cost of other construction related activities
k400	cost of structure - services
k410	cost of sewerage, water and gas systems
k420	cost of heat supply systems
k430	cost of air treatment systms

k440	cost of power installation systems
k450	cost of telecommunications and other communication systems
k460	cost of transport systems
k470	cost of function related equipment and fitments
k480	cost of building automation
k490	cost of other services related work
LI	BKI location index
LRM	Linear regression models
LSE	Least square estimate
m310	Excavation volume (cubic meters)
m320	Foundation area (square meters)
m330	External walls area (square meters)
m340	Internal walls area (square meters)
m350	Floors and ceilings area (square meters)
m360	Roof area (square meters)
MAD	Mean absolute deviation
MAER	Mean absolute error rate
MAPE	Mean absolute percentage error
MC	Market conditions
MLF	Multi layer feedforward neural network
MLP	Multi layer perceptrons
MSE	Mean square error
NoS	Number of storeys
NoSag	Number of storeys above ground
NoSbg	Number of storey below ground
PA	Plot area
RMSE	Root mean squared error
RUQ	Reference unit quantities
SA	Site access
SMEs	Small and Medium Size Enterprises
TOP	Topographical conditions
€	Currency unit of European Union

List of Figures

1.1	Three main success criteria of projects	2
1.2	Hierarchical structure of DIN 276-1	3
1.3	Current cost planning practice in Germany (Source: HOAI (2009) and DIN 276-1 (2008))	4
1.4	Proposed frameworks for cost estimation	8
1.5	Proposed frameworks for construction duration prediction	10
3.1	Main stages of experimental design (Source: Fellows and Liu (2008))	34
3.2	A typical feedforward network (MLP)	42
3.3	GRN Net block diagram (Specht, 1991)	46
4.1	An example of a development plan issued by Stadtplanungsamt	52
4.2	Scatterplot of external walls area and gross external floor area	54
4.3	Scatterplot between predicted values of external walls area and actual values of cost of external walls	55
6.1	Summary statistics of test sample for absolute percentage errors of cost framework #1 via linear regression models	95
6.2	Summary statistics of test sample for absolute percentage errors of cost framework #1 via artificial neural network models	96
6.3	Summary statistics of test sample for absolute percentage errors of cost framework #2 via linear regression models	96
6.4	Summary statistics of test sample for absolute percentage errors of cost framework #2 via artificial neural network models	97
6.5	Summary statistics of test sample for absolute percentage errors of cost framework #3 via linear regression models	98
6.6	Summary statistics of test sample for absolute percentage errors of cost framework #3 via artificial neural network models	98
6.7	Summary statistics of test sample for absolute percentage errors of cost framework #4 via linear regression models	99
6.8	Summary statistics of test sample for absolute percentage errors of cost framework #4 via artificial neural network models	100
6.9	Comparison of cost estimation frameworks with respect to method of analysis	101
6.10	Summary statistics of test sample for absolute percentage errors of construction duration framework #5 via linear regression models	102
6.11	Summary statistics of test sample for absolute percentage errors of construction duration framework #5 via artificial neural network models	103

6.12 Summary statistics of test sample for absolute percentage errors of construction duration framework #6 via linear regression models 103
6.13 Summary statistics of test sample for absolute percentage errors of construction duration framework #6 via artificial neural network models 104
6.14 Comparison of construction duration frameworks with respect to method of analysis 105

7.1 BKI Object 1300-082 108
7.2 Relative impacts of input nodes for external wall area prediction via MLF net 112
7.3 Relative impacts of input nodes for cost of external wall area prediction via MLF net 114
7.4 Relative impacts of input nodes for construction duration prediction via GRN net 118
7.5 Residual plots of linear regression model for cost of external walls 122

List of Tables

1.1	Cost groups of structure and corresponding reference quantities (Source: DIN 276-1 and DIN 277-3)	3
2.1	Linear regression model results of cost models by Emsley et al. (2002)	18
2.2	Neural network model results by Emsley et al. (2002)	18
2.3	Input variables of Lowe *et al.* (2006); Emsley *et al.* (2002)	19
2.4	Variable included in the regression models of Lowe *et al.* (2006)	20
2.5	Model determinants to predict cost of structural system per square meters (Murat Gunaydin & Zeynep Dogan, 2004)	20
2.6	Candidate predictors of cost models (Karanci, 2010)	21
2.7	Classification of the key studies with respect to method of analysis	22
2.8	Model parameters and effect measures for tendered and actual road contracts in England	25
2.9	Model parameters and effect sizes of buildings categorized by client, contract and tendering type in England	25
2.10	Estimated and actual values and resulting effect sizes for public and private buildings in Hong Kong	26
2.11	Correlation and regression results of the time-cost relationships of building projects in Hong Kong	26
3.1	Notation for the data used in regression analysis	37
4.1	Qualitative attributes of the sample	50
4.2	Comparison between training and test samples	51
4.3	Available quantitative predictor variables to an estimator at the basic evaluation stage of HOAI	52
4.4	Response variables and their descriptive statistics	53
5.1	Candidate predictors for reference unit quantity models	59
5.2	Estimated parameters and closeness of fit for final linear regression models of reference unit quantities	60
5.3	Prediction performance of linear regression models of reference unit quantities	63
5.4	Architecture and prediction performance of ANN models for reference unit quantities	64
5.5	Candidate predictors of second level cost groups of structure models	67
5.6	Estimated parameters and closeness of fit for final linear regression models of second level cost groups of structure, construction works	69
5.7	Estimated parameters and closeness of fit for final linear regression models of second level cost groups of structure, services	72

5.8 Prediction performance of linear models for second level cost group of structure 76
5.9 Architecture and prediction performance of ANN models for second level cost groups of structure, construction works 77
5.10 Architecture and prediction performance of ANN models for second level cost groups of structure, services 79
5.11 Candidate predictors of linear regression models for the first level cost groups of structure via multi-way approach 81
5.12 Estimated parameters and closeness of fit for final linear regression models of first level cost groups of structure via multi-way approach 82
5.13 Prediction performance of linear regression models of first level cost groups of structure via multi-way approach 84
5.14 Architecture and prediction performance of ANN models for first levels cost groups of structure via multi-way approach 84
5.15 Candidate predictors of linear regression models for the first level cost groups of structure via one-step ahead approach 86
5.16 Estimated parameters and closeness of fit for final linear regression models of first level cost groups of structure via one-step ahead approach 87
5.17 Prediction performance of linear regression models of first level cost groups of structure via one-step ahead approach 88
5.18 Architecture and prediction performance of ANN models for first levels cost groups of structure via one-step ahead approach 88
5.19 Candidate predictors of construction duration models 89
5.20 Estimated parameters and closeness of fit for final linear regression models of construction duration 90
5.21 Prediction performance of linear regression models of construction duration 91
5.22 Architecture and prediction performance of ANN models for construction duration 92

6.1 Comparison of cost estimation frameworks with respect to method of analysis 94
6.2 Framework comparison of the duration models 102

7.1 Model inputs, that were available to the architect, of 1300-082 at the basic evaluation stage of HOAI. 109
7.2 Predicted values of reference unit quantities for 1300-082 111
7.4 Predicted values of cost groups of structure for 1300-082 116
7.5 Analysis of variance for cost of sewerage, water and gas systems (k410) linear model 117
7.6 95% confidence intervals for expected values of 1300-082 119

Abstract

Construction cost and duration are two strategic determinants of building projects along with quality. Objective estimations of these determinants are crucial through establishing the basis of a project, hence they serve as foundation for budgeting, planning, executing, monitoring and even for any litigation aims. The research is designed, having two primary aims for the population of interest, in this case new building construction in Germany.

First, to form an objective basis for conceptual estimates, empirical investigation of historical project information is conducted, using two alternative method of analysis, multiple linear regression and artificial neural networks. This incorporates with development of predictive models that can further be employed as an objective ground for practitioners.

Second, to increase prediction accuracy of conceptual estimates to an extent, a novel solution, *multi-way* approach, is adopted along with traditional *one-step ahead* approach. It is hypothesized that estimating building element quantities and later employing them as additional inputs along with meagre information through establishing the basis of a project, multi-way approach may influence a *substantial* decrease in prediction error when compared to conventional, one-step ahead approach. Four alternative frameworks of cost estimation and two alternative frameworks of duration estimation are considered, adopting alternative methods of analysis.

Results reveal that adopting multi-way approach, approximately 19% improvement in prediction accuracy can be accomplished over one-step ahead approach, when the aim is predicting cost of structure. The increase in prediction accuracy is 10%, when duration estimation is considered. However, it is underlined that formal statistical test denies the differences as *substantial* at 5% significance level due to insufficient size of the test sample. Results of analysis also demonstrates, linear regression models provide slightly smaller prediction errors than neural network models. What is more to the point, linear regression models offer significantly lower disperse in error of predictions; and therefore preferred over neural network models. Lastly, confidence intervals of linear regression estimations are computed on a case project to demonstrate application of range estimates. For predictions, like the ones relevant to this research, the possible prediction error inherent in the process itself and therefore the security of the prediction has to be kept in mind.

The models developed within the content of this research relies on objective assessment and scientific methodology and therefore can be regarded as highly reliable for practical implementation by German architects aiming to predict cost of structure and construction duration through establishing the basis of a project. New insights to the German architects are offered by proposing alternative relevant factors in developed models. These factors can be taken into account to predict, monitor and maintain budget and schedule. More to the point, along with observed increase in prediction accuracy compared to conventional practice, employing multi-way approach German architects are able to provide expected average values of building element quantities and hierarchical cost groups according to relevant German standards.

Keywords: Germany; Modeling; Predictions; Cost of structure; Construction duration; Linear regression; Artificial neural networks; Multi-way forecasting

Zusammenfassung

Baukosten und Ausführungsdauer sind neben der Qualität zwei entscheidende Einflussfaktoren für Bauprojekte. Die objektive Schätzung dieser Faktoren ist daher bei der Grundlagenermittlung eines Projekts unabdingbar, um als Basis für Budgetierung, Planung, Ausführung, Bauüberwachung und sogar Rechtsstreitigkeiten zu dienen. Die vorliegende Arbeit untersucht Neubauprojekte in Deutschland als entsprechende Datengrundlage mit zwei vorrangigen Zielen.

Um eine objektive Grundlage für frühzeitige Schätzungen aufzubauen, werden Informationen von fertiggestellten Projekten empirisch untersucht. Dabei kommen zwei alternative Analysemethoden zum Einsatz, die multiple lineare Regression und künstliche neuronale Netze. Daneben werden Prognosemodelle entwickelt, die als objektive Grundlage in der Praxis Anwendung finden können.

Um die Vorhersagegenauigkeit der frühzeitigen Schätzungen möglichst zu erhöhen, wird als neuartige Herangehensweise ein mehrstufiger Ansatz neben dem traditionellen einstufigen Ansatz verfolgt. Es wird angenommen, dass der mehrstufige Ansatz, dass heißt die Schätzung der Elementmengen der Gebäude und anschließende Verwendung dieser Ergebnisse als zusätzliche Inputfaktoren neben einigen weiteren Informationen bei der Grundlagenermittlung eines Projekts, im Vergleich zum konventionellen einstufigen Ansatz eine erheblichen Verringerung des Prognosefehlers bewirken kann. Vier alternative Schemata für die Schätzung der Baukosten und zwei alternative Schemata für die Schätzung der Ausführungsdauer werden dazu anhand zweier alternativer Analysemethoden betrachtet.

Die Ergebnisse belegen, dass bei Anwendung des mehrstufigen Ansatzes im Vergleich zum einstufigen Ansatz bei der Prognose der Bauwerkskosten eine um etwa 19% höhere Vorhersagegenauigkeit erreicht werden kann. In Bezug auf die Abschätzung der Ausführungsdauer beträgt die Erhöhung der Vorhersagegenauigkeit etwa 10%. Es wird jedoch betont, dass ein formaler statistischer Test diese Differenzen aufgrund der unzureichenden Größe der Teststichprobe bezogen auf das 5% Signifikanzlevel nicht als signifikant bewertet. Die Ergebnisse der Analyse zeigen auch, dass die Modelle der linearen Regressionsanalysen geringfügig kleinere Prognosefehler liefern als die der neuronalen Netze. Wesentlich ist weiterhin, dass die Modelle der linearen Regressionsanalyse eine erheblich geringere Streuung beim Vorhersagefehler liefern und aufgrund dessen den neuronalen Netzen vorgezogen werden können. Abschließend werden die Konfidenzintervalle der linearen Regressionsanalysen anhand eines Beispielprojektes errechnet, um die Anwendbarkeit der Bandbreitenschätzung zu veranschaulichen. Bei den Vorhersagen dieser Forschungsarbeit darf der mögliche prozessimmanente Vorhersagefehler und daraus folgend die Zuverlässigkeit der Vorhersage nicht außer Acht gelassen werden.

Die innerhalb dieser Forschungsarbeit entwickelten Modelle stützen sich auf objektive Datenerhebungen und wissenschaftliche Methoden und können daher als sehr zuverlässig für die praktische Umsetzung durch deutsche Architekten betrachtet werden, wenn es um die Abschätzung der Bauwerkskosten und Ausführungsdauer im Rahmen der Grundlagenermittlung eines Projekts geht. Durch die Darstellung alternativer, relevanter Einflussfaktoren in den entwickelten Modellen werden den deutschen Architekten neue Erkenntnisse geboten. Die Faktoren können für die Planung, Steuerung und Kontrolle von Kostenbudget und Zeitplan herangezogen werden. Wichtiger noch, neben dem ermittelten Anstieg der Vorhersagegenauigkeit im Verhältnis zur herkömmlichen Praxis, sind deutsche Architekten bei Anwendung des mehrstufigen Ansatzes in der Lage, die zu erwartenden Durchschnittswerte für Elementmengen von Gebäuden und hierarchischen Kostengruppen nach den einschlägigen deutschen Normen zu ermitteln.

Schlüsselwörter/Stichwörter: Deutschland; Modellierung; Prognose; Bauwerkskosten; Ausführungsdauer/Bauzeit; lineare Regression; künstliche neuronale Netze; mehrstufige Vorhersage

Contents

Table of Contents vii

List of Abbreviations ix

List of Figures xi

List of Tables xiii

Abstract xv

Zusammenfassung xvii

1 Introduction 1
- 1.1 Research Background 1
- 1.2 Problem Statement 2
- 1.3 Scope, Aims, and Hypotheses 6
  - 1.3.1 Research scope 6
  - 1.3.2 Research aims 7
  - 1.3.3 Frameworks and hypotheses 7
  - 1.3.4 Research objectives 11
- 1.4 Structure of the Manuscript 12

2 Literature Review 15
- 2.1 Construction Cost Estimation 16
  - 2.1.1 Overview 16
  - 2.1.2 Key Studies 16
  - 2.1.3 Conclusion 22
- 2.2 Construction Duration Estimation 23
  - 2.2.1 Overview 23
  - 2.2.2 Bromilow's time-cost model and related studies 24
  - 2.2.3 Critics on BTC model and other models 27
  - 2.2.4 Conclusion 30

3 Methodology 33
- 3.1 The Empirical Work 33
  - 3.1.1 A General outlook 33
  - 3.1.2 Motivation 34
  - 3.1.3 Modeling 35
  - 3.1.4 Quantitative approach 35

3.2 Linear regression 36
3.2.1 Overview 36
3.2.2 The standard regression assumptions 38
3.2.3 Detection of violations in model assumptions 40
3.3 Artificial Neural Networks 41
3.3.1 Overview 41
3.3.2 Multi layer perceptrons 41
3.3.2.1 The network architecture 42
3.3.2.2 Activation function 43
3.3.2.3 Training algorithm 44
3.3.2.4 Data normalization 44
3.3.2.5 Training and test sample 45
3.3.2.6 Performance measures 45
3.3.3 General regression neural network 45
3.4 Concluding Remarks 47

4 Sample 49
4.1 Overview 49
4.2 Restrictions and Predictor Variables 51
4.3 Response Variables 53
4.4 Supporting the Hypotheses - Pilot Analysis 54
4.5 Concluding Remarks 55

5 Results of Analysis 57
5.1 Reference Unit Quantities 58
5.2 Cost of Structure 65
5.2.1 Second level cost groups of structure 65
5.2.2 First level cost groups of structure 80
5.2.2.1 Multi-way forecast 81
5.2.2.2 One-step ahead forecast 85
5.3 Construction Duration 89

6 Comparison of the Frameworks 93
6.1 Cost Estimation Frameworks 93
6.2 Duration Estimation Frameworks 101

7 Implementation 107
7.1 Story of 1300-082 107
7.2 Point Estimates 108
7.2.1 Reference unit quantities 109
7.2.2 Cost of structure 112
7.2.3 Construction duration 117
7.3 Confidence Intervals 119

8 Conclusions 125

References 131

1 Introduction

1.1 Research Background

> *"The physical substance of a house is a pile of materials assembled from widely scattered sources. They undergo different kinds and degrees of processing in large numbers of places, requires many types of handling over periods that vary greatly in length, and use the services of a multitude of people organized into many different sorts of business entity."*

Cox & Goodman (1956) underlined characteristics of the construction industry approximately 60 years ago in a widely known study of the distribution of house building materials. One of the main conclusions is that the number of possible permutations and combinations of specific places and entities is numerous, even for one product. Similarly, the complexity of the industry have been emphasized more recently. Winch (1987) claimed that construction projects are amongst the most complex of all undertakings. Gidado (1996) further remarked and suggested that there is a continuous increase in the complexity of construction projects. The industry's way of functioning and its performance are shaped by these underlying conditions (Dubois & Gadde, 2002). Now and then construction industry participants are blamed for inefficiency of operations (Cox & Thompson, 1997). A number of authors have argued that construction has failed to adopt techniques that have improved in other industries, such as just-in-time (Low & Mok, 1999), total quality management (Shammas-Toma *et al.*, 1998), partnering with suppliers (Cox, 1996), supply chain management (Vrijhoef & Koskela, 2000), and industrialization of manufacturing processes (Gann, 1996). Main properties of construction industry shall be described to reveal why this figure is so.

Cox & Thompson (1997) have argued that construction is inherently *a site specific project-based activity*. Parallel, Shirazi *et al.* (1996) concluded that construction is mainly about coordination of specialized and differentiated tasks at the site level. Gidado (1996), on the other hand, suggested complexity in construction systems caused from interdependence among tasks, and represents those sources of complexity that originate from different parts together to form a work flow. Accordingly, Dubois & Gadde (2002) suggested two central features of construction: (1) the focus on individual projects, in terms of decentralized decision making and financial control, (2) the need for the local adjustment at the construction site. They argued local adjustments are necessary because of the three remaining uncertainty factors: (1) lack of complete specification for the activities at the construction site, (2) lack of uniformity of materials, work, and teams with regard to place and time *(every project is unique)*, and (3) *an unpredictable*