APPRECIATION FOR DEVELOPMENT COST ESTIMATION IN EARLY DESIGN PHASES

D. Hellenbrand, K. Helten and U. Lindemann

Keywords: cost estimation, development costs, early design phases, uncertainty

1. Introduction and motivation

In early phases of product development designers want to predict as many attributes and features of the final product as possible. This especially includes the estimation of the product costs because they are highly relevant for the success of the project. The costs of a product consist in a first approximation of the production costs (manufacturing and assembly) and the development costs. Other life-cycle costs, e.g. recycling or waste disposal, are in this fist attempted not of relevance and therefore neglected.

There are many approaches in engineering design that estimate the production costs. Furthermore there are methodologies which consider the complete life-cycle, but there are almost no approaches that focus on development costs. Following literature a high percentage of the costs of a later product are predefined in the early design phases. However the actual costs really occur during the production of a product [Ehrlenspiel et al. 2007]. In typical mechanical engineering for example the used material is highly responsible for the later production costs. As a result in many companies this lead to the conclusion that the development costs are not of any relevance. Therefore all these costs are covered by defined factors in the overhead calculation and are not calculated or documented separately.

On closer examination these costs are of relevance, too. The companies in most cases have no idea which dimension the real costs for certain development project have. This is especially important in case of development projects that never lead to a final product. Furthermore there is a strong relation between costs and time effort of a development process. Following the ideas of process costing [Horváth & Mayer 1989] and activity based costing [Cooper and Kaplan 1988] the costs of a development project are approximately dependent on the effort of deployed hours of designers. This is due to the fact that the only used resource during a development process is the time of the involved designers. As a result the estimation of development costs is strongly connected to the prediction of the development effort in terms of time. This information is very important for the planning of resources and the number and chronology of development projects.

A major problem in early design phases is that many of the attributes of the final product are not exactly known. These uncertainties have to be considered during the cost estimation and the used method must be able to deal with it. Therefore a methodology is needed which allows for a fast and simple but nevertheless reliable estimation of the development effort expressed in terms of costs. Therefore an analysis of different cost estimation methods and their transferability to development costs has been carried out. Based on the results an approach for the development cost estimation has been created which especially deals with the uncertainties in early design phases.
2. Basis and related research

As already mentioned in engineering design there are a lot of methods that deal with the prediction of the production costs of a product [e.g. Ehrlenspiel et al. 2007]. There are almost no methods and tools available which allow for the calculation of development costs as well. Some tools are able to deal with the complete life cycle but the focus is on production costs. At the moment the costs for development are usually considered by an overhead calculation and added on the production costs. In most companies there is no documentation of the actual development costs of the products. Besides the well introduced traditional production cost prediction methods like similarity analysis, weight calculations or the usage of geometric parameters [e.g. Ehrlenspiel et al. 2007] there are some approaches that are in general suitable to estimate the costs of a development project. Most of them are in the field of software engineering to predict costs and time effort. The most common is the CoCoMo model which was originally published in 1981 [Boehm et al. 2000]. This approach is based on a regression model and uses the number of lines of code as the only input variable. During the last years a large number of methods and tools based on the same principle were created which used different metrics and variables to describe software projects. However most of this methods are limited to software engineering projects and there has been no adaption to engineering design. A more detailed description of the approaches is given in section 3.3.

Besides there are other approaches, methods and tools in different fields of engineering design available which allow for a description and optimization of development processes. Following the idea of process costing, these methods are in general suitable for the cost estimating problem. So in the beginning a widespread analysis of different methods and tools was carried out regarding their usability for the prediction of development costs. The methods have been allocated in four major groups dependent on their basic working principle: Process costing, statistic simulation, regression analysis and DSM-based methods. A summary of the analyzed methods and their characteristics is given in the following. The results are the basis for the developed estimation approach in this paper.

3. Cost estimation of development processes

3.1 Process costing

Traditional cost calculation methods and their adoptions like process costing [Horváth and Mayer 1989] or activity based costing [Cooper and Kaplan 1988] are able to allocate costs to different task. The basic idea is the modelling of the process steps and the costs of every single step. Afterwards the complete costs of the examined process can be calculated by adding the values of all process steps. A special remark during the analysis was laid on a calculation using individual pathways [Gahr 2006] which is an adoption of process costing especially for individualized products. The approach uses so called individual pathways which can easily be adapted to a new or changed product. Further the model allows for a process description on different levels of abstraction. The designer can start on a very abstract level and the more detailed steps are added afterwards. The costs of a process step on a higher level are calculated out of the costs of the belonging detailed ones. The general idea is presented in figure 1.

![Figure 1. Individual pathways as an example of process costing (following Gahr 2006)](image)
These approaches work well for the estimation of production costs. In this case it is possible to specify distributions for every task to improve results. In case of the individual pathways a triangular distribution is used [Gahr 2006]. In case of independent process steps (which can be assumed in production processes) it is possible to derive an overall distribution by adding all values of the distributions. A major problem dealing with development processes is that in reality the single development tasks are not independent [e.g. Browning 1998]. Therefore it is not possible to derive an overall distribution of the development costs by simply adding all single distributions. In this case only one single value for every task can be used. That makes the results quite unusable because designers are not able to estimate costs for every task well enough.

3.2 Statistic simulation
An approach to overcome these limitations is the use of statistic simulation (Monte-Carlo simulation) to derive an overall cost distribution. In this case a number of independent simulation runs are carried out. This allows for the specification of cost distributions for every single task on the most detailed level. In every simulation run one single value which fits with the specified distribution is calculated and therefore these values can simply be added.

So the general principle of this approach follows the idea of process costing as well. The basic steps are shown in figure 2. After defining a development process (possible on different abstraction levels) the costs and distributions of the single tasks have to be estimated by the designer. In terms of costs and time a triangular distribution fits best in most cases [Browning 1998]. Afterwards the simulation is carried out to derive a distribution for the overall development costs. A major advantage of this method is that the interrelations of the activities do not have to be understood and modelled. Theoretically it is possible to include interdependencies between the tasks as well by using a covariance matrix. At the moment there are no guidelines available for setting up this matrix and how the entries should be interpreted. This offers an opportunity for further research.

<table>
<thead>
<tr>
<th>Process step</th>
<th>Costs [x1000€]</th>
<th>pessimistic</th>
<th>presumably</th>
<th>optimistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Define Project</td>
<td>11.17</td>
<td>1200</td>
<td>1100</td>
<td>10500</td>
</tr>
<tr>
<td>2. Check feasibility</td>
<td>4.57</td>
<td>5000</td>
<td>4500</td>
<td>4200</td>
</tr>
<tr>
<td>3. Define Requirements</td>
<td>6.00</td>
<td>7000</td>
<td>6000</td>
<td></td>
</tr>
<tr>
<td>4. Plan project</td>
<td>2.57</td>
<td>3000</td>
<td>250</td>
<td></td>
</tr>
<tr>
<td>5...</td>
<td>13.33</td>
<td>15000</td>
<td>1300</td>
<td></td>
</tr>
</tbody>
</table>

There are different tools available to perform Monte-Carlo simulations. In this case the Excel based tool “Crystal Ball” by Oracle [Oracle 2009] was used. It has a very intuitive user interface, allows an easy definition of distributions and offers different possibilities to analyze the results.
3.3 Regression analysis

Regression analysis in general is used to describe the relation of a certain output (response) by different input variables (predictors) [Chatterjee and Hadi 2006]. The methodology is well introduced and used for the description of different kinds of systems. To set up a regression model a significant number of projects (data sets) has to be analyzed to identify the significant describing predictors. Afterwards the regression model is created. So the regression analysis can be used to describe a correlation between attributes of a product and is therefore suitable for development cost estimation.

A problem within the regression analysis is that a lot of data is required to get a reliable model. On the other hand all methods used in software engineering like the Constructive Cost Model (CoCoMo) and later adoptions [Boehm et al. 2000] as well as the production costs estimation models [Ehrlenspiel et al. 2007] are based on regression analysis. It also offers the possibility to interpret the dependencies and to identify relevant predictors (cost drivers). That makes it very promising to adopt this method to cost estimation in engineering design. Besides it is very close to similarity analysis which is also often successfully used in practice. In figure 3 the CoCoMo model originally published in 1981 is presented as an example for regression analysis. It is based on the analysis of a large number of software projects and allows for an easy estimation of effort, time and required persons [Boehm et. al. 2000]. The used metric to describe the product is the estimated lines of code. In addition only a basic classification of the project (organic, semi-detached, embedded) is needed to calculate the effort of the project. Later adoptions like CoCoMo II (Boehm et al. 2000) introduce some more project specific factors to achieve better results but the basic idea remains the same.

Another related approach in this context is the Constructive Systems Engineering Cost Model (CoSysMo) [Valerdi 2005]. This method is used for the estimation of systems engineering effort in large systems. Accordingly it is not limited to software but allows an effort prediction in both hardware and software. The CoSysMo model is also bases on regression analysis and provides equations to calculate the effort. Like CoCoMo II the model includes different factors like size and cost drivers or effort multipliers to describe the boundary conditions of the examined system [Valerdi 2005].

![Figure 3. Software development effort estimation using CoCoMo](following Boehm et al. 2000)

It is also possible to use fuzzy logic or neural networks to describe a relation between some characteristic attributes of a product and its development costs. So these methods can therefore be adopted to estimate costs. The main difference to regression models is that the exact interrelation of input and output is not visible to the user. These methods also have the general problem that there is a lot of data required to train them but at the same time there is no possibility to analyze the dependencies and thereby verify them. So it becomes very difficult to identify the significant input variables for the estimation of development costs.

3.4 DSM-based approaches

Methods based on the Design Structure Matrix (DSM) [Steward 1981] use a matrix to describe relations between the elements of a product or (development) process. There are approaches that focus on modelling and optimisation of the process itself [e.g. Wynn and Clarkson 2009] or which use a
separate matrices for product and process as well as coupling matrices [e.g. Gärtner et al. 2008]. In both mentioned cases probabilities are used to describe interdependencies or effects among the product/process elements. They are represented by belonging entries in the matrices. This allows for a statistic simulation of the process and the analysis of the effects. So on a first view these approaches seem quite perfect for cost estimation.

The major problem that all DSM-based approaches have in common is that they require a lot of effort and experience to set up the needed matrices. In particular it is of special relevance to specify and model iterations in an adequate way. This requires a lot of experience with the development task and it is not possible to conclude generalized outcomes. Furthermore the representation using matrices is not very intuitive and requires a lot of cognitive capabilities as well.

To sum it up the DSM-based approaches are more specialized on optimizing a given process than in the estimation of time and cost effort in absolute terms. The effort to set up a reliable model of the product and/or development is much too high and a lot of the needed information is not available during the early design phases.

3.5 Results of the analysis

The analysis of the four different categories of methods to estimate development costs points out that all available methods have specific strengths and weaknesses. The process costing approach is in general suitable for any kind of process. The basic idea is well introduced and the usage for triangular distributions fits well into the needs of design engineers. But there are problems within the interconnections of single process steps. Especially the assumption of independent tasks is not applicable to development processes. So the method can only be used with single values for every task.

A way to get around this limitation is the use of statistic simulation. Thereby a high number of independent simulation runs are carried out. The costs are varied according to the specified distributions and the result is a distribution of the development costs. In this case the summation is mathematically allowed because for every single run only one value is created. There is also a tool available to support this approach which is based on an Excel sheet. The modelling is very easy and intuitive for the designer and allows for the use on different levels of abstraction and dealing with different distributions. It is also possible to model interdependencies by using a covariance matrix.

Furthermore different approaches using regression analysis were presented. A number of production cost estimation methods are based on this method. It describes a relation between product attributes and its development costs. In software and systems engineering there are the CoCoMo and CoSyMo models that illustrate the practical relevance. Until now there is no adaption to engineering design processes. A major advantage is that iterations and interdependencies of the process steps are an implicit part of the model. On the other hand a lot of data is required to create a reliable model. Finally DSM-based approaches show that they are generally suitable for development cost estimation. But the creation of a reliable model requires a high effort and a lot of experience with the method.

To sum it all up there is no single method that is suitable in early phases especially when dealing with uncertainties. Uncertainty in this context means that many attributes and characteristics the final products are yet not known to the designer. A lot of decisions and refinements that have effects on the
costs are made in later phases of the project. Therefore a new approach is needed which allows for a fast estimation of development costs regarding uncertainties and easy interpretation of the results. The derived approach is presented in the next chapter.

4. Cost estimation in early phases regarding uncertainties

The analysis of different approaches of in principle suitable methods for cost estimation shows that Monte-Carlo simulation and regression analysis offer the most promising opportunities to estimate costs of development processes. The regression analysis thereby appears very promising because it is successfully used in software and systems engineering projects. This indicates that it could be adapted to engineering design processes as well. So a set of significant predictors to describe the relation between development costs and attributes of the final product have to be identified. With regard to specific development situations in different companies also process attributes describing the process should be included. To describe the specific situation in a single company also attributes describing the development process should be included.

The major problem of methods using regression models is the uncertainty in early design phases. In most cases the designer has no detailed information of final properties of a product. On the other hand the regression analysis is only able to describe a strict relation between input and output variables (in this case costs). Therefore a methodology is needed which allows the usage of uncertain input variables and thereby leads to meaningful and reliable results.

These goals can be achieved by combining regression analysis and Monte-Carlo simulation in one cost estimation model. To include the uncertainties of the later product attributes not only one simulation is carried out but a number of independent simulations. The input variables of the regression are varied in a certain range with regard to specified distributions to deal with the uncertainty. The ranges of the predictors and their distributions have to be defined by the designer and depend on the specific company, the development situation and the product. The proceeding of cost estimation using a combined approach of regression analysis and statistic simulation in presented in the following sections.

The process can generally be divided into two phases. First step is the regression analysis and the model creation. The setting up of a regression model is not described in detail because there is a lot of literature available in this field [e.g. Chatterjee and Hadi 2006]. Afterwards the distributions of the predictors are defined and the cost estimation is carried out. In consequence the following steps can be seen as an outline of a reference procedural model which leads to the cost model. The process and its belonging steps are also presented in figures 5 and 6.

To illustrate the steps of the approach the example of an oil-cooler is used. In this case it is an adoption of an existing product to a new field of application. Hence it is not a complete new development and the basic working principles and the system structure are generally known. So the new product is comparable and it is possible to use data out of former projects for setting up the regression model. The project is highly relevant for the company and therefore a prediction of development costs was carried out.

4.1 Setting up the regression model

The first step is the creation of a regression model as the basis for the following Monte-Carlo simulation. Therefore a set of data of completed development projects is needed with include the available data to characterize the projects. To achieve customized and most reliable results these data sets should include different categories of variables. On the one hand there are variables that describe the final product, on the other hand variables that characterize the development process.

Afterwards the data sets of finished projects are analyzed using regression analysis methods. Thereby the significant parameters have to be identified which have the main effects on the costs of the later product. The other variables are not taken into further consideration to create a most simple regression model. This is especially important for performance and the time effort of the following Monte-Carlo simulation (see section 4.2). The result of this step is a regression model that describes the development costs depending on the identified product and process attributes (see figure 5). This procedure is standard for all regression analysis.
The example in figure 5 shows the development cost analysis of the oil cooler. In this case, there were six parameters chosen to describe the project. Five attributes to describe the product (weight, cooling medium, performance, complexity and profit) and one that is related to the process (involved persons). The set of finished projects (similar to the new one) is analyzed to identify the significant attributes. In this case those are weight, performance, involved persons and profit. With these predictors the final regression model is created.

![Identification of relevant parameters and regression analysis](image)

There are a number of upcoming questions concerning the presented regression analysis. The first one deals with the adequate kind and number of product and process attributes to describe the project. Theoretically all available information should be analyzed because in the beginning the effects on the later product are not known. For the practical use of the method a guideline has to be created which helps to identify fitting attributes depending on the product and the development situation. Therefore more research has to be done and the approach has to be applied to different products in different companies. It has to be figured out how these identified variables and attributes should be coded to gain the desired results and which metrics can be used. The aim is the creation of a generic guideline to identify possible fitting attributes. These results are used to generate a generic guideline for setting up cost estimation models of development processes.

A further problem that might occur in the practical use of the presented procedure is the identification of the development costs which are needed in all cases. As already stressed these costs are not documented explicitly in most companies. A possible solution is the use of development effort in terms of time. With the known price per hour for every designer it is possible to deduce the development costs.

The aim of this paper is to present a general approach of dealing with uncertainties in early phases. Hence most of these questions have not been solved yet and offer some space for further research. This especially includes a procedural model for the needed attributes of the product and process.

### 4.2 Statistic simulation of the model

Afterwards the identified product and process describing predictors have to be varied to regard the uncertainties in early phases. So the next step is the definition of distributions for the predictors. According to Browning [1998] in terms of costs and time a triangular distribution fits best. For all other predictors the designer has to find fitting distributions based on his experience or information out of similar projects. In other words he has to make an educated guess.
It is possible to define different distributions for every input parameter (predictor) or to use the same distribution (with different parameters) for all variables. First experiences with the approach indicate that in most cases the triangular or normal distributions are fit the needs of the user best. These distributions are well known to most developers and therefore they should be able to specify their parameters.

Finally the Monte-Carlo simulation which varies the predictor variables according to the specified distributions is carried out. This leads to an overall distribution of the development costs. The results of the simulation can be analyzed using statistical methods. It is possible to calculate the envelope of a simulated project or to derive a representative distribution that fits the results best. Besides the expected value more parameters, like the variance, of the distribution of the overall development costs are visible. This allows for statements e.g. that the development costs are lower than a specific value on a defined level of significance. For example the tool “Crystal Ball” [Oracle 2009] which was presented in chapter 3.2 supports a brought variety of analysis methods.

By using an average hourly price rate it is also possible to calculate the effort in terms of time (hours, month) very easily. It is also possible to go the other way around and calculate the time effort first and derive the development costs afterwards. This is a second possibility in case the development costs are not available for the regression analysis. The process of effort estimation regarding uncertainty is shown in figure 6.

![Figure 6. Monte-Carlo simulation with distributed predictors](image)

An option to deal with the problem of selecting and defining fitting distributions for the predictors is to carry out a number of Monte-Carlo simulations instead of a single one. For every run a different combination of distributions can be used. In case of easy regression models and a not too high number of simulation runs these will not cost much time. Afterwards the results can be compared and differences can be analyzed. With this sensitivity analysis the effects of using different distribution can be identified and the results become more reliable.

In the presented oil cooler example a linear regression model is used (figure 6). The details of the model are not relevant in this context so only a generic term is presented. Afterwards there are two different distributions used to describe the uncertainty in the predictor variables. In one case all attributes are described using a triangular distribution; in the other case a normal distribution is used. After the simulation runs the results of the different simulations were analyzed and compared. In this case there is a difference between triangular and the normal distribution. Furthermore there could be more simulation runs using more or different distributions within one simulation run. In the presented example with six predictors and 10,000 simulation runs the simulation took less than a minute on standard computer hardware. So the simulation time is not the limiting factor. The time to specify the distributions is much higher and thereby limits the number of different simulation runs.

In case of the oil cooler the results of different simulations seem very promising at the moment. A first plausibility check of the results implies that the model leads to suitable results. Furthermore results do
that these costs are of relevance because they are strongly connected to time effort.

So an analysis of different methods which are in general suitable to estimate development costs (process costing, statistic simulation, regression analysis and DSM-based approaches) was performed. The results point out that there are some successful and practical relevant approaches available to estimate costs and time effort in software engineering and for production costs. These methods use different metrics and input variables to specify a project and are based on regression analysis. Therefore the general idea of these approaches should be assigned to the field of development cost estimation in engineering design.

Based on these findings an approach was created which combines regression analysis with Monte-Carlo simulation and allows for the estimation of development costs in early design phases. It uses product and process attributes to specify the development project and is able to deal with uncertainties at the same time. It is also possible to estimate the time effort of a project by using average cost rates. As a result the method is also interesting for resource planning in development departments and the chronological organization of development projects.

The approach consists of two major parts. In a first step a regression analysis is performed to identify the relevant attributes and to describe a relation between these attributes and the development costs. The second step is the definition of the distributions for the predictor variables to regard uncertainties in the later product attributes. The result is an overall distribution of the development cost which can be analyzed using statistic methods. A major advantage of using a regression model is that interdependencies and iterations in the process do not have to be modelled as they are an implicit part of the model. The approach was applied to a basic and very simple example and the results are very promising.

The major problems or limitations of the approach are connected to the creation of the regression model. First of all a generally similar or comparable former product is needed. Otherwise the data does not lead to a fitting regression model. In industries most projects are adoptions to changed conditions and in most cases suitable products should be available. For setting up universal regression models for a special type of products (like in CoCoMo, see section 3.3) a much higher effort and number of projects is needed. The second and more critical problem deals with the acquisition of the data for the regression model. The needed data might be not available in the company’s systems or to the designer, the quality is not adequate or there are simply too less data sets. In this case the model cannot be created or will not lead to reliable results. Therefore it has to be analyzed which minimum quality of data is needed for the cost estimation problem. Another possible way out might be the use of single static simulation (see below).

During the next steps it has to be analyzed which attributes and metrics are adequate to characterize engineering development projects (product and process related) and how these variables should be coded to achieve the expected results. This includes the identification of relevant categories of attributes to improve the practical use of the approach. The aim is the creation of a generic guideline for the identification and modelling of fitting attributes for the cost estimation in early phases.

Besides the refinement of the presented approach the usage of a statistic simulation using software tools like Crystal Ball (see section 3.3) will be analyzed in more detail. The method is closely related to process costing and thereby suitable for the cost estimation of development processes. The modelling of interdependencies and iterations is not generally necessary because these dependencies can be represented by the chosen distributions. The single use of statistic simulation is suitable if the required data for the regression model is not available (see above). In this case a model still can be
created by describing the process based on designers experience and defining distributions for the single process steps.

The use of a covariance matrix seems to be a promising way to model interdependencies and to improve the results. There are lot of open questions dealing with the interdependencies of process steps. In a first analysis it has to be clarified if the correlation matrix is suitable to describe engineering design processes and leads to better results. This might be possible by comparing the results to different DSM approaches. Another question deals with the correlations matrix itself. There have to be further analysis what in this context is expressed by the entries of the matrix, what the means for the process are and how dependencies can be quantified. Furthermore a practical way which allows for an easy creation the matrix has to be found because designers are not so much familiar with this kind of description. To sum it up a guideline is needed which describes how the dependencies among process steps can be quantified and expressed within the correlation matrix.

Acknowledgement
This research is supported by the Bundesministerium für Bildung und Forschung (BMBF), Germany, through the project “VireS – Virtuelle Synchronisation von Produktenentwicklung und Produktionssystementwicklung” (engl. Virtual Synchronisation of Product Development and Production System Planning”).

References

David Hellenbrand
Scientific Assistant
Technische Universität München, Institute of Product Development
Boltzmannstr. 15, 85748 Garching, Germany
Telephone: +49.89.289.15124
Telefax: +49.89.289.15144
Email: david.hellenbrand@pe.mw.tum.de
URL: http://www.pe.mw.tum.de