Abstract

Empirical studies differ in what they report as the underlying relation between project size and percent cost overrun. As a consequence, the studies also differ in their project management recommendations. We show that studies with a project size measure based on the *actual cost* systematically report an increase in percent cost overrun with increased project size, whereas studies with a project size measure based on the *estimated cost* report a decrease or no change in percent cost overrun with increased project size. The observed pattern is, we argue, to some extent a statistical artifact caused by imperfect correlation between the estimated and the actual cost. We conclude that the previous observational studies cannot be considered as providing reliable evidence in favor of an underlying project size related cost estimation bias. The more robust evidence from controlled experiments, limited to small tasks, suggests an increase in underestimation with increased project size.

Keywords: Cost estimation, magnitude bias, statistical artifacts, regression analysis, random error, non-random samples

Is there a Magnitude Bias in Project Cost Estimation

Percentage cost estimation error in projects may be measured as the difference between the actual and the estimated cost divided by the estimated cost. If the percentage cost estimation error is positive there is a cost overrun and if negative there is a cost underrun. A frequently reported estimation bias is the tendency towards higher percentage cost overrun on larger than of smaller projects. This finding is reported for tasks with sizes ranging from simple, small scale tasks, e.g, the paper sheet counting tasks in (Roy and Christenfeld 2008), to large engineering projects (Heemstra and Kusters 1991; Gray, MacDonell et al. 1999; Moløkken-Østvold, Jorgensen et al. 2004; Yang, Wang et al. 2008). Explanations in support of an increase in cost overrun with increasing project size may involve both human biases, e.g., increased over-confidence with increased size and or complexity (Grieco and Hogarth 2009), and biases resulting from rational estimation strategies, e.g., the variance shrinkage effect that typically occurs if people base their estimates on the average cost of similar projects (Hatton 2007).

There are studies that report the opposite bias, i.e., a decrease in cost overrun with increased project sizes. The opposite bias is, for example, reported in several engineering project contexts (Odec 2004; Creedy 2006; Bertisen and Davis 2008). In addition, there are studies showing no significant relationship between project size and percentage cost overrun, e.g., the rail and road construction projects in (Flyvbjerg, Skamris Holm et al. 2004). The difference in results has not surprisingly led to differences in recommendations. Odeck (2004), who found an increase in percentage cost overrun with decreasing project size, recommend that managers should pay special attention to cost control of the smaller projects. Sauer et al. (2007a), who found the opposite bias, recommend on the other hand the managers to keep the projects small to avoid cost overruns. The common belief among practitioners, as far as we have experienced, is that the percentage cost overruns of larger projects tends to be higher than those of smaller ones. The belief that larger projects are less predictable and manageable than smaller ones, is for example consistent with the creation of software development methods based on so-called "incremental development" (Larman and Basili 2003), i.e., methods based on splitting a large projects into smaller ones.

There may be natural reasons for the studies´ differences in reported relationship between project size and percentage cost overrun. There may, for example, be differences in the ability to and complexity in handling large projects in different contexts. It is, however, also possible that the statistical analyses typically used to support the claimed relation between project size and cost estimation bias are problematic and causes artificial variations in the results. In that case we should be very careful about making claims about the underlying (causal, "true") relationship and recommendations based on the analyses. In this paper we claim that there are indeed problems with the analyses. The analyses do not, we argue, enable a separation of statistical artifacts from underlying relationships. This is not necessarily a problem when the results are used for prediction purposes, but can easily lead to incorrect interpretations of underlying relationships.

The remaining part of this paper is organized as follows: First, we review relevant empirical studies and report that the studies' difference in results seems to be strongly related to their choice of project size variable. Then, we apply regression analysis mathematics to show that the observed project size related difference in reported results is an expected consequence (statistical artifact) of imperfect correlation between the estimated and the actual cost. Examinations of problems related to random error in project size measurement and non-random samples suggest that project size variables related to the estimated and to the actual cost are likely to lead to interpretation problems. We discuss the interpretation problems and argue that robust knowledge about the underlying relationship between project size and cost estimation bias may require other types of studies and analyses, e.g., controlled experiments with fixed project size variables and studies aiming at better in-depth understanding of the involved mechanisms.

A Review of Empirical Studies

In this section we review empirical studies that indicate an underlying relationship between project size and percent cost overrun. The main purpose is to give an initial assessment of the validity of general claims regarding the impact of project size on percentage cost overrun. The assessment is based on the assumption that a change from one meaningful project size measure to another should not lead to a substantial change of the reported underlying relationship between project size and percent cost overrun. Otherwise, the results are likely to be results of properties of the analyses, rather than the underlying phenomenon.

Through searches in various digital libraries and reference lists of relevant papers we identified thirteen observational studies reporting results on the relation between project size and percent cost overrun. The studies report from engineering projects, particularly software and infrastructure construction projects. There may be relevant studies from other domains that we were unable to identify, but the selection should be sufficient to demonstrate the methodological problems.

There are several potentially meaningful measures of project size. In our review we found, however, mainly two types of project size measures being used: one related to the projects´ *actual cost* and another related to the projects' *estimated cost*. We denote these two measures ACT (actual cost) and EST (estimated cost). Other measures, such as measures related to number of people involved or calendar time (duration) were sometimes included, but never as the main project size measure in the analyses of percent cost overrun. This may be understandable given that such measures may be perceived as less connected with project size. A project may, for example, involve more people or last for a longer period of time than another and still be perceived as smaller. Project size measures related to the produced output were not applied in any of the identified studies. This may be a consequence of lack of measures of produced output that enable comparison across different project types. The studies′ choices of project size measure, i.e., the choices between the estimated and the actual cost, were typically *not* motivated by an explicit argumentation. An exception is the study presented in (Flyvbjerg, Skamris Holm et al. 2004), where the decision to use the estimated (forecasted) cost as project size measure is based on the following argumentation:

> "*First, cost escalation* [percent cost overrun] *is statistically confounded with actual construction cost being part of it, whereas forecast construction costs are not. Second, the decision about whether to go ahead with a given project is based on forecast construction costs; this is the decision variable, not actual cost.*"

Both arguments are in our opinion debatable. Both the estimated and the actual cost are included as part of cost overrun measures, e.g., when defining cost estimation overrun as the ratio of actual to estimated cost. As a consequence, both project size measures are statistically confounded with the cost overrun measure, i.e., there is a mathematical coupling between the estimation error measure and the project size measure (Tu, Maddick et al. 2004). We discuss the consequence of this coupling in more detail in Section 3. The validity of the second argument depends on the purpose of the analysis. If the purpose is to *predict* the percent cost overrun, it is a reasonable argument. If, however, the purpose is to gain knowledge about the underlying relationship between project size and percent cost overrun, which we understand is the purpose of the analyses in (Flyvbjerg, Skamris Holm et al. 2004), the argument is of lesser relevance. The problematic nature of the choice of project size measures may be illustrated by the extreme case where a project is estimated to be small, but ended up being very large. Is this an example of a cost overrun of a small or a large project? It is possible to argue in favor of the use of the actual size as project size measure in such cases, since it may be more meaningful to understand project size as the size the project turned out to have, rather than the size the organization thought the project would have.

Studies on the underlying relationship between project size and degree of cost overrun naturally focus on the ratio between the estimated and the actual cost, since it is unquestionable that the absolute difference between the actual and the estimated project cost tend to increase with increased project size. A frequently used cost estimation error measure is the ratio of actual (ACT) to estimated (EST) cost:

$$(1) \quad \frac{ACT}{EST}$$

If this ratio is higher than 1 there is a cost overrun, if it is less than 1 there is a cost underrun. Some of the reviewed studies use the inverse value, i.e., the ratio $\frac{EST}{ACT}$, some use the value 1 - $\frac{ACT}{EST}$, so that zero cost overrun corresponds with the value 0, and some use other variants derived from this measure. The reviewed studies sometimes measure the project size and the cost overrun as ordered categories (ordinal scale), e.g., they measure cost estimation error and/or the project size on the scale “small” - “medium” - “large”. When both the cost estimation error and the project size are measured as ordinal scaled variables the presented analyses is typically based on cross tabulation, e.g., Chi-square analyses. When the cost estimation error is measured as a ratio scaled and the project size as an ordinal scaled variable, the analysis is typically based on comparing the mean or median cost estimation error of projects in different size categories, e.g., through ANOVA tests. When both variables are measured as ratio scaled variables, the analysis is typically based on regressing project size on cost estimation error. We denote these three types of analyses TAB (cross tabulation), CAT (size category), and REG (regression-based analysis) for the purpose of the discussion in this article.

The software development projects included in the review use project size and cost estimation error measures related to work-effort, not cost. Since work-effort is, by far, the dominant cost driver in such projects, we decided to use the term cost for those studies, as well. The methods used to derive the cost estimates in the reviewed studies are likely to be expert judgment-based (Jørgensen 2004; Odec 2004).

If a study reports that there is an increase in $\frac{ACT}{EST}$ with increased project size, we code this finding as "ICO". This corresponds to an increase in percentage cost overrun (or decrease in percentage underrun) with increased project size. The opposite, a decrease in $\frac{ACT}{EST}$ with increased project size, is coded as "DCO". If a study reports that there is neither an increase nor a decrease in $\frac{ACT}{EST}$ with increased project size, this is coded as "CCO". To simplify the presentation we have not distinguished between findings that are reported to be statistically significant and findings where no such statistical information is available. We find this defendable since we were more concerned with the combined results of a family of studies, and not so much the statistical significance of each individual study.

In cases where the data set was available for reanalysis, we present both the original and the "alternative" result. The alternative result is the result derived from the analysis when changing the project size variable either from actual to estimated cost or from estimated to actual cost. We try to keep the alternative analysis as close as possible to the original analysis on all other aspects. The value "-" means that we did not have access to the data set and, as a consequence, an alternative analysis was not possible. The review of the thirteen identified studies according to the above variables and coding is presented in Table 1.

<Please, insert Table 1 here>

The results presented in Table 1 suggest that a study is highly likely to report an increase in percentage cost overrun with increased project size (ICO) when measuring the project size as the actual cost, while it most likely will report a decrease (DCO) or no change (CC0) when measuring project size as the estimated cost. The perhaps strongest argument for an effect of the choice of project size variable is, however, the observation of simultaneously increase (ICO) and decrease (DCO) in cost overrun with increased project size dependent on whether the actual or the estimated cost is used as project size measure in two of the data sets. A strong impact from the choice of project size variable on the reported results seems to be independent on whether the analyses are based on cross tabulating categories of project size and cost overrun (TAB), comparison of mean cost overrun of different project size categories (CAT) or regression analysis (REG). We will briefly discuss the reasons for this in Section 3.

We interpret the review summarized in Table 1 as implying that there may be some problems with the robustness of the results and that the reported results to some extent may be statistical artifacts of the choice of project size measure. The next section aims at explaining when and why we should expect the project size measure dependent cost estimation bias results observed in Table 1.

The Effect of Choice of Project Size Variable

The relationship between project size and cost estimation error may be modeled as follows, using the actual cost (ACT) as the project size variable:

$$(2)\quad \frac{ACT}{EST} = \alpha_1 ACT^{\beta_1}$$

Correspondingly, we may model the relationship using the estimated cost (EST) as the project size variable:

(3) $\frac{ACT}{EST} = \alpha_2 EST^{\beta_2}$

The expression in (2) implies that there is an increase in percentage cost overrun (disproportional more under-estimation or less over-estimation) with increased actual cost if $\beta_1 > 0$, a constant estimation error if $\beta_1 = 0$ and a decrease in percentage estimation error if $\beta_1 < 0$. The expression in (3) has the same implication for the estimated cost. When we in the following refer to cost estimation error or cost overrun, we refer to the ratio $\left(\frac{ACT}{EST}\right)$, unless otherwise stated.

Log-transforming both sides of (2) and (3) has the advantage of resulting in linear models and, when we use regression to estimate the model parameters (α and β), the distributions of data values typically better resemble normal distributions than the non-transformed values in project cost estimation contexts. The log-transformation gives:

(4) $\ln\left(\frac{ACT}{EST}\right) = \ln(ACT) - \ln(EST) = \ln(\alpha_1) + \beta_1 \ln(ACT)$

(5) $\ln\left(\frac{ACT}{EST}\right) = \ln(ACT) - \ln(EST) = \ln(\alpha_2) + \beta_2 \ln(EST)$

The relationships in (4) and (5) can be expressed as:

(6) $\ln(EST) = -\ln(\alpha_1) + (1 - \beta_1)\ln(ACT) = \alpha_1' + \beta_1' \ln(ACT)$

where $\alpha_1' = -\ln(\alpha_1)$ and $\beta_1' = (1 - \beta_1)$

(7) $\ln(ACT) = \ln(\alpha_2) + (1 - \beta_2)\ln(EST) = \alpha_2' + \beta_2' \ln(EST)$

where $\alpha_2' = \ln(\alpha_2)$ and $\beta_2' = (1 + \beta_2)$

Thus, if we estimate the model parameter $\beta_1^{'}$ of (6) from a dataset using linear regression giving an estimate $b_1^{'}$, we can estimate the model parameter β_1 of (2) as $b_1 = 1 - b_1^{'}$. Furthermore a test of $b_1 \neq 0$ which, if true, would mean that cost estimation error was correlated with actual project size, corresponds to a test of $b_1^{'} \neq 1$. Similarly, a test of $b_2 = (b_2^{'} - 1) \neq 0$ which, if true, would imply that cost estimation error was correlated with estimated project size, corresponds to a test of $b_2^{'} \neq 1$. If $b_1' < 1$ then $b_1 > 0$ which implies an increase in cost overrun with increased actual cost. If $b_2' < 1$ then $b_2 < 0$ which implies a decrease in cost overrun with increased estimated cost. Consequently, an examination of the effect of a change of project size variable from actual cost in (2), to estimated cost in (3) corresponds to a change of the linear model from (6) to (7). Projects are estimated before they are completed, so it may look strange to study the time-reversed relationship. It is, however, meaningful to analyze the time-reversed relationship in (6) for purposes of examining the statistical robustness of findings. Campbell and Kenny (1999, p 158-159), for example, recommend a time-reversed analysis as a means to check for regression artifacts: "*When both the original and the time-reversed analysis yield essentially the same result* [similar b-values]*, then we should lose confidence in the interpretation of the original analysis.*"

If we apply a mean center transformation to ln(EST) and ln(ACT), which will not change the $\beta^{'}$ values, then both $\alpha_1^{'}$ and $\alpha_2^{'}$ equal zero. Assuming that (6) and (7) display the same underlying relationship and a deterministic relationship, we then have that:

(8) $\beta_1^{'} = \frac{\ln(EST)}{\ln(ACT)} = \frac{\ln(EST)}{\beta_2^{'}\ln(EST)} = \frac{1}{\beta_2^{'}}$ and $\beta_2^{'} = \frac{\ln(ACT)}{\ln(EST)} = \frac{\ln(ACT)}{\beta_1^{'}\ln(ACT)} = \frac{1}{\beta_1^{'}}$

Combining these models we have that

(9) $\beta_1^{'}\beta_2^{'} = 1$

However, when we estimate the model parameters using linear regression, based on the usual relationship between the estimate of the model slope and the correlation between variables, what we have is:

(10) $b_1^{'}b_2^{'} = r\frac{s_{\ln(ACT)}}{s_{\ln(EST)}} r\frac{s_{\ln(EST)}}{s_{\ln(ACT)}} = r^2$

where is r the (Pearson´s) correlation between the logarithm of the estimated and the logarithm of the actual cost, $s_{\ln(ACT)}$ the standard deviation of the logarithm of the actual cost and $s_{\ln(EST)}$ the standard deviation of the logarithm of the estimated cost. (The expression in (10) also shows that the correlation between ln(*EST*) and ln(*ACT*) can be interpreted as the geometric mean of the regression slopes estimates b_1' and b_2'.)

Comparing the expressions in (9) and (10) we see that it is only when there is a perfect correlation between the actual and the estimated cost, i.e. $r^2 = 1$, that we can expect results from analyses based on the estimated cost as project size measure to equal the results of analyses based on the actual cost as project size measure. The lower the correlation, the more we should expect the results of these two analyses to deviate. A re-expression of (10) in the highly likely context where there is a positive correlation between estimated and actual cost and imperfect cost estimates gives:

(11) $b_1^{'} = \frac{r^2}{b_2^{'}} < \frac{1}{b_2^{'}}$

The expression in (11) implies that, whenever there is an imperfect correlation between the estimated and the actual cost, we will find a stronger tendency towards increased cost overrun when using the actual size as project size measure than when using the estimated size. When there is no strong underlying increase or a decrease in cost estimation overrun with increased project size together with far from perfect correlation between estimated and actual cost, a change in result from increase to decrease in cost overrun when changing from actual to estimated cost is to be expected and a candidate to explain much of the difference in results of the studied in Table 1.

The data set in (Hill, Thomas et al. 2000) may serve as an illustration of the relevance of the above expressions. The correlation $r_{\ln(ACT),\ln(EST)}$ is 0.755 and the regression slope $b_1^{'}$ is 0.642 when analyzing the projects of that data set giving $b_1 = 1 - 0.642 = 0.358$, which suggest a strong *increase* in cost overrun with increased project size when measured as the actual cost. By re-expressing (11) we find that $b_2^{'} = \frac{r^2}{b_1^{'}} = 0.887$, giving $b_2 = .887 - 1 = -0.113$ which suggest a strong *decrease* in cost overrun with increased project size when project size is measured as the estimated cost.

The above expressions are based on the situation where the underlying relationship between project size and cost estimation bias is examined through regression analysis (REG). The results are however relevant to explain both the category- (CAT) and cross-tabulation-based (TAB) analyses. ANOVA and regression-based analyses are both subsets of the general linear model and uses equivalent procedures (Hays 1981; Wampold and Freund 1987). We may, for example, generate the same output as in an ANOVA-model by introducing proper dummy variables representing size categories in a regression model. An analysis of the frequently high similarity between rank (ordered categories) and regression-based analyses is given in (Anatolyev and Kosenok 2009). It is consequently no surprise that the project size measure dependent results in Table 1 are independent on whether they are derived from REG, CAT or TAB-based analyses.

There are alternative ways of explaining the relation between choice of project size measure and reported results. A perhaps especially interesting explanation is the one based on "regression to the mean" (Galton 1889). The interpretation problems we examine in this paper are, as we see them, of the same type as those misleading Galton when he observed that children of tall parents on average were smaller than their parents (but still taller than the total average) and understood this to mean that there was a biological (underlying) force that made people get more and more close to average in height. Later, by reversing the analysis, he found that the parents of children taller than the average height had parents that were smaller than them (Stiegler 1997). This implied that children became less and less close to average in height, i.e., the opposite of what he found in the original analysis. In addition, if there were an underlying biological force leading to more average heights, soon all people would be of average in height and the variance in height would decrease. In reality, the variance of heights was quite stable. Finally, Galton concluded that the regression to the mean effect was a result of imperfect correlation and not a consequence of underlying mechanisms leading to more and more average heights of people[1]. The problematic interpretation of the regression to the mean observation of Galton may be seen as similar to the interpretation problems in our situation, replacing the height of the parents with the estimated cost and the height of the children with the actual cost. Projects have a random element in their actual and estimated cost. Assume a neutral situation where it is as likely to over-estimate as under-estimate the actual cost of a project, i.e., a situation with unbiased cost estimates. Projects estimated to be unusually large, tend to have more than average random

[1] It is important to separate the prediction that extreme observations tend to be followed by less extreme ones, from the interpretation that there is an underlying force that causes the subsequent observation to tend to be more average. The first is a valid prediction, the second is an incorrect interpretation of why subsequent observations tend to be less extreme.

elements in favor of high estimates. Since the actual effort is expected to have average random elements in favor of high and low actual cost, those projects are likely to be over-estimated. Projects estimated to be unusually small, will, on the other hand, tend to have more than average random elements in favor of low estimates and most likely be under-estimated. The consequence is that we will typically observe decreasing cost overrun with increased estimated cost. A similar argumentation can be made to show that we should expect projects with unusually high actual cost to be under-estimated and unusually low actual cost to be over-estimated. As in Galton´s case, we cannot use this observation to claim the existence of an underlying bias, e.g., a biological force or human disposition. It could instead be a result of random variation in the values of the analyzed variables. The regression to the mean phenomenon has confused numerous researchers. Milton Friedman (Friedman 1992) once wrote that "*I suspect that the regression fallacy is the most common fallacy in the statistical analysis of economic data*". Its sometimes counter-intuitive nature also makes it mislead many real-life expectations and decisions. Rewarding very good and punishing very poor performance together with the regression towards the mean effect may, for example, lead to the impression that rewards has a negative effect, i.e., subsequent performance is on average worse, while punishment has a positive effect, i.e., subsequent performance is on average better (Kahneman and Tversky 1973). A more average performance after unusually good or poor performance is, however, as expected when there is a random variation in measured performance values and do not require any impact from reward or punishment to occur.

The analysis in this section shows that the choice between the estimated and the actual cost as project size measure is likely to give different, sometimes reversed, interpretations with regards to whether increased project size causes more or less cost estimation overruns. While this project size measurement dependency of the results may be sufficient to doubt the robustness of the presented results, it is not sufficient to claim that we cannot trust any of the results. If all the analysis problems were related to one of the size measures and none with the other, one may argue that the interpretation problems can be solved by selection of the proper project size variable. The next two sections argues that this is not likely to be the case, i.e., there will be interpretation problems with both types of project size measures. We discuss two of the elements potentially leading to interpretation problems, i.e., random error in the measurement of project size and non-random sampling. There are several other elements that potentially contribute to additional interpretation problems, e.g., omitted explanatory variables, non-linear relationships and multicollinarity. Our goal is to argue that the current types of analyses are insufficient to make claims about the underlying relationship between project size and cost estimation bias. For that purpose, we believe that an examination of random error and non-random sampling is sufficient.

Random Error in Measurement of Project Size

Measurement of the estimated and actual cost of projects will typically be exposed to random error. This is a violation of an essential assumption of ordinary regression analysis, which requires that the independent variables are "fixed" (or at least have no random error). A fixed variable is one were the values of the independent variables are set and not just observed as a result of sampling projects from a population. If the random error of the independent variables is substantial, the interpretation of the parameters as indicating underlying relationships is problematic.

The Size of the Random Error

It is, we argue, likely that the measurements of actual and the estimated cost of many projects are exposed to substantial random error. The measurement of actual cost is, amongst others, likely to be influenced by incorrect or inconsistent logging of cost, misunderstandings in the interpretation of cost data, and, strategically misleading reports of cost data. Even more randomness in actual cost results from the sometimes very large difference in cost when different providers complete the "same" project, i.e., projects that under the fixed variable assumption would be set to the same size. There are of obvious reasons not much data from different companies completing the same project. One exception is the study reported in (Anda, Sjøberg et al. 2009) where four companies where asked to complete the same software project, i.e., based on the same requirements, same client and same product acceptance criteria. The CV (coefficient of cost variance), i.e., the standard deviation of actual cost / mean actual cost, was as low as 0.29. The minimum actual cost was only 55% of the maximum cost of the projects building the same software. The same study also reports that the CV-value of the estimated cost was 0.65 for a set of software projects and between 0.1 and 0.3 for different sets of infrastructure projects. The sometimes very large random variance in the software project cost estimates are also reported in (Grimstad and Jorgensen 2007). In that study, the software developers estimated the effort they would need on the same task on several occasions over a six months period. The mean deviation between the effort estimates on the same task by the same developer was as high as 71%! This difference may be explained by variation in task estimation sequence, e.g., due to assimilation effects, differences in accessibility of previous experience and other effects that leads to a, from an observer´s point of view, high random variation of how large the developer perceives the task to be. There may be contexts where there are meaningful project size measures that can be measured with little random error, e.g., situations where the measurement of number of produced output units have close to zero random error and is a meaningful measure of project size. Judging from the identified studies, such situations may however be rare in real-world project situations.

The Effect of the Random Error

To illustrate how the random error in the measure of project size affects the regression model assume that:

(12) $ACT = ACT_{TRUE}e_{ACT}$

(13) $EST = EST_{TRUE}e_{EST}$

The expression in (12) corresponds to a situation where the measured actual cost (ACT) is a function of the true actual cost (ACT_{TRUE}) and a multiplicative error term e_{ACT}. It is, as debated in for example (Brenner 2000), frequently not clear what the true values refers to, but this is not essential for the illustration in this section. It may, similar to what is common in true score measurement (Cronbach 1972), be considered as the mean or median cost of the "same" project, when repeated without learning several times. The true actual cost will hardly ever be known in real-life context. A similar argumentation can be made for the true estimated cost (EST_{TRUE}) in expression (13).

Log-transforming the expressions in (12) and (13) give:

(14) $\ln(ACT) = \ln(ACT_{TRUE}) + e'_{ACT}, where\ e'_{ACT} = \ln(e_{ACT})$

(15) $\ln(EST) = \ln(EST_{TRUE}) + e'_{EST}, where\ e'_{EST} = \ln(e_{EST})$

When we estimate the β-values of (6) and (7) adjusted for random error in the independent variable, i.e., b_1^{adj} and b_2^{adj} can be expressed as (derived from (Greene 2003), see also en.wikipedia.org/wiki/Errors-in-variables_models):

(16) $b_1^{adj} = b'_1\left(1 + \frac{s^2_{e'_{EST}}}{s^2_{\ln(EST)}}\right),$

$where\ s_{e'_{EST}} is\ the\ standard\ deviation\ of\ e'_{EST}\ and\ s_{\ln(EST)}\ the\ standard\ deviation\ of\ \ln(ES$

$$(17)\ \ b_2^{adj} = b_2' \left(1 + \frac{s_{e'_{ACT}}^2}{s_{\ln(ACT)}^2}\right),$$

where $s_{e'_{AC}}$ *is the standard deviation of* e'_{ACT} *and* $s_{\ln(ACT)}$ *the standard deviation of* $\ln(A$

The above equations are based on several assumptions not necessarily met, e.g., independence between the random error and the independent variable and an unlimited sample, and it is mainly included for illustrative purpose. A comprehensive discussion about the effect of random error in the independent variables can be found in, for example, (Kendall 1951). We see from (16) and (17) that when the random error in the independent variable is large compared to the total variance of the values of that variable, the reported b-values will be lower than the true (underlying) β-value. A deflated b_1'-value means that studies will report a too high increase in cost overrun with increase project size when using actual cost as project size variable. A deflated b_2'-value means, on the other hand, that studies will report a too low decrease in cost overrun with increased project size when using estimated cost as project size variable. Earlier, we argued that there is likely to be a substantial random error in the measurement of both the estimated and the actual cost of a project. This means that the random error in the estimated and the actual cost may explain much of the observed project size measurement dependent results presented in Table 1. To adjust the b-values for random error using (16) or (17) we either need to know the standard deviation of the random error or the ratio of standard deviation of the random error to the standard deviation of the total variance of one of the variables. In typical project contexts, we are unlikely to know any of these values.

Simulation of the Effect of the Random Error

To illustrate the effect of what we think is in some project contexts a realistic amount of random error in the measurement of actual and estimated cost, in a situation with no underlying effect of project size on cost estimation error, i.e., a CCO-relationship, we produced a data set based on the following scenario:

1. The true actual cost values (ACT_{TRUE}) of 5000 projects are randomly drawn from a uniform distribution with values between 1000 and 10 000 cost units.
2. The true estimated cost (EST_{TRUE}) is set equal to the true actual cost. This ensures that the underlying relationship between estimated and actual cost is a perfect linear relationship, i.e., $b_{TRUE} = 1$.
3. The random errors of the actual cost (e_{ACT}) and the estimated cost (e_{EST}) are randomly drawn from a normal distribution with mean 1 and standard deviation 0.2. This implies a CV-value of 0.2, which may be similar to what is reported for some types of real life projects (Anda, Sjøberg et al. 2009). These error terms are, see (12) and (13), multiplicative and represent a situation where 68% of the measurements deviate less than 20% from the true value. The ratio of the square of the standard deviation of the random error to the total standard deviation is in this case 0.11. The error distributions are symmetric and identical for all project size values, which is consistent with an underlying (true) CCO-relationship between project size and cost overrun.
4. The observed actual and estimated costs are calculated applying (12) and (13), respectively. This results in a correlation between ACT and EST of 0.83 and between ln(ACT) and ln(EST) of 0.89.
5. The regression analysis of the original (6) and the alternative (7) model is conducted on the log-transformed values of actual and estimated cost. The distributions of the error terms of the log-transformed models are then not perfectly normally distributed, but sufficiently normal for our simulation purposes.

Since the standard deviation of ln(ACT) equals that of ln(EST) we will have that $b_1' = b_2' = r = 0.89$. While $b_1' = 0.89$ corresponds to $b_1 = 1 - 0.89 = 0.11$ implying noticeable increase in cost overrun with increased project size, $b_2' = 0.89$ corresponds to $b_2 = 0.89 - 1 = -0.11$ implying the opposite relationship.

The Effect of Narrow Project Size Intervals and Random Error

As can be derived from (16) and (17), the deviation between the observed (0.89) and the true (1.0) regression slope equals the ratio of the squared standard deviation of the random error and the squared total standard deviation correspond to the difference. The need for b-value adjustment to reflect underlying relationships is consequently quite sensitive to the interval of project sizes studied. If we, keeping all other values equal, had restricted the simulation to project sizes between 1000 and 2000 cost units, b_1' and b_2' would have been 0.47, while the same simulation with project sizes between 1000 and 100000 cost units would have led to b_1' and b_2' values of 0.94. Consequently, not only the choice of project variable is of potential relevance for the reliability of the analyses when there is random error in the variables, but also the size interval of the studied projects. In the extreme case where all projects are of almost the same true actual size, nearly all variance in the cost overrun would be due to random error in the project size variable and the need for adjustment of the b-values would be very strong. We should consequently be especially careful in our interpretation of the parameters of regression models when the projects do not vary much in size and there is likely to be random error in the independent variables. Consistent with the relevance of the effect of narrow project size interval and illustrating the potential relevance of the above argumentation, we find for example a substantial correlation (0.54) between the rank of the b_1' values and the rank of the standard deviation of ln(ACT).

The relevance of narrow intervals can be further illustrated by an analysis of the dataset in (Creedy 2006). We split this data set into two equally sized sets of projects based on estimated cost higher or lower than the median estimated cost. This had the consequence that the standard deviation of the budgeted cost of the smaller projects is quite low, while the standard deviation of the data set of larger projects is not much affected. The relation between project size measured as actual cost and estimation error goes from CCO (b_1 = -0.01) to ICO ($b_{1,SMALL}$ = 0.13) for the data set of smaller projects, while the change in b_1-value for the data set of larger projects is only minor ($b_{1,LARGE}$ = 0.01). It is, of course, possible that there is a different relationship between project size and cost estimation bias for smaller and larger projects. The problem is that we cannot know to what extent the results are consequences of non-random sampling and to what extent it represents underlying relationships.

Results from Experiments with no Random Error

A study presented in (Roy and Christenfeld 2008) is interesting because it is based on random allocation of treatment and fixed task size variable. The use of fixed task size variables removes the problems of random error in the independent variable and the use of random allocation reduces the problem of omitted independent variables in the regression models. In that study, the participants were randomly allocated to the effort estimation and execution of the counting of a pile of paper sheets fixed on the task sizes 50, 100, 250 or 500 sheets. The original analysis in the paper is exposed to random error effects, i.e., it uses the log-transformed actual effort as the task size variable and the log-transformed ratio of actual to estimated effort as the cost bias variable. We received the data set from the authors and found the same pattern as in the studies in Table 1 when using the random task size variables. Larger tasks tended to be under-estimated, while smaller tasks tended to be over-estimated (b_1 = 0.185) when using the actual effort as project size variable, while larger tasks tended to be over-estimated and smaller tasks under-estimated (b_2 = -0.135) when using the estimated effort as project size variable. If, however, we instead of using the random variables, i.e., the estimated or actual effort, use the logarithm of the fixed variable “number of paper sheets counted” as our project size variable, we find the regression slope 0.241, i.e., support for the increase of under-estimation with increased task size. Another study with small tasks, applying fixed variables, finding the same increase is described in (Halkjelsvik, Jørgensen et al. 2001). That study found that larger reading tasks, as defined by number of pages, were associated with a decrease in the expected reading effort per page. Although actual reading performance was not assessed, the results indicate more underestimation (less overestimation) for larger tasks.

These two studies indicate an underlying relationship where an increase in task size produces more estimation overrun. While limited to two small tasks, none of the very interesting in themselves, these results are statistically more robust than those based on the observational (non-experimental) project data in Table 1. The problem is, however, that it is hardly feasible to do similar random treatment experiments with more ecologically valid, larger projects.

Non-random Sampling

A correlation between the model error (the non-explained variance of the dependent variable) and one of the independent variables leads to biased model parameters. This type of correlation may be a result of incorrectly specified model, e.g., through omitted variables, but may also be a result of non-random sampling (Heckman 1979). Non-random sampling may be a consequence of practical concerns, e.g., data collection limitations, or caused by processes inherent in the phenomenon under study. As is the case with incorrectly specified models, the direction and size of the total bias is in general hard to predict (Berk 1983). We will in this section argue that non-random sampling contributes to the interpretation problems of observational studies of the relation between project size and cost estimation error. An extensive list of selection biases leading to non-random samples is given in (Delgodo-Rodriguez and Llorca 2004). In the context of project cost estimation, the following two selection biases may be particularly interesting:

1. Data collections limited to projects within a particular size interval, e.g., exclusion of projects with budgets or actual cost smaller or larger than a certain value.
2. Over-representations of projects with over-optimistic cost estimates. Project bids based on over-optimistic cost estimates are more likely to result in a project. This is evident from, for example, the results reported in (Hinze, Selstead et al. 1992) where the average cost overrun was found to increase with the number of bidders. In competitive bidding rounds this bias may be substantial and is frequently termed "the winner´s curse" or the "optimizer's curse" (Smith and Winkler 2006), but the bias will also be present in situation where a company has to prioritize between projects and selects the one with the estimated best cost-benefit. An over-representation of projects with high cost overruns means that we do not evaluate all project cost estimates. When the set of completed projects is the population of interest this may not be considered a non-random sample. In our case, however, we are interested in estimation biases. Clearly, it would be incorrect to make claims about estimation biases, if the observed effect is mainly a project selection bias.

Figures 1-2 illustrate, inspired by the presentation in (Berk 1983), how the above two situations may affect the regression estimate, $b_1^{'}$, of the parameter $\beta_1^{'}$ in model (6), i.e., the model, $\ln(EST) = \alpha_1^{'} + \beta_1^{'} \ln(ACT)$ and the regression estimate, b_1, of the parameter β_1 in model (4), i.e., the model $\ln\left(\frac{ACT}{EST}\right) = \ln(ACT) - \ln(EST) = \ln(\alpha_1) + \beta_1 \ln(ACT)$. As before, a test of $b_1^{'} \neq 1$ in first model corresponds to a test of $b_1 \neq 0$ in the second model. The choice between (4) and (6) in the two figures is therefore mainly motivated by ease of illustration. We use the same data set as in the previous simulation (Section 3). This data set represents, as before, a situation where the underlying (true) relationship is true slope is 1, but the observed slope is $b_1^{'} = 0.89$ (and consequently $b_1 = 1 - 0.89 = 0.11$), which suggest an increase in cost overrun with increased project size. The boxes in the figures represent the variance of the dependent variable. The regression line in the middle of the box is the line found when including the whole data set, while the dotted line is the regression line resulting from the reduced, non-randomly selected, data set. The reduced data set is the full data set (the whole box) without the shaded area.

<*Please, insert Figure 1 here*>

Figure 1 illustrates that an exclusion of projects with low cost estimates leads to a further decrease of the slope, i.e., to a slope even more deviating from the true slope than the observed slope, which was already deflated from 1 to 0.89 due to the presence of random error in the measurement of actual cost. Similarly, exclusion of projects with high estimated cost would also deflate the $b_1^{'}$ -value. In this simulated case, a non-random selection based on thresholds of the actual cost would not affect the regression slope. When using the alternative model in (5), it is however a non-random selection based on thresholds of the actual cost that deflates the regression slope. Data sets based on threshold on either the estimated or the actual cost will consequently contribute to the diverging, project size dependent, results reported in our review in Table 1. It can be derived from this that even when we are not really interested in the smaller or larger projects, they may have to be included to enable proper interpretation of parameters of the regression models. For the same reason, the narrower the project size interval studied, the stronger the biasing effect of non-random sampling.

<*Please, insert Figure 2 here*>

Figure 2 represents the winner's curse situation, which is common when, for example, bidding for large infrastructure and software development projects. Clearly, the actual situation is far from as simple as presented in Figure 2, but the effect may be similar, i.e., the slope will be deflated. In this case we have that the slope of the sub-sample is closer to the true slope (which is a slope of 0). However, the total effect of all types of non-random samples is typically hard to predict. The non-randomness illustrated in Figures 1 and 2, for example, seems to impact in different directions and the total effect depends on which of the project selection biases is the strongest.

Most studies do not say much about their inclusion and exclusion criteria or discuss the effects of potential selection biases. The potential relevance of non-random samples may nevertheless be illustrated by the use of a data set only including projects which had exceeded budget by more than 10% in (Creedy 2006) (which is likely to reduce the effect of project size of cost overrun), the non-randomness of the data when relying on publicly available data and voluntarily responses from surveys as in (Heemstra and Kusters 1991; Flyvbjerg, Skamris Holm et al. 2004; Odec 2004; Creedy 2006; Sauer, Gemino et al. 2007b; Bertisen and Davis 2008; Yang, Hu et al. 2008), and the cut off value of larger projects as a results of a focus on maintenance tasks and smaller work packages as in (Gray, MacDonell et al. 1999; Hill, Thomas et al. 2000; van Oorschot, Bertrand et al. 2005). In additions comes the unavoidable selection biases related to the "winner's curse" and lack of inclusion of projects that were started, but never completed. Projects started, but not completed may affect the analysis, since projects with high actual cost tend to have lower survival rate (Sauer, Gemino et al. 2007b), i.e., the larger the project and the cost overrun gets, the less likely it is that the cost overrun will be measured and included in the data set.

Discussion and Conclusion

We have in the previous sections argued that there are reasons to doubt the robustness of observational studies on the relation between project size and cost estimation bias. There are a couple of, laboratory-based, small tasks, studies that avoids the interpretation problems through fixed size variable values and random treatment. These two studies suggest that there is an increase in effort (cost) overrun with increased task (project) size. In the other reviewed studies, however, we cannot know with confidence whether the reported results are statistical artifacts or represent underlying relationships. As a consequence, our empirical knowledge about how project size affects cost estimation bias in real-world project situations seems to be weak.

The problems with statistical analyses of observational data in situations with random error in the independent variables and incorrectly specified models (including omitted variables and non-random sampling) are not unique for cost estimation studies and several solutions have been proposed. To correct for random error in the independent variable we may, for example, use information about the random error of the independent variable, the ratio of the random error to the total variance, or the ratio of random error of the dependent to the independent. Orthogonal regression analysis (Carrol and Ruppert 1996), for example, assumes that the level of random error is the same in the dependent and the independent variable and the method proposed in (Blomquist 1986) that the ratio of the variance of random errors of the dependent to the variance of independent variable is known. We may also use “method of moments”, “instrumental variables” (Fuller 1987), the geometric mean functional relationship as suggested in (Barker, Soh et al. 1988), and, analyses of the change in variance (Oldham 1962). To what degree some of these methods solve the problems related to random error in the independent variable in the type of analyses discussed in this article, without introducing new problems, is hard to tell. A general observation is that the methods seem to require information difficult to collect or assumptions difficult to evaluate.

The corrections could therefore easily lead to even larger analyses biases (Hausman 2001). As an illustration of to what extent the error in the independent variable is random is hard to evaluate, which may have as consequence that "corrections" based on this assumptions does not improve the analyses, see (Reichardt 2000). Even if we managed to control for the random error in the independent variable, there will be remaining problems, particularly the problem of incorrectly specified model. Similarly to the problems with random error in the independent variable, there are numerous methods proposed to correct for misspecification. The main strategies are related to inclusion of additional, relevant variables, collection of additional information which enable better specification of the model and adjustments for missing variables (Marais and Wecker 1998). However, as pointed out in for example (Clarke 2005), these strategies may sometimes increase rather than decrease the bias of the analyses. As an illustration, Griliches (1977) argues that small amounts of measurement error in the control variables (variables added to avoid the omitted variable bias) "are magnified as more variables are added to the equation in an attempt to control for other possible sources of bias". In addition, adding more variables may lead to problems with collinarity (Whiteside and Narayanan 1989) and there is a complex relationship between sampling strategies and proper regression analysis, as pointed out in for example (Winship and Radbill 1994).

A solution to the analysis problems does not only have to solve the challenges related to one problem at the time, but all problems simultaneously. For that purpose, it may be necessary to rely on controlled experiments with random allocation of project sizes and fixed size variables. The main limitation of controlled experiments is however, as pointed out earlier, that it is hard to study projects of the sizes that are typical in industrial contexts, e.g., in cost estimation of engineering projects. Alternatively, one might try to better understand the underlying cognitive strategies when estimating the cost of projects. This way one may, for example, find situations where people typically assume linear relations between size and cost, while there in reality is a non-linear relationship. The result presented in (Staats, Milkman et al. 2011) suggests for example that there is an underestimation of coordination work and that this underestimation increases the more people involved in the team work. If this is a robust finding, it means that it is likely to see an increase in cost overrun with increase in project size when the increase in project size also means an increase in number of people involved. More in-depth studies of this type, with more realistically sized projects, may lead to a better understanding of the relationship between project size and cost overrun.

Even if we were able to identify project size-based cost estimation patterns, we should be careful about interpreting them as human biases. Several rational cost estimation strategies would lead to a pattern with increased cost overrun with increased actual cost. The use of analogy-based estimation is one such strategy. This estimation strategy is based on the selection, from memory or databases with historical information, of projects that have characteristics similar to the one to be estimated and use the actual cost of these projects as input to the cost estimate of the new project (Jørgensen, Indahl et al. 2003). One may for example choose to use the mean cost of similar, previously completed projects as the estimated cost of the new project. This is likely to lead to lower variance of the cost estimates (variance shrinkage) than of the actual cost values. This, in turn, will contribute to more overestimation (less underestimation) with increased actual cost. In order to find the opposite relationship in the regression models, i.e., that the cost overrun does not change or decreases with increased actual cost, the variance of the estimated cost has to increase compared to that of the actual cost so that:

(18) $s_{\ln(EST)} \geq \dfrac{s_{\ln(ACT)}}{r}$

The expression in (18) shows that in situation when the correlation between the estimated and the actual cost is low, the variance of the estimated cost has to be much higher than that of the actual cost to observe a decrease in cost overrun with increased actual cost (DCO). Assume, for example, a correlation of 0.8 between ln(EST) and ln(ACT). In this case, the standard deviation of the logarithm of the estimated cost has to be 25% higher than that of the actual cost to observe a constant or decreasing cost overrun with increased project actual cost. We have not been able to identify meaningful cost estimation strategies that would lead to this substantial level of increase in variance from actual to estimated cost. In other words, there are reasons to expect an increase in cost overrun with increased actual project size even when rational estimation strategies are applied. An examination of the studies in Table 1 where we had the data sets available (six studies) indicates that there is typically a decrease in the variance of the estimated compared to the actual cost. The mean standard deviation of ln(EST) is 6.2% lower than that of ln(ACT). Only one of the studies had an increase in the standard deviation of the estimated cost and that increase was small (2%).

So, what do we currently know about the effect of project size on cost estimation bias? Is there valid evidence to support either a division of larger projects into smaller ones or to join projects into larger ones to increase the planning abilities? Unfortunately, we do not have strong evidence either in support of either division or joining of projects. The limited evidence that is likely to be methodologically sound point in the direction of that there may be an increase in cost overrun with increased project size, but this evidence is mainly valid for very small tasks that are not closely related to engineering projects.

Acknowledgement: Thanks to Les Hatton, John Hill, Kim E. van Oorschot and Michael M. Roy for making their data sets available.

References:

Anatolyev, S. and Kosenok, G. (2009). "Tests in contingency tables as regression tests." Economics letters **105**: 189-192.

Anda, B. C. D., Sjøberg, D. I. K. and Mockus, A. (2009). "Variability and reproducibility in software engineering: A study of four companies that developed the same system." IEEE Transactions of Software Engineering **35**(3): 409-429.

Barker, F., Soh, Y. C. and Evans, R. J. (1988). "Properties of the geometric mean functional relationship." Biometrics **44**: 279-281.

Berk, R. A. (1983). "An introduction to sample selection bias in sociological data." American Sociological Revew **48**: 386-398.

Bertisen, J. and Davis, G. A. (2008). "Bias and error in mine project capital cost estimation." Engineering economist **53**(2): 118-139.

Blomquist, N. (1986). "On the bias caused by regression towards the mean in studying the relation between change and initial value." Journal of clinical peridontology **13**: 34-37.

Brenner, L. (2000). "Should observed overconfidence be dismissed as a statistical artifact? Critique of Erev, Wallste, and Budesco (1994)." Psychological Review **107**(4): 947-949.

Campbell, D. T. and Kenny, D. A. (1999). A primer on regression artifacts. New York, The Guilford Press.

Carrol, R. J. and Ruppert, D. (1996). "The Use and Misuse of Orthogonal Regression in Linear Errors-in-Variables Models." The American Statistician **50**(1): 1-6.

Clarke, K. A. (2005). "The phantom menace: Omitted variable bias in econometric research." Conflict Management and Peace Science **22**: 341-352.

Creedy, G. D. (2006). Risk factors leading to cost overrun in the delivery of highway construction projects. Queensland University of Technology. **Doctoral thesis**.

Cronbach, L. J., Gleser, G. C, Nanda, H., Rajaratnam, N. (1972). The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York, John Wiley & Sons.

Dantata, N. A., Touran, A. and Schneck, D. C. (2006). Trends in U.S. rail transit project cost overrun. Transportation Research Board 2006 Annual Meeting, Washington, DC.

Delgodo-Rodriguez, M. and Llorca, J. (2004). "Bias." Journal of Epidemiologic Community Health **58**: 635-641.

Flyvbjerg, B., Skamris Holm, M. K. and Buhl, S. L. (2004). "What causes cost overrun in transport infrastructure projects?" Transport review **24**(1): 3-18.

Friedman, M. (1992). "Do old fallacies ever die?" Journal of economic literature **30**: 2129-2132.

Fuller, W. A. (1987). Measurement error methods.

Galton, F. (1889). Natural inheritance. London, Macmillan and company.

Gray, A., MacDonell, S. G. and Shepperd, M. J. (1999). Factors systematically associated with errors in subjective estimates of software development effort: the stability of expert judgment. IEEE 6th International Metrics Symposium, Boca Raton, Fl., IEEE Computer Society.

Greene, W. (2003). Econometric analysis (5th edition). New Jersey, Prentice Hall.

Grieco, D. and Hogarth, R. M. (2009). "Overconfidence in absolute and relative performance: the regression hypothesis and Bayesian updating." Journal of Economic Psychology **30**: 756-771.

Griliches, Z. (1977). "Estimating the results to scholloing: Some econometric problems." Econometric **45**: 1-22.

Grimstad, S. and Jorgensen, M. (2007). "Inconsistency of expert judgment-based estimates of software development effort." Journal of Systems and Software **80**(11): 1770-1777.

Halkjelsvik, T., Jørgensen, M. and Teigen, K. H. (2001). "To read two pages, i need 5 minutes, but give me 5 minutes and i will read four: how to change productivity estimates by inverting the question." Applied Cognitive Psychology **25**: 314-323.

Hatton, L. (2007). "How Accurately Do Engineers Predict Software Maintenance Tasks?" IEEE Computer **40**(2): 64-69.

Hausman, J. (2001). "Mismeasured variables in econometric analysis: Problems from the right and problems from the left." Journal of Economic Perspectives **15**(4): 57-67.

Hays, W. L. (1981). Statistics. New York, Holt, Rinehart & Winston.

Heckman, J. H. (1979). "Sample selection bias as a specification error." Econometrica **47**(1): 153-161.

Heemstra, F. J. and Kusters, R. J. (1991). "Function point analysis: Evaluation of a software cost estimation model." European Journal of Information Systems **1**(4): 223-237.

Hill, J., Thomas, L. C. and Allen, D. E. (2000). "Experts' estimates of task durations in software development projects." International Journal of Project Management **18**(1): 13-21.

Hinze, J., Selstead, G. and Mahoney, J. P. (1992). Cost overruns on state of Washington construction contracts. Transportation Research Record. Washington**:** 87-93.

Jørgensen, M. (2004). "A review of studies on expert estimation of software development effort." Journal of Systems and Software **70**(1-2): 37-60.

Jørgensen, M., Indahl, U. and Sjøberg, D. I. K. (2003). "Software effort estimation by analogy and "regression toward the mean"." Journal of Systems and Software **68**(3): 253-262.

Kahneman, D. and Tversky, A. (1973). "On the psychology of prediction." Psychological Review **80**(4): 237-251.

Kendall, M. G. (1951). "Regression, structure and functional relationship. Part I." Biometrika **38**(1/2): 11-25.

Larman, C. and Basili, V. (2003). "Iterative and Incremental Development:A Brief History." IEEE Computer(June): 47-56.

Marais, M. L. and Wecker, W. E. (1998). "Correcting for omitted-variables and measurement-error bias in regression with an application to the effect of lead in IQ." Journal of the American Statistical Association **93**(442): 494-505.

Moløkken-Østvold, K., Jorgensen, M., Tanilkan, S. S., Gallis, H., Lien, A. C., Hove, S. E. and ieee computer, s. (2004). A survey on software estimation in the Norwegian industry. 10th International Symposium on Software Metrics, Chicago, IL, Ieee Computer Soc.

Odec, J. (2004). "Cost overruns in road construction-what are their sizes and determinants?" Transport policy **11**: 43-53.

Odeck, J. (2004). "Cost overruns in road construction—what are their sizes and determinants?" Transport Policy **11**: 43-53.

Oldham, P. D. (1962). "A note on the analysis of repeated measurements of the same subjects." Journal of chronic diseases **15**: 969-977.

Reichardt, C. S. (2000). "Regression facts and artifacts." Evaluation and Program Planning **23**: 411-414.

Roy, M. M. and Christenfeld, N. J. S. (2008). "Effect of task length on remembered and predicted duration." Psychonomic Bulleting & Review **15**: 202-207.

Sauer, C., Gemino, A. and Blaize, H. R. (2007a). "The impact of size and volatility on IT project performance." Communications of the ACM **50**(11): 79-84.

Sauer, C., Gemino, A. and Reich, B. H. (2007b). "The impact of size and volatility - On IT project performance." Communications of the Acm **50**(11): 79-84.

Smith, J. E. and Winkler, R. L. (2006). "The optimizer's curse: skepticism and postdecision surprise in decision analysis." Management Science **52**(3): 311-322.

Staats, B. R., Milkman, K. L. and Fox, C. R. (2011). "Group work planning fallacy: Team size matters more than you think." submitted for publication.

Stiegler, S. M. (1997). "Regression towards the mean, historically considered." Statistical methods in medical research **6**: 103-114.

Tu, Y.-K., Maddick, I. H., Griffiths, G. S. and Gilthorpe, M. S. (2004). "Mathematical coupling can undermine the statistical assessment of clinical research: illustrations from the treatment of guided tissue regeneration." Journal of Dentistry **32**(2): 133-142.

van Oorschot, K. E., Bertrand, J. W. M. and Rutte, C. G. (2005). "Field studies into the dynamics of product development tasks." International Journal of Opertaions & Production Management **25**(8): 720-739.

van Oorschot, K. E., Bertrand, J. W. M., Rutte, C. G. (2005). "Field studies into the dynamics of product development tasks." International Journal of Operations & Production Management **25**(8): 720-739.

Wampold, B. E. and Freund, R. (1987). "Use of multiple regression in counseling psychology research: A flexible data-analytic strategy." Journal of counceling psychology **34**(4): 372-382.

Whiteside, M. M. and Narayanan, A. (1989). "Reverse regression, collinearity, and employment discrimination." Journal of Business & Economic Statistics **7**(3): 403-406.

Winship, C. and Radbill, L. (1994). "Sampling weights and regression analysis." Sociological Methods & Research **23**(2): 230-257.

Yang, B., Hu, H. J. and Jia, L. X. (2008). "A Study of Uncertainty in Software Cost and Its Impact on Optimal Software Release Time." Ieee Transactions on Software Engineering **34**(6): 813-825.

Yang, D., Wang, Q., Li, M. S., Yang, Y., Ye, K., Du, J. and Acm (2008). A Survey on Software Cost Estimation in the Chinese Software Industry. ACM/IEEE International Sympsoium on Empirical Software Engineering and Measurement, Kaiserslautern, GERMANY, Assoc Computing Machinery.