Journal of Engineering, Project, and Production Management
2014, 4(1), 26-35

# Estimating Construction Duration for Public Roads During the Preplanning Phase

Agata Czarnigowska[1] and Anna Sobotka[2]

[1]Assistant Professor, Faculty of Civil Engineering and Architecture, Lublin University of Technology, 20-618 Lublin, Nadbystrzycka 40, Poland. E-mail: a.czarnigowska@pollub.pl (corresponding author).

[2]Professor, Department of Geomechanics, Civil Engineering and Geotechnics, AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Krakow, Poland. E-mail: sobotka@agh.edu.pl

Project Management
Received May 12, 2013; revision August 28, 2013; accepted August 30, 2013
Available online December 15, 2013

---

**Abstract:** Estimates of the construction time are of key importance in the early phases of the project – they serve as a basis for the decision whether to commence with planning or not, and are used as input for budgets and programmes. Usually, such estimates base on experience with similar projects completed in the past. This experience may be recorded in the form of mathematical models that relate project characteristics to construction time. The aim of the research was, basing on real-life cases, to develop a model of public road building projects duration. The research comprised collection of input for the analyses, preselection of project features correlated with duration, and construction of three models: a simple statistical regression, a multifactor regression and a regression tree. The models were then compared to each other and to the models presented in the literature with respect to their predictive ability. With the assumed set of potential predictors of construction duration, the regression models were found statistically correct, though not precise enough be used as decision-support tools.

**Keywords:** Construction duration estimates, regression tree, multiple regression.

---

## 1. Introduction

The literature on planning construction projects with respect to time focuses mostly on tools and techniques related with detailed scheduling problems of on-site operations. The objects of researchers' interest are, among others, improvement of network techniques, schedule optimisation and schedule reliability improvement (e.g. Jaśkowski and Sobotka, 2012; Biruk and Jaskowski, 2010, Ko and Chen, 2012, Liu and Wang, 2012). However, before any scheduling can be conducted, the project constraints have to be defined, and the key one is the overall duration of the construction phase. It needs to be assumed early in the project preparation stage to serve as a basis of any future planning, feasibility checks, and negotiations with project participants. The literature on the methods of defining the required construction duration at early stages of project planning is scarce. The estimates base on a compromise between the client's expectations towards completion date and on "technical" feasibility of the construction makespan. The latter comes from the planner's experience with similar projects.

In the case of early cost estimates, there exist an abundance of methods – from intuitive case-based reasoning to complex mathematical modelling – that use cost records of completed projects to predict cost of new schemes. Examples of parametric cost models can be found e.g. in Cheung and Skitmore (2006) and ISPA (2008). With the development of information technology, non-parametric cost models gain on popularity: neural networks were used e.g. by Adeli and Wu (1998), Elhag and Boussabaine (1999), Leśniak (2004), Juszczyk (2008), and simulation techniques by Chau (1995) or Lai et al. (2008). Interestingly, databases of construction schedules, or even records of overall construction duration, are less common. The models of project duration based on historical data are also quite rare in the project management literature.

The paper investigates into the duration of public road projects and the potential of utilising data of real-life cases from the past to forecast durations of similar projects in the near future by means of simple regression models.

## 2. Regression Models of Construction Duration

Cost is a generalized measure of any project's scale and complexity. The existence of a relationship between construction time and cost has been considered obvious: the time-cost-performance triangle appears in practically all project management handbooks (e.g. Kerzner, 1984). Assuming that a reliable estimate of the project cost is possible to be made at early stages of planning, the cost may be considered known at the moment when project duration is to be decided. This rather optimistic assumption was the foundation of numerous models that could be used for predicting project duration on the basis of project cost. The first time-cost regression model of construction projects is attributed to Australian researchers who, having analyzed cost and duration of a sample of construction projects completed during late 1960ies, proposed the following model, later referred to as the Bromilow's time-cost model (Kaka and Price, 1991):

$$L = K \cdot C^B, \tag{1}$$

or its equivalent:

$$\ln L = \ln K + B \cdot \ln C, \tag{2}$$

where $L$ is the number of working days from the contractor's possession of the building site to the completion of works; $C$ – actual value of works as paid by the client, expressed in A$ million; $K$ and $B$ – constants.

Bromilow's findings were checked by other researchers on the basis of new samples (Kaka and Price, 1991; Chan, 2001; Yousef and Baccarini, 2001; Ogunsemi and Jagboro, 2006). The form of the time-cost function (1) was confirmed to match sample data better than other function types tried, though determination coefficients obtained by the authors were low (for large samples of non-uniform projects below 0.75). Large yearly fluctuations of the constants B and K were reported, though without any particular trend (Skitmore and Ng, 2001).

Statistical significance of the time-cost relationship gave rise to numerous attempts to create a multifactor regression model of construction duration that would incorporate project qualities other than cost and provide a better fit than the Bromilow's model. Table 1 provides an overview of selected findings presented in the literature that regard factors correlated with construction duration and duration models, where $L$ stands for construction duration expressed in days, and $b_i$ are constants. Generally, there was no agreement on what factors should be the basis for estimating the duration. With few exceptions (Skitmore and Ng, 2003, Love et al., 2005, Stoy et al., 2007), cost was usually considered the most important independent variable present in multifactor models. The models presented in the literature were claimed to be statistically correct and significant. However, the authors often came to contradictory conclusions: some found that e.g. the client's sector (public/private), building function or size strongly affected the construction duration, others excluded them as insignificant.

The initial selection of factors considered was also a matter of assumption, as the models were not always aimed at duration predictions – some were by-products of search for factors correlated with duration, some were used for measuring the project time performance. Some researchers focused on management factors, other preferred more "tangible" qualities, either known well ahead of commencement with works, or possible to be determined only after the project was finished.

The log-log relationship between time and cost in these multifactor models was widely argued to provide best fit, though some different functions were also proposed (Stoy et al., 2007; BCIS, 2004a; Martin et al., 2006; BCIS 2009). The authors were rarely specific about the quality measures of their models. The prediction and confidence intervals for the estimates can be found only in Stoy et al. (2007), BCIS (2004a) and BCIS (2009). Naturally, the larger and more diversified the samples, the greater errors were observed.

Most researchers analysed projects related with construction of buildings, so there are only a few works devoted to civil engineering projects. Kaka and Price (2001) analysed 140 UK road projects and found that the form of contract (fixed price vs. adjusted price) affects strongly the Bromilow's time-cost model parameters. Yousef and Baccarini (2001) conducted similar work on the basis of 46 sewerage projects in Australia, but did not considered factors other than costs. Irfan et al. (2011), disposing of large samples, focused on highway projects and created separate regression models for different project types (maintenance, resurfacing, construction, bridge construction, traffic infrastructure) that used planned cost and contract type as predictors of duration.

Most authors claimed that it was possible to apply regression models to estimating construction time on the basis of cost, so they assumed that it was easier to estimate construction cost than construction time, and that the cost estimates were accurate enough to provide the basis for time estimates. However, there is an abundance of evidence on discrepancies between early budgets and costs at completion. Case studies focus on most striking examples (Potts, 2005; Polonski, 2006; Magnussen and Olsson, 2006), but there exist statistical overviews of the scale and frequency of cost miscalculations: quite alarming by Flyvbjerg et al (2002), and a number of less pessimistic (Ng et al. 2001; KPI UK, 2003; BCIS, 2004b; KPI New Zealand, 2005). Another issue is the reliability of the winning-bid price as a measure of the project scope and scale. There is evidence that contractors' bids are sensitive to intensity of competition, subjective risk perception and even season of the year when a call for tenders is announced. This can be observed in bid spreads in public procurement procedures. In Poland, they are expressed by an average bid dispersion factor $W_z$ (Borowicz, 2005):

$$W_z = \frac{1}{n}\sum_{i=1}^{n}\frac{C_{i\max}}{C_{i\min}}, \tag{3}$$

where $n$ is the number of tender procedures investigated, and $C_{imax}$ , $C_{imin}$ are, respectively, the highest and the lowest bid in each procedure. For instance, for the years 2000-2007, the average bid dispersion factors of public projects in Poland were from 1.23 to 1.43, and bid dispersion in particular cases reached even 250% (Borowicz, 2005 and 2008). Thus, the relationship between the contract price and actual value of works may be rather loose.

Under these circumstances, there are reasons to question both the "as-planned" and "contractual" cost in the role of independent variable for planning construction time. Moreover, it occurs that predictability of cost is generally no better than predictability of time (Martin et al., 2006; KPI UK, 2003). This provides more arguments against using cost as a predictor of time. However, before it is rejected, it would be interesting to check the model's sensitivity to cost miscalculation. In general, values of the constant $B$ in the models presented in the literature range from 0.2 to 0.5, and the smaller $B$, the smaller the effect of cost on the value of the time estimate (see formula 1). The time-cost models are thus not very sensitive to the cost estimate errors.

**Table 1.** Selection of multifactor models (selected works), $L$ – construction duration, $b_i$ – constants.

| Author, sample | Significant factors | Regression function, quality measures |
|---|---|---|
| Kaka and Price 1991,<br>UK<br>661 buildings from BCIS database,<br>140 civil engineering projects | cost<br>client's sector (public/private)<br>contract type (fixed-price/other)<br>procurement route (open tender/restricted tender/negotiations)<br>project type (building/civil eng.) | simple time-cost regression:<br>$\log L = b_0 + b_0 \log C$<br>separate equations for sub-sets divided according to significant factors other than cost<br>assumed that building function does not affect model parameters |
| Walker, (1995),<br>33 non-residential buildings, new-built | cost ($C$)<br>ratio of time extension ($x_1$)<br>scope (fit-out/other) ( $x_2$)<br>level of quality requirements ($x_3$)<br>management style ($x_4$)<br>design and construction teams communication quality ($x_5$)<br>efficiency of IT use ($x_6$) | $\log L = b_0 \log C + b_1 x_1 + b_2 x_2 + \ldots + b_6 x_6$<br>determination coefficient $R^2$=0.9987<br>and percentage errors of estimate (for particular observations, not summarised) |
| Chan and Kumaraswamy, (1999),<br>56 blocks of flats of the same standardised design system, the same public client, | different set of factors describe as-planned and actual duration; for actual duration $L$:<br>cost ($C$)<br>project type (flats for sale / rent) ($x_1$)<br>facade type (prefab or other) ($x_2$)<br>volume of the building ($x_3$)<br>gross floor area ($x_4$)<br>number of storeys ($x_5$) | $\ln L = b_0 + b_1 \ln C - x_1 + x_2 + b_3 \cdot x_3 - b_4 \frac{x_4}{x_5}$<br>percentage errors of the estimate of $L$ for each observation, maximum errors of +/-7%, the model's *MAPE*=2,51% |
| Skitmore i Ng (2003)<br>Australia,<br>93 buildings, new | contractual time (*CT*)<br>contract type (lump sum / other) ($x_1$)<br>procurement method ($x_2$)<br>cost excluded from the analysis | $\log L = b_0 + b_1 \log CL + x_2 + x_3$,<br>Adjusted determination coefficient $R^2$=0.938 |
| Love et al. (2005)<br>Australia.<br>126 buildings, new or refurbished | usable floor area ($x_1$)<br>number of storeys ($x_2$)<br>cost excluded from the analysis, as cost not known until the project is finished | $\log L = b_0 + b_1 \log x_1 + b_2 \log x_2$,<br>Adjusted $R^2$=0.96<br>*MAPE*=50% |
| Stoy et al. (2007),<br>Germany,<br>200 buildings from BKI database, 16 buildings for verification | gross floor area ($x_1$)<br>number of winters ($x_2$)<br>project planning time in months ($x_3$)<br>dependent variable is logarithm of construction speed,<br>cost excluded | $\ln \frac{x_1}{L} = b_0 + b_1 \ln x_1 - b_2 x_2 - b_3 \ln x_3$,<br>Adjusted $R^2$=0.915, *MAPE*=20%,<br>Errors of estimate for test sample projects range (-29%;9%) |
| Hoffman et al. (2007),<br>USA,<br>616 military buildings, new or refurbished | cost ($C$)<br>client type ($x_2$-$x_4$)<br>management (own / contr. out) ($x_5$)<br>design type (own / contr. out) ($x_6$) | $\ln L = b_0 + b_1 \ln C - b_2 x_2 - b_3 x_3 +$<br>$- b_4 x_4 - b_5 x_5 - b_6 x_6$<br>Adjusted $R^2$=0.374 |
| BCIS (2004a)<br>UK,<br>1500 new buildings, from KPI database | cost ($C$)<br>procurement system ($x_3$)<br>contractor selection ($x_4$)<br>client type ($x_5$)<br>function ($x_6$)<br>region ($x_7$) | $\sqrt{L} = b_0 - b_1 \log C + b_2 \log^2 C + b_3 x_3 +$<br>$+ b_4 x_4 + b_5 x_5 + b_6 x_6 + b_7 x_7$ |
| BCIS (2009),<br>UK,<br>4500 buildings, new or refurbished, BCIS database | factors as above<br>Separate regression functions for new-builts and refurbishments<br>Winter period does not affect duration significantly | equation as above<br>the "calculator" provides prediction and confidence intervals for the estimates; separate equations for new-built and refurbishment projects |

## 2. Methods

### 2.1. Aim and Scope of Research

The aim of the research was to create a model of the road construction duration based on relationships between the project qualities. These relationships were assumed to be determined while analysing project qualities likely to be known or estimated at early planning stages, without consideration of organisation of works or construction methods. The model could be applied to estimating construction duration at the stage of feasibility checks.

For the purpose of this study, a project was defined as a scope of works contracted in one public procurement procedure and covered by one contract, supplemented with change orders and contract annexes, if applicable. Construction duration was measured as the number of calendar days from the day of commencement with works to the date of completion of the contract. Construction cost was the amount payable to the contractor – reduced or increased by penalties, change orders etc.

Stages of research were as follows: literature review, interviews with the construction clients – to determine their approach to estimating construction durations and budgets at the planning stage (outside the scope of this paper), data collection (by analysing project records – no ready-made databases were available), preselection of project qualities correlated with duration by means of the regression tree, construction of regression models, and finally assessing quality of the models.

Three types of models were selected for comparison:

1. simple regression (parameters calculated by means of the least squares method) between functions of actual duration (L) and actual cost (C): $f(L) = b_0 + b_1 f(C)$, where $b_0$ and $b_1$ are model parameters,

2. CART regression tree based on 25 project qualities recorded at data collection stage and likely to be known at early stage of project planning;

3. multiple linear regression model (parameters calculated by means of the least squares method) relating a function of actual duration ($L$) and functions of predictors selected from the same project qualities as used to construct Model 2, $f(L) = b_0 + b_1 f(x_1) + b_2 f(x_2) + \ldots + b_n f(x_n)$, where $b_0$ – $b_n$ are model parameters, and $x_1$ – $x_n$ are predictors.

Calculations were conducted by means of Statistica 8.0.

### 2.2. The Sample

The sample comprised 100 public road projects, completed between 2003 and 2008 in three neighbouring regions in south-eastern Poland. The projects considered differed in scope and type (Fig. 1), and their cost ("as planned", including VAT) ranged from PLN 800 thousand to PLN 500 Million. The sample was considered representative of road projects from the analyzed period and location, and its size was at least 15% of the size of the population (imprecise due to non-uniform reporting methods used by the public clients).

One of the early assumptions of the research was to analyse projects of one kind, such as new circular roads. As occurred during the data collection process, the number of such projects was too small to be used for statistical analyses, and the majority of works contracted in the analyzed period consisted in modernization of the existing infrastructure. Therefore, projects varying in scope and type were included in the sample. Their similarity consisted in overall conditions: the clients were from public sector and acting under similar budgetary constraints, the regions were similar in terms of natural and economic environment and level of infrastructure development, the works were contracted according to the public procurement law, the only criterion of contractor selection was the lowest price, and contract duration was enforced by the client. A diversified sample implies that the model derived from the data would be a far going generalisation.

Prior to construction of the models, the cost and time predictabilities of the cases included in the sample were compared to check if cost was not significantly less predictable than duration (if so, duration estimates based on cost would be questionable). Fig. 2, compares these predictabilities.

Predictability of construction cost at planning stage was defined as the difference between the actual cost $C$ and the as-planned cost $C_{PL}$ assumed by the client, expressed as a percentage of the as-planned cost, in accordance with the definition of Cost Predictability – Construction presented in KPI UK (2003):

$$PC_{PL} = \frac{C - C_{PL}}{C_{PL}} \cdot 100\% \qquad (4)$$

Predictability of construction cost at contract signing was defined as the difference between the actual cost $C$ and the contractual cost $C_C$, expressed as a percentage of the contractual cost. Predictability of construction duration was defined as the difference between the actual duration $L$ and the as-planned (contractual) duration $L_{PL}=L_C$, expressed as a percentage of the planned duration.

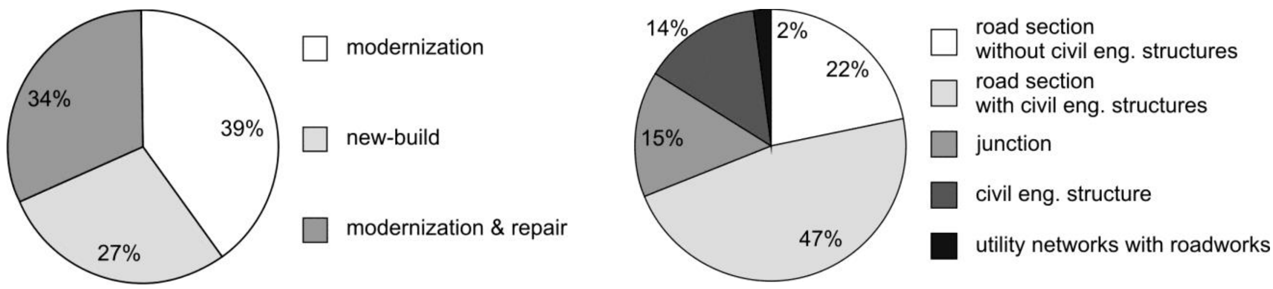


**Fig. 1.** Sample structure according to project type and scope of works

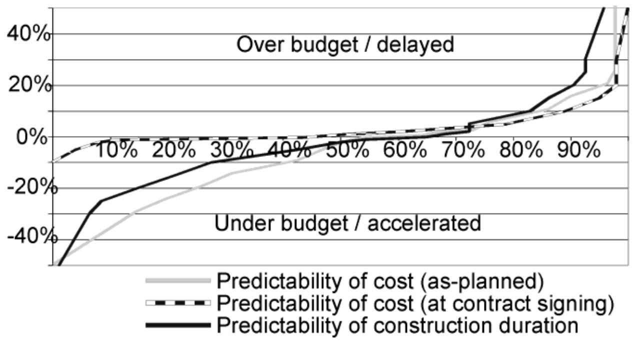


**Fig. 2.** Duration and cost predictabilities in the sample

Actual cost exceeded the range of planned cost by +/- 10% in about 50% of cases, while actual duration occurred to be slightly more predictable – time miscalculation greater than 10% of the planned duration took place in about 45% of cases. However, at contract signing, the predictability of cost was evidently better than that of duration: only 20% of cases were outside the contractual cost +/- 10% brackets. This may be due to the client's preference on fixed-price contracts (33% of cases) and restrictions on public spending.

One can conclude that, for predictability reasons, both the planned (the client's estimate) and the contractual cost may serve as a predictor of duration: the planned costs' predictability is comparable with the predictability of time, and the contractual cost's predictability is better. Thus, there is no reason to question construction cost as a potential predictor of construction duration in the sample.

## 3. Results

### 3.1. Simple Linear Regression Model

Analysis of scatter diagrams (Fig. 3) and experiments with several functions confirmed thatthe Bromilow's model provides the best fit for the analysed sample, and that it is statistically correct. The model (Formula 5) is significant (F-test) and of significant parameters (t-tests), the residuals are normally distributed, with constant variance and expected value of 0. Normality was checked by analysing residual histograms, scatter diagrams, and by normality tests: Kolmogorov-Smirnov/Lilliefors' and Shapiro-Wilk's. Homoscedasticity of residuals was checked by analyzing residual scatter diagrams, and by Lagrange test (Stanisz 2007). The Bromilow's model for the sample (Model 1) is described by the following equation:

$$\ln L = 1{,}2067 + 0{,}4749 \cdot \ln C \qquad (5)$$

The model's adjusted determination coefficient $\overline{R}^2 = 0.636$, and standard error SEE=0.504. The mean absolute percentage error *MAPE*=44,84% is comparable with the scale of errors of time-cost models using logarithm transformation of duration, presented in the literature.

### 3.2. Regression Tree

While collecting input, data on 25 project qualities were collected. These qualities were considered likely to be known at early stages of project planning, and were of various types: categorical and quantitative, related with geometric parameters of the road, scope of works, road class, location, number of bridges and many more (listed in Fig. 5). All these potential predictors were used to construct regression tree (*CART*) models. The method consists in recursive division of the set of observations into subsets (two subsets at a time), according to one quality at a time, to obtain the greatest possible reduction of heterogenity of observations in the subset (Gatnar, 2001). Here, the heterogeneity was measured by the variance of durations of projects in the subset. The best tree was selected according to Breiman's procedure (Gatnar, 2001). The best-fit model (Model 2), presented in Fig. 4, uses seven predictors: assumed number of winters during construction, construction cost, number of culverts along the route, client type (either national or regional road office), total length of civil engineering structures, number of intersections, number of parking/bus bays. If applied in practice, it would assign a project one of ten durations: 91, 161, 166, 291, 396, 487, 490, 573, and 810 days. Some of them differ by only a few days.

The model's adjusted determination coefficient $\overline{R}^2 = 0.924$ (Gatnar, 2001) indicates that the model is well fitted to the sample. Considering the relatively small number of observations used to create the model, and their being diverse, this is not automatically an advantage when it comes to using the model for predictions.

There may be doubts about using the number of winters as a predictor of construction duration, as hard to estimate as the duration itself. However, interviews with the client's representatives indicated that the clients decided to fit a project in a certain number of years at the beginning of the project planning process, which arised from budgetary constraints and long-term planning of public organisations. Therefore, the number of winters can be considered defined in advance. Selection of this variable was prompted also by other research (Stoy et al 2007; BCIS 2004a) – where it was to allow for seasonal changes in speed of works.

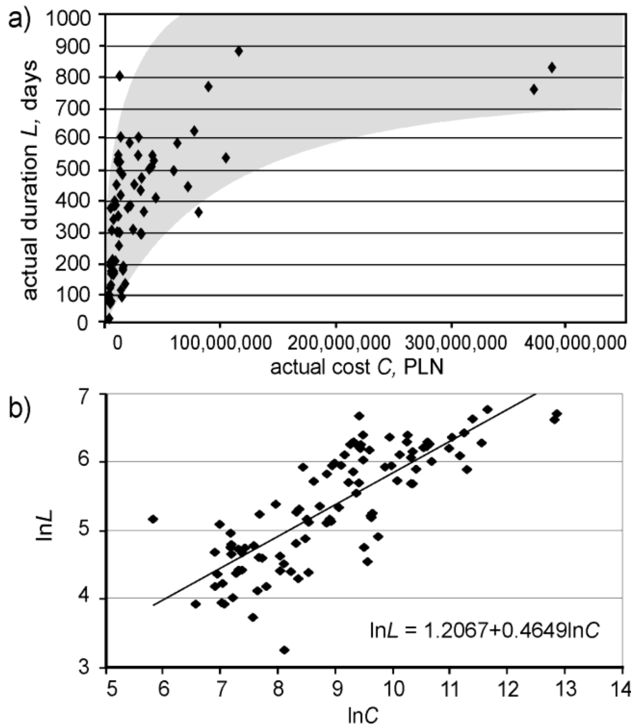


**Fig. 3.** Scatter diagram of actual construction duration $L$ against actual cost $C$ (a), and scatter diagram of log-log values with regression line (b)

### 3.3. Multiple Regression Model

It was assumed that a linear multiple regression model would be looked for, and its parameters were to be determined by the least squares method. With only 100 cases in the sample, using the stepwise regression to select the most suitable of 25 potential predictors (or actually over 40, as categorical variables were converted into binary variables) was considered inefficient. However, while constructing regression trees, one can identify variables that are potentially strongly correlated with the predicted variable, but not necessarily present in the best regression tree (Gatnar, 2001). Fig. 5 presents the relative importance of the potential predictors, determined in the procedure of constructing regression trees.

For further investigations, nine potential predictors were selected arbitrarily: six of the "most important" defined in the *CART* analysis (Fig. 5), and additionally those present in the best regression tree.

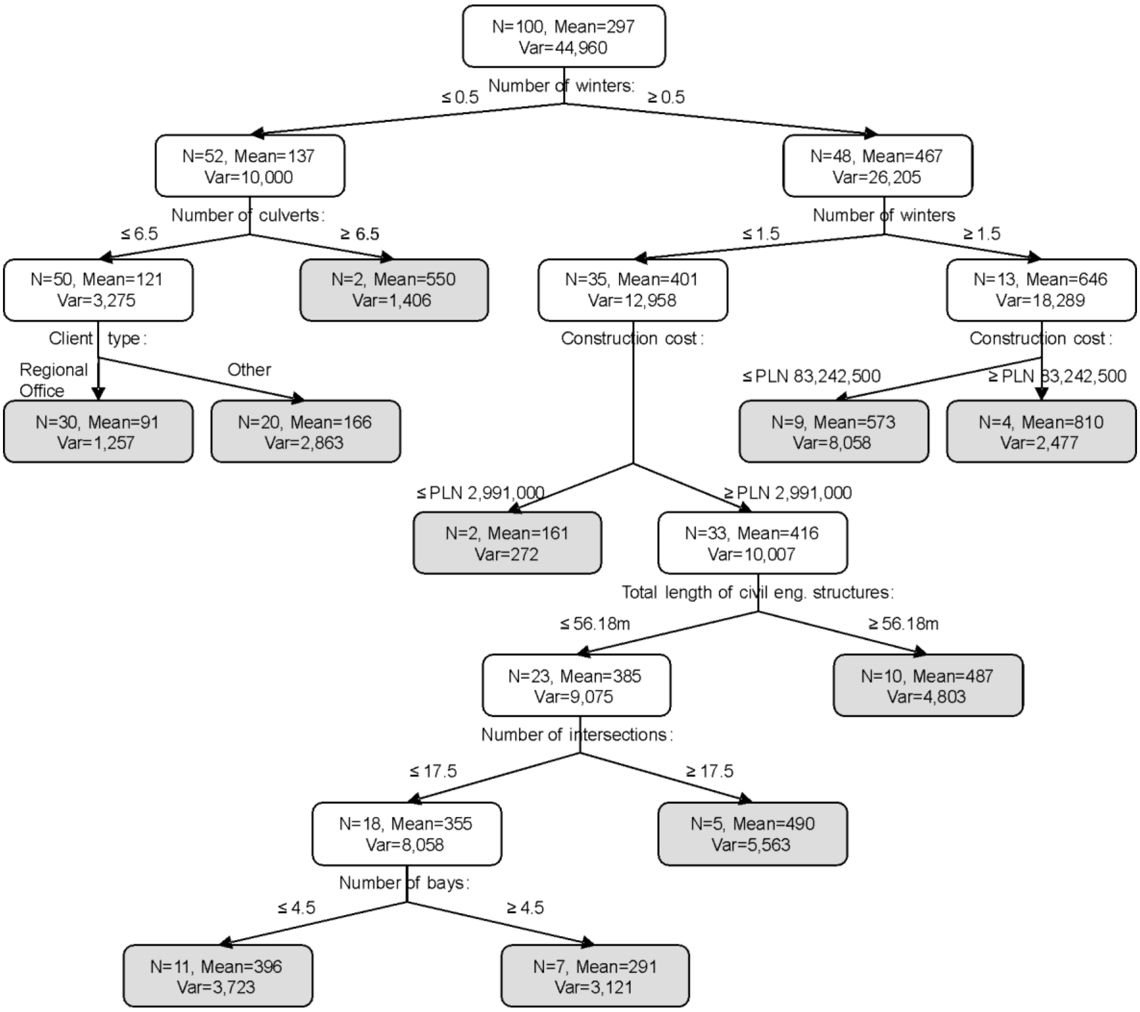


**Fig. 4.** Model 2 – regression tree

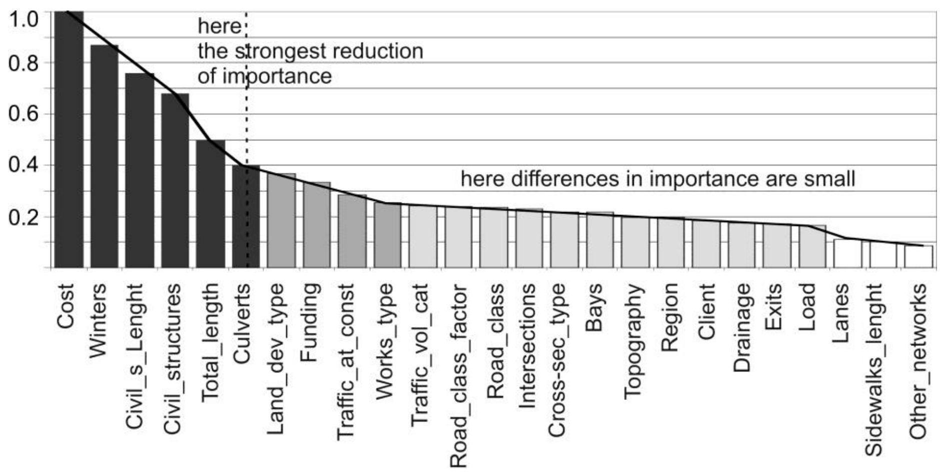


**Fig. 5.** Relative importance of variables defined while constructing regression trees

These nine factors were then used for constructing regression models by means of stepwise regression (forward selection and backward elimination):

- construction cost,
- number of winters,
- length of civil engineering structures in the scope of a project,
- number of civil engineering structures,
- total length of roads covered by the project,
- number of culverts along the route,
- number of bays,
- number of intersections,
- client type (either regional or national road agency)

Several models were tried, differing in transformations of variables. The best fitted was Model 3, with four predictors: cost ($C$), civil engineering structures length (*Civil_s_length*), number of winters (*Winters*), and number of civil engineering structures (*Civil_s*):

$$\sqrt{L} = -4.43 + 1.89\ln C + 0.56\ln Civil\_s\_length + \\ + 4.44Winters - 0.28Civil\_s \quad (6)$$

The model fulfils the assumptions of the least squares method, $\overline{R}^2 = 0.867$ , *MAPE*=13,17%, *SEE*=2.28. Considering the parameters of Equation 6, one can observe that the estimate is very sensitive to the number of winters that can be hard to asses.

### 3.4. Quality of the Models

The models use different transformation of predicted value (ln$L$ in the case of simple regression, $L$ for regression tree, and $\sqrt{L}$ for multifactor regression). Due to this fact, the statistics of adjusted determination coefficient $\overline{R}^2$ , standard error *SEE*, or mean absolute percentage error *MAPE* cannot be directly compared. As the models are meant to be used for predicting duration expressed in days ($L$), errors expressed in days were calculated ($L_i$ is the observed duration, and $\hat{L}_i$ – duration calculated on the basis of the model):

$$e_i^{days} = L_i - \hat{L}_i \quad (7)$$

Analysing them, one can see the scale of dispersion between expected vs. observed – for the set of observations used to build the models. Values of these errors, and the mean absolute percentage error, $MAPE^{days}$ :

$$MAPE^{days} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{e_i^{days}}{L_i}\right|, \quad (8)$$

are directly comparable, though not normally distributed.

Fig. 6 compares the scale of errors in days for all models considered. For practical applications, the best model would be the one of lowest dispersion. In this case, it is the regression tree. Its $MAPE^{days}$ is 23%. Model 1 has $MAPE^{days}$ of 45%, and Model 3 – 28%.

To compare the models' predictive ability, duration estimates (in days) were calculated for seven projects not included in the initial sample. Their qualities stayed within the ranges covered by the models. Fig. 7 shows predicted vs. observed values of project durations. Again, for practical applications, the best model would be the one of lowest error.

The test sample is small, which affects reliability of the conclusions. In the case of these particular projects, it is Model 3 that seems to provide most precise predictions, as the observed values are quite close to predicted values. To express it in numbers, one can calculate mean absolute percentage errors in days for the test sample: the "best-looking" prediction model (Model 3) has $MAPE_{test}^{days}$ = 9%, the second-best is Model 1 (Bromilow's) with $MAPE_{test}^{days}$ =21%. The regression tree (Model 2) provides the least accurate estimates with $MAPE_{test}^{days}$ =22%. This is due to the fact that it is too well adjusted to the initial sample, and the test sample simply does not follow the same pattern. The quality of predictions based on regression functions cannot be judged without prediction and confidence limits for the estimates of durations. However, in the case of non-parametric Model 2, there are no grounds to calculate prediction and confidence intervals in a way that could be compared with Models 1 and 3. Fig. 8 presents these data for parametric models (expressed in months for better readability), assuming 95% confidence level.

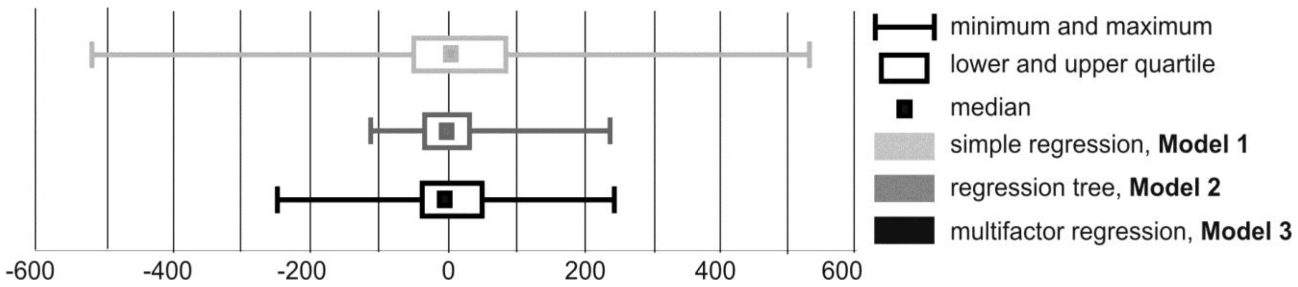


**Fig. 6.** Comparison of model errors expressed in days

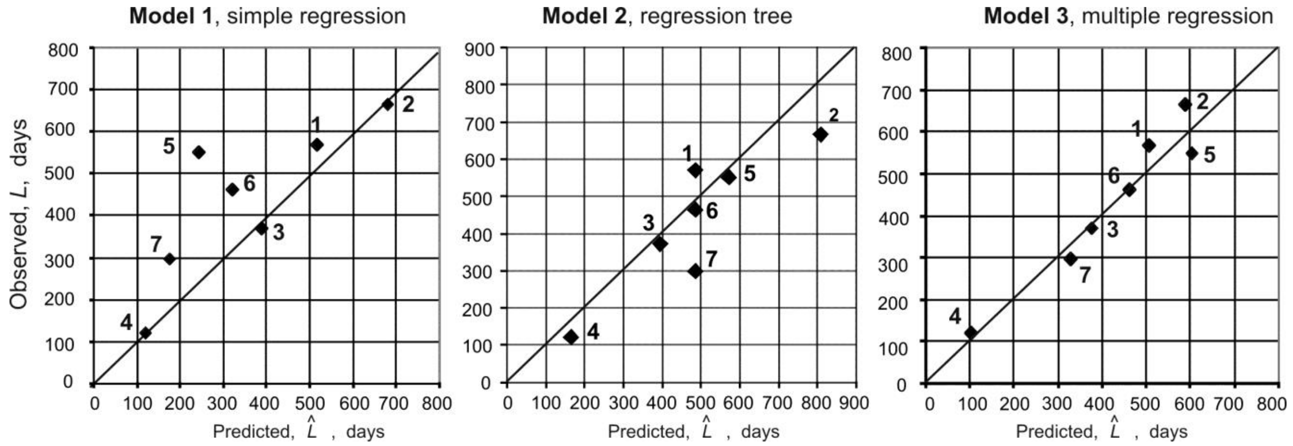


**Fig. 7.** Predicted against observed durations (days) of test sample

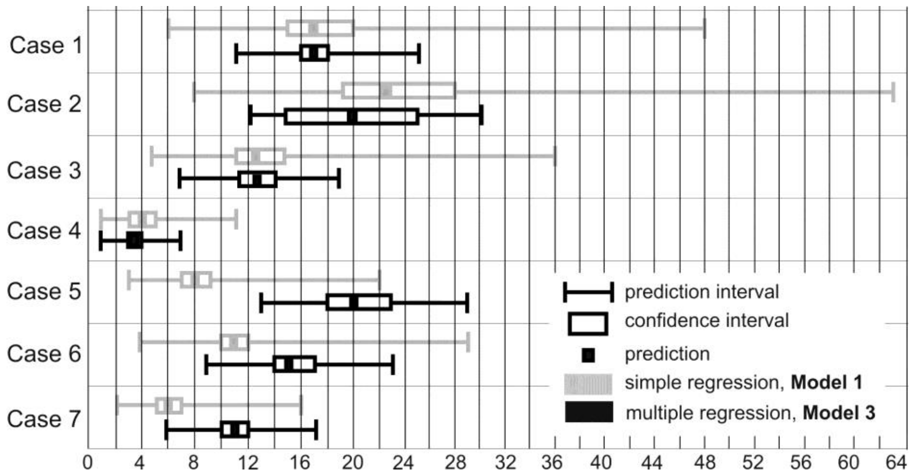


**Fig. 8.** Comparison of confidence and prediction intervals at 95% confidence for Model 1 and Model 3, test sample projects, durations expressed in months

Model 1 seems too inaccurate to find any practical application. This may be illustrated by Case 3 from test sample: the user can be 95% sure that the expected duration for the project of such cost is between 11 and 15 months (confidence interval for regression), but can be also 95% sure that a particular project of such cost may take between 5 and 36 months (prediction interval). The multiple regression Model 3 is certainly more accurate, but confidence and prediction intervals are still quite broad: for Case 3, the confidence limits for the regression are 11 and 14 months, and prediction limits – 7 and 19 months.

## 4. Summary and Conclusions

Developing a model that could be used to estimate the overall duration of construction works on the basis of a few data available at the earliest stages of project preparation is not an easy task. As there are many factors affecting construction duration, and a considerable number of them are related with events and decisions ocurring at later stages, it is hard to expect that such models would be reliable as tools for predictions – they are more to record what is likely to be feasible judging by experience with past projects. The paper was aimed at recording such experience in the case of Polish public road construction projects in the form of regression models. Another aim of the paper was to check if such models could find any practical application.

The sample considered in the paper was small and diversified. However, some statistically significant relationships between construction duration and other project qualities have been found. Calculations confirmed the universal character of the Bromilow's time-cost model proposed in nineteen-sixties. Certainly, the model has some advantages: it is simple and, at least for the considered sample, statistically correct. However, its errors are high (mean absolute percentage error in days is 45%) , and the prediction and confidence intervals impractically broad.

Using a non-parametric method of regression trees, 25 qualities of the analysed project were checked with regard to their relationship with construction duration. Only four of them (the most important according to the non-parametric analysis) stayed in the final multifactor regression model: construction cost, number of winters within the construction period, total length of civil engineering structures (as bridges) in the project, and the number of these civil engineering structures. The predictors are different than these presented in the literature – this is of course specific to the type of projects analysed (the literature focuses mostly on buildings, not roads) and initial assumptions on what factors to consider. What is interesting, the form of a multifactor regression equation most frequent in the literature (Table 1),

$$\ln L = b_0 + b_1 \ln C + b_i \ln x_i + b_j x_j , \tag{9}$$

where $L$ is duration, $C$ is cost, $x_i$ is a continuous variable, $x_j$ represents a discreet variable, and $b_i$ – parameters, did not provide the best fit for the analysed sample. The following equation proved more appropriate:

$$\sqrt{L} = b_0 + b_1 \ln C + b_i \ln x_i + b_j x_j . \tag{10}$$

It is statistically correct, immune to outliers, of lower errors and narrower predictions and confidence intervals, and also not very sensitive to errors of the predictors' estimates, with the exception of the number of winters.

A non-parametric model of regression trees (*CART*) also provides a good fit, though its predictions tested on a small sample occurred less accurate than the predictions of the classic regression models. Interestingly, the regression tree uses a different set of predictors than the multifactor regression model: instead of number of civil engineering structures, there appeared: client type, number of culverts, number of intersections and number of bays. However, with the sample being small (100 observations used to construct the model and 7 to validate it) and diversified, such models are not reliable.

Time-cost regression models for repeatable projects (e.g. buildings of the same function, structure type, similar layout and location) could be more precise. Similarly, if more independent variables were considered, and samples were larger, better models could be provided. A number of researchers report their achievement in this field (Irfan et al. 2011) and there exists at least one commercial regression-based duration "calculator" (BCIS 2009). This may serve as evidence of practical applicability of parametric models in planning construction duration. Such models have some advantage over other models based on experience, such as "black box" expert systems, or neural networks – they are portable: regression models are expressed as equations, and to use them, one does not need to dispose of the whole database or software. Moreover, the reasoning process behind the model is quite obvious. This may be the reason why, in the time of quick development of artificial intelligence methods, statistical analyses do not loose on popularity. Further research in the field may include: investigation on other factors, constructing other model types (here, specification of regression functions was based on results presented in the literature and scatter diagram analyses, and the simplest approach of least squares method was used), and applying artificial intelligence tools to create better models. This however requires expanding the database.

**Acknowledgments**

The authors would like to thank Mr Joe Martin, Executive Director, BCIS, for his help and the information on the Guide to Building Construction Duration, and the General Directorate for National Roads and Motorways, Divisions in Lublin, Krakow and Rzeszow, the Lublin, Krakow and Rzeszow Directorates of Regional Roads for providing access to project documentation.

The work has been financed from core statutory research funds of Polish Ministry of Science and Higher Education (S/63/2013)

**References**

Adeli H., Wu M. (1998). Regularization Neural Network for Construction Cost Estimation. *Journal of Construction Engineering and Management*, 124(1), 18-24.

BCIS. (2004a). Guide to Building Construction Duration, *BCIS*, London.

BCIS. (2004b). Time and Cost Predictability of Construction Projects. Analysis of UK Performance. *BCIS*, London.

BCIS. (2009). Construction Duration Calculator, www.bcis.co.uk/ construction_duration_calculator, accessed in April 2009

Biruk S. and Jaśkowski P. (2010). On the problems of modelling and reliability assessment of construction projects duration. *Reliability: Theory & Applications* (Electronic Journal Reliability & Risk Analysis: Theory & Applications), 4(19), 6-14.

Borowicz A. (2005). Public procurement of construction works (Statistical analysis). *Konferencja naukowo-techniczna Budownictwo Polskie w rok po wstapieniu do Unii Europejskiej*, Gdansk 9-11 czerwca, 41-47. (in Polish)

Borowicz A. (2008). Tendencje i uchybienia w kalkulacji cen zamowien publicznych na roboty budowlane. *Materiały XIV Konferencji Naukowo-Technicznej*, Ciechocinek, 15-17 października, 9-29 (in Polish)

Chan A.P.C. (2001). Time-cost relationship of public sector projects in Malaysia. *International Journal of Project Management*, 19, 223-229.

Chan D.W.M. and Kumaraswamy M.M. (1999). Modelling and predicting construction durations in Hong Kong public housing. *Construction Management and Economics*, 17, 351-362.

Chau K.W. (1995). Monte Carlo simulation of construction cost using subjective data. *Construction Management and Economics*, 13, 369-383

Cheung F.K.T. and Skitmore R.M. (2006). A modified storey enclosure model. *Construction Management and Economics*, 24, 391-405.

Chou, J.-S. and Tseng, H.-S. (2011). Establishing expert system for prediction based on the project-oriented data warehouse. *Expert Systems with Applications*, 38, 640-651.

Elhag T.M.S. and Boussabaine A. (1999). Tender price estimation: Neural networks vs regression analysis. Cobra 1999, *RICS Research Foundation*, 114-123.

Flyvbjerg B., Holm M.S., and Buhl S. (2002). Underestimating Cost in Public Works Projects: Error or Lie? *Journal of the American Planning Association*, 68(3), 279-295.

Gatnar E. (2001). *Nieparametryczna metoda dyskryminacji i regresji*. PWN, Warszawa.

Hoffman G.J., Thal A.E., Webb T.S., and Weir J.D. (2007). Estimating Performance Time for Construction Projects. *Journal of Management in Engineering*, *ASCE*, 23(4), 193-199.

Irfan, M., Khursid, M.B., Anastasopoulos, P., Labi, S., and Moavenzadeh, F. (2011). Planning-stage estimation of highway project duration on the basis of anticipated project cost, project type, and contract type. *International Journal of Project Management*, V 29 (1), 78-92.

ISPA. (2008). Parametric Estimating Handbook. Fourth Edition. Vienna, VA, USA, ISPA/ SCEA Joint Office.

Retrieved from: https://iceaaonline.org on December 25, 2012.

Jaśkowski P. and Sobotka A. (2012). Using soft precedence relations for reduction of the construction project duration. *Technological and Economic Development of Economy*, 18, 262-279.

Juszczyk M. (2008). Szacowanie kosztow realizacji obiektow budowlanych z zastosowaniem sztucznych sieci neuronowych. *Materiały Konferencji Naukowo-Technicznej Technologia i Zarządzanie w Budownictwie Ladek Zdroj*, 5-8 czerwca.

Kaplinski O. (1997). Modelling of Construction Processes: A Managerial Approach. KILiW PAN, IPPT, Studia z Zakresu Inzynierii, 43, Warsaw.

Kerzner H. (1984). *Project management*. A systems approach to planning, scheduling and controlling. Van Nostrand Reinhold, New York.

Ko, C.H. and Chen, Y.C. (2012). Evaluating Production Time Buffer for Precast Fabrication. *Journal of Engineering, Project, and Production Management*, 2(2), 101-111

KPI UK. (2003). All Construction Industry Key Performance Indicators Wallchart 2003. *Construction Best Practice*, Watford

KPI New Zealand. (2005). The New Zealand Construction Industry National Key Performance Indicators Wallchart 2005 Data. *Centre for Advanced Engineering Christchurch*, New Zealand.

Kaka A. and Price A.D.F. (1991). Relationship between value and duration of construction projects. *Construction Management and Economics*, 9, 383-400.

Lai Y-T., Wang W-C., and Wang H-H. (2008). AHP- and simulation-based budget determination procedure for public building construction projects. *Automation in Construction*, 17, 623-632.

Lee J-R, Hsueh S-L, and Tseng H-P. (2008). Utilizing data mining to discover knowledge in construction enterprise performance records. *Journal of Civil Engineering and Management*, 14(2), 79-84.

Leśniak A. (2004). Metoda kalkulacji kosztow posrednich robot budowlanych z zastosowaniem sztucznych sieci neuronowych. *Rozprawa doktorska, Politechnika Krakowska*, Krakow.

Liu S.S. and Wang C.J. (2012). Optimizing linear project scheduling with multi-skilled crews. *Automation in Construction*, 24, 16-23.

Love P. E. D., Tse R. Y. C., and Edwards D.J. (2005). Time–Cost Relationships in Australian Building Construction Projects. *Journal of Construction Engineering and Management*, *ASCE*, 131(2), 187-194.

Magnussen O.M. and Olsson N.O.E. (2006). Comparative analysis of cost estimates of major public investment projects. International Journal of Project Management 24, 281-288

Martin J., Burrows T.K., and Pegg I. (2006). *Predicting Construction Duration of Building Projects*. Shaping the Change XXIII FIG Congress, Munich, Germany, October 8-13.

Ng S.T., Mak M.M.Y., Skitmore R.M., Lam K.C., and Varnam M. (2001). The predictive ability of Bromilow's time-cost model. *Construction Management and Economics*, 19, 165-173.

Ogunsemi D.R. and Jagboro G.O. (2006). Time-cost model for building projects in Nigeria. *Construction Management and Economics*, 24, 253-258.

Polonski M. (2006). The analysis of the reliability of realization costs and investment's time limits in Warsaw. *Electronic Journal of Polish Agricultural Universities*, Topic Civil Engineering 9(4) www.ejpau.media.pl/volume9/issue4/art-10.html (accessed in May 2010).

Potts K. (2005). The New Scottish Parliament Building – A Critical Examination of the Lessons To Be Learned. *Third International Conference on Construction in the 21st Century (CITC-III) Advancing Engineering, Management and Technology*, 15-17 September, Athens, 534-539.

Skitmore R.M. and Ng S.T. (2001). Note. Australian project time-cost analysis: statistical analysis of intertemporal trends. *Construction Management and Economics*, 19, 455-458.

Skitmore R.M. and Ng S.T. (2003). Forecast models of actual construction duration and cost. *Building and Environment*, 38(8), 1075-1083.

Stanisz A. (2007). Przystepny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. Tom 2. Modele liniowe i nieliniowe. Statsoft Polska, Krakow.

Stoy Ch. and Polalis S. (2007). Early estimation of building construction speed in Germany. *International Journal of Project Management*, 25, 283-289.

Walker D.H.T. (1995). An investigation into construction time performance. *Construction Management and Economics*, 13, 263-274.

Yousef G. and Baccarini D. (2001). Developing a cost-time model for estimating construction durations of sewerage projects in Perth, Western Australia. *Journal of Construction Research*, 2, 213-220.

Agata Czarnigowska is Assistant Professor at the Department of Construction, Lublin University of Technology, Poland, where she received PhD in 2010. Her research interests include construction project management, especially cost planning and project control.

Anna Sobotka is Professor in the Department of Geomechanics, Civil Engineering and Geotechnics, AGH University of Science and Technology in Cracow, Poland. She received her degrees at the Warsaw University of Technology and Poznan University of Technology. Her research focuses on construction project management, supply chain management, construction logistics and sustainable development in construction.