A remedy for the instability in the relationship between cost overruns and project size: evidence from the Norwegian petroleum industry

Sindre Lorentzen
University of Stavanger
sindre.lorentzen@uis.no

Abstract

Project investment size is frequently used to explain cost overruns, but the literature is inconsistent with regard to its effect. This inconsistency might be caused by the idiosyncratic property of projects coupled with the use of ordinary least square estimation. Through a Monte Carlo simulation and an empirical example using data from offshore oil and gas projects on the Norwegian continental shelf (NCS), I show that the more efficient least absolute deviation estimation approach is more appropriate for inferring the relationship between project size and cost overruns.

Keywords: cost overruns, project size, bias.

JEL classification: C180**,** C210, C510**,** D220

1.0 Introduction

Cost overruns, defined as the relative change between the initial and final project budgets, are prevalent across both different industries and time periods. Much effort has been dedicated in the past literature to quantifying and explaining the occurrence of cost overruns. In this pursuit, perhaps one of the most frequently used variables for explaining cost overruns is project investment size. While other variables are certainly of interest, much of the empirical literature has been restricted to univariate regressions where project size seems to be frequently used. Data related to project execution are notoriously difficult to obtain, which might explain the popularity of investment size as an explanatory variable – it is necessary for calculating the overrun in the first place, which typically indicates that this variable is available.

Based on theoretical considerations, the relationship between cost overruns and project size is expected to be positive. In other words, ceteris paribus, bigger projects tend to have larger cost overruns. There are two main reasons for this. First, size is commonly regarded as a proxy for the complexity of the project. Intuition would dictate that larger projects tend to be more difficult to estimate and to coordinate, and thus more likely to incur cost overruns (Ansar et al, 2016). Hence, we expect less of a cost overrun for smaller projects, all other things being equal. Second, larger projects are typically more difficult to terminate when the budget is overspent (Merrow, 2011). The sunk cost of a megaproject is usually harder to swallow.

The empirical literature appears to be more divergent than theory would suggest. Empirical research across different industries and periods has found every possible relationship between cost overruns and project size, from a significant positive or negative link to none at all. This apparent lack of consensus is interesting and warrants further investigation, since policy recommendations will depend on the exact relationship. If the relationship is positive (negative), for instance, more attention should be dedicated to larger (smaller) projects. While the possibility always exists that the effect of project size depends on the type of industry concerned, this seems less likely since the statistical characteristics of cost overruns appear to conform across different samples. As pointed out by Jørgensen et al (2013), an alternative explanation for the lack of consensus could be the econometric methodology used to investigate this relationship.

Further inspection of the literature reveals a tendency to utilise univariate ordinary least square (OLS) to study this relationship. OLS is well known for its rigid assumptions and limitations. It is, for instance, an approximation of the average conditional mean and, as with all averages, is sensitive to outliers. Since sample size tends to be limited in cost overrun studies, owing to data being notoriously difficult to obtain, a few outliers could potentially be enough to change the outcome of regressing cost overruns on project size. As Taleb (2007) puts it, cost overruns are what we call a scalable random variable – a single outlier is potentially enough to change the average significantly. To illustrate this, consider an example provided by Taleb:

> This scalable/nonscalable distinction allows us to make a clear-cut differentiation between two varieties of uncertainties, two types of randomness. Let's play the following thought experiment. Assume that you round up a thousand people randomly selected from the general population and have them stand next to one another in a stadium. […] Imagine the heaviest person you can think of and add him to that sample. Assuming he weighs three times the average, between four hundred and five hundred pounds, he will rarely represent more than a very small fraction of the weight of the entire population (in this case, about a half of a percent). You can get even more aggressive. If you picked the heaviest biologically possible human on the planet (who yet can still be called a human), he would not represent more than, say, 0.6 percent of the total, a very negligible increase. And if you had ten thousand persons, his contribution would be vanishingly small. In the utopian province of Mediocristan, particular events don't contribute much individually – only collectively. I can state the supreme law of Mediocristan as follows: When your sample is large, no single instance will significantly change the aggregate or the total. The largest observation will remain impressive, but eventually insignificant, to the sum. But you cannot so easily rule out extreme variations with a different

> brand of quantities, to which we turn next. […] Consider by comparison the net worth of the thousand people you lined up in the stadium. Add to them the wealthiest person to be found on the planet – say, Bill Gates, the founder of Microsoft. Assume his net worth to be close to $80 billion – with the total capital of the others around a few million. How much of the total wealth would he represent? 99.9 percent? Indeed, all the others would represent no more than a rounding error for his net worth, the variation of his personal portfolio over the past second. For someone's weight to represent such a share, he would need to weigh fifty million pounds! […] In Extremistan, inequalities are such that one single observation can disproportionately impact the aggregate, or the total.

All studies can potentially contain outliers. However, data on cost overruns are perhaps especially likely to contain them owing to the inherit characteristics of projects. Projects are one-offs – they are by definition a unique set of tasks implemented within a given scope and time frame. In other words, they are idiosyncratic. If they were not, they would not be projects but business as usual. By the very nature of projects, we should expect data on cost overruns obtained from projects to be both heteroscedastic and riddled with outliers. Regardless of the true relationship between cost overruns and project size, outliers can significantly influence the econometric results – to the extent that the opposite relationship is obtained.

The presence of outliers may distort regression results in either direction, but it can be argued that the bias is more likely to be negative. Consider the following. The characteristics of both cost overruns and project size have are two noteworthy aspects. First, cost overrun is a constrained variable in the sense that it can only take on values between -100 per cent and positive infinity, [-100, ∞]. We could have a 200 per cent overrun, but a 200 per cent underrun would be impossible. Furthermore, it is far easier to overspend the budget by 50 per cent than to decrease it by an equal amount. Combined, these two aspects imply that severe outliers can only be positive. Second, an outlier is more likely to occur with smaller projects. Arguably, megaprojects tend to fewer than smaller ones. Hence, we have more opportunities to obtain outliers in the sample for smaller projects. If that is the case, even with a positive relationship between cost overrun and project size, the threat exists that outliers with a high cost overrun and small project size will bias the result so that the estimated coefficient becomes either insignificant or significantly negative. Having established that outliers represent a prevalent threat in studies on project cost overruns, it becomes apparent why OLS might not be the most suitable estimator in seeking to quantify the relationship between cost overruns and project size. OLS aims to estimate the conditional expectation. It provides the best average fit but, as established, this conditional mean might be sensitive to outliers. Estimating the conditional median function by utilising quantile regression would make more sense, since the median is more robust to outliers than the OLS estimator.

To summarise, the crux of the argument is as follows. I suggest that, owing to the idiosyncratic nature of projects, characteristics exist with cost overrun data which make OLS infeasible for estimating the relationship between cost overruns and project size. In other words, the inconsistency in the cost overrun literature could be a result of faulty methodology. By using offshore oil and gas development projects, I will empirically demonstrate in this paper how OLS is incapable of producing a significant relationship, while outlier-robust methodologies such as quantile regression are.

This article is structured as follows. Section 2 will review relevant literature on cost overruns and project size. The data set utilised for further analysis is presented in section 3. Econometric results are presented in section 4. Finally, section 5 concludes.

2.0 Literature review

Cost overruns have been an active area of research for decades, and much attention has been paid to investigating the relationship between them and project size. As pointed out by Jørgensen et al (2012), however, the empirical literature has not established a consensus regarding the effect of project size on cost overruns. Table 1 shows an overview of the empirical results obtained in several articles on the subject. As shown, results range from significantly positive or negative to insignificant. Jørgensen et al (2012) argue that this inconsistency in the literature is partly attributable to statistical artefacts related to the operationalisation of project size.

Table 1: Overview

Author	Sample size	Project	Relationship
Creedy (2006)	231	Road	↓
Dantata et al (2006)	37	Rail	↑
Flyvbjerg et al (2004)	131	Rail	–
Gray et al (1999)	77	Software	↑
Hatton (2007)	957	Software	↑
Heemstra and Kusters (1991)	388	Software	↑
Hill et al (2000)	506	Software	↓
Moløkken-Østvold et al (2004)	42	Software	↑
Odec (2004)	620	Road	↓
Pickrell (1989)	10	Rail	–
Sauer et al (2007)	412	IT	↑
Shehu (2014)	359	Construction	↑
Skitmore and Drew (2003)	89	Construction	↑
Van Oorshot et al (2005)	108	Software	–
Yang et al (2008)	112	Software	↑
Lorentzen et al (2017a)	79	Petroleum	↑

A few case studies have been conducted on cost overruns in oil and gas projects on the NCS. For instance, the Investment Committee conducted a thorough investigation of 13 projects between 1994 and 1998. Another study by the Norway's Office of the Auditor General presented a similar analysis of three projects between 1995 and 1996. Finally, the Norwegian Petroleum Directorate (NPD) investigated five projects executed between 2006 and 2008. These case studies conclude that cost overruns are a widespread phenomenon among oil and gas projects on the NCS. Unrealistic assumptions, insufficient planning, and underestimating risks and challenges related to new technology have been identified as some of the root causes for such overruns. Cost overruns in oil and gas projects on the NCS have also been studied in an econometric setting. Emhjellen and Osmundsen (2003), Lorentzen & Osmundsen (2017), Dahl et al (2017), Lorentzen et al (2017a) and Lorentzen (2017b), for instance, confirm their prevalence. Emhjellen and Osmundsen (2003) suggest that the occurrence of cost overruns can partly be explained by errors in estimating and reporting capital expenditure costs. Dahl et al (2017) and Lorentzen et al (2017a) find no evidence for a significant linear relationship between cost overruns and project size, but suggest that the inverse of project size has a significant effect on overruns. Lorentzen and Osmundsen (2017) investigate a more complex nonlinear relationship using multivariate fractional polynomials. Finally, Lorentzen et al (2017b) find no support for a linear relationship when applying OLS, but evidence of a significant and positive effect when applying a least absolute deviation (LAD) approach.

3.0 Data

The data set consists of 160 different offshore oil and gas development projects on the NCS between 1988 and 2016. All data have been extracted from the National Budget and the Investment Committee. The data set utilised consists of current control estimates (CCE_{it}) for each year (t) in the execution of a given project (i). The CCEs are inflation-adjusted to 2015. The initial budget (CCE_0) available at project sanctioning and the latest available budget (CCE_T) are used as inputs for calculating the relative cost overrun. See equation (1). It should be noted that not all projects in the data set were fully implemented at the time the cost was registered. Hence, projects at the end of the sample period are subject to change. Furthermore, the timing of the initial budget has been contested in the literature. If cost overruns are regarded as a tool for evaluating past decision-making, it makes sense to use the budget available at the time the decision was made to undertake the project in question. On the other hand, if the project scope were to change so that it can be regarded in part as a different project, the initial budget becomes (partly) invalidated. Given the limitations in the data set, change of scope cannot be accounted for. Hence, I opt to utilise the budget available when the project is sanctioned.[1]

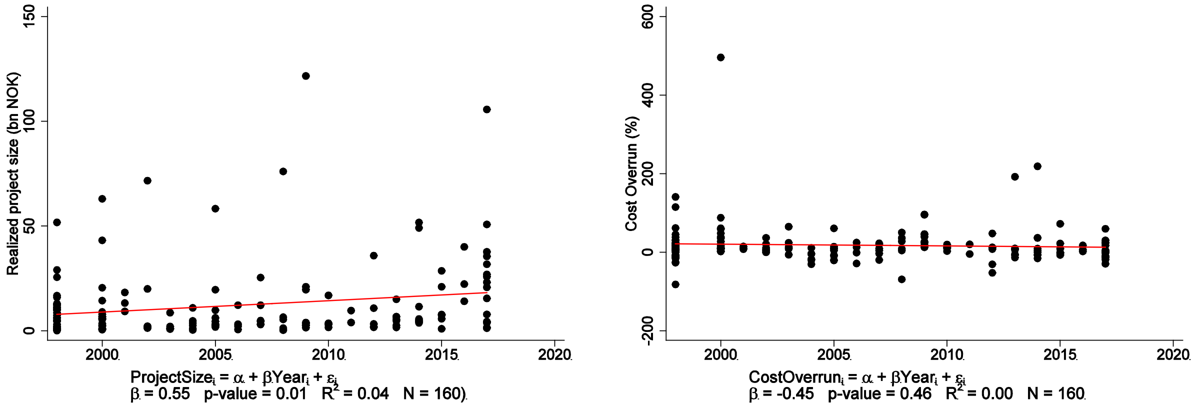
$$CostOverrun_i = \frac{CCE_{iT}}{CCE_{i0}} - 1 \tag{1}$$

Figure 1 shows the temporal development in project size and cost overruns. Project size is measured here as the investment size of the project in NOK billion adjusted to 2015 value at the end of project implementation (CCE_T). As can be observed from figure 1 (a), project size contains some variability, but it can be inferred from regressing size on time that the size of the projects tends to increase over time. In figure 1 (b), however, we see no significant trend in cost overruns, indicating that the ability to predict project execution costs accurately has neither increased nor decreased. This is consistent with the findings of Dahl et al (2017) and Lorentzen et al (2017a, 2017b).

[1] Further information on the content of the budgets can be found in the guidelines for a plan for development and operation of a petroleum deposit (PDO) and a plan for installation and operation of facilities for transport and utilisation of petroleum (PIO) of 4 February 2010.

Figure 1: Temporal development in project size and cost overruns on the NCS (1988-2016)

(a) Project size (NOK bn) (b) Cost overruns (%)

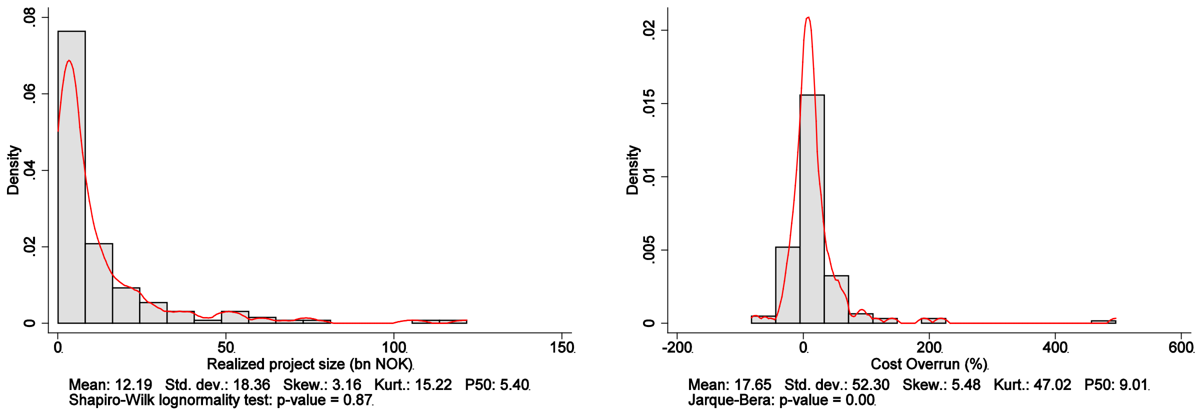


The figure shows temporal development in project size and cost overrun through a scatter plot with a fitted OLS regression line. The end of project execution is marked on the x axis.

Inspection of the statistical distribution of project size and cost overruns reveals some interesting characteristics of the behaviour of these variables. See figure 2 (a) for project size. We observe that the distribution has similarities with a lognormal one. That implies we have a high number of smaller projects compared with larger ones. Analogously, we observe in figure 2 (b) that cost overruns are characterised by a distribution with a positive mean (17.65 per cent), a large standard deviation (52.3 percentage points), leptokurtosis (47.02) and positive skewness. Positive skewness implies that the distribution exhibits a long right tail. In other words, instances occur with very large cost overruns. A further implication is that the average cost overrun becomes a poor description of how the typical project performs. While the mean cost overrun is 17.65 per cent, for instance, the median is 9.01 per cent.

Figure 2: Distribution of project size and cost overruns on the NCS (1988-2016)

(a) Project size (NOK bn) (b) Cost overruns (%)



The figure shows distributions approximated through a histogram and Epanechnikov kernel density plot.

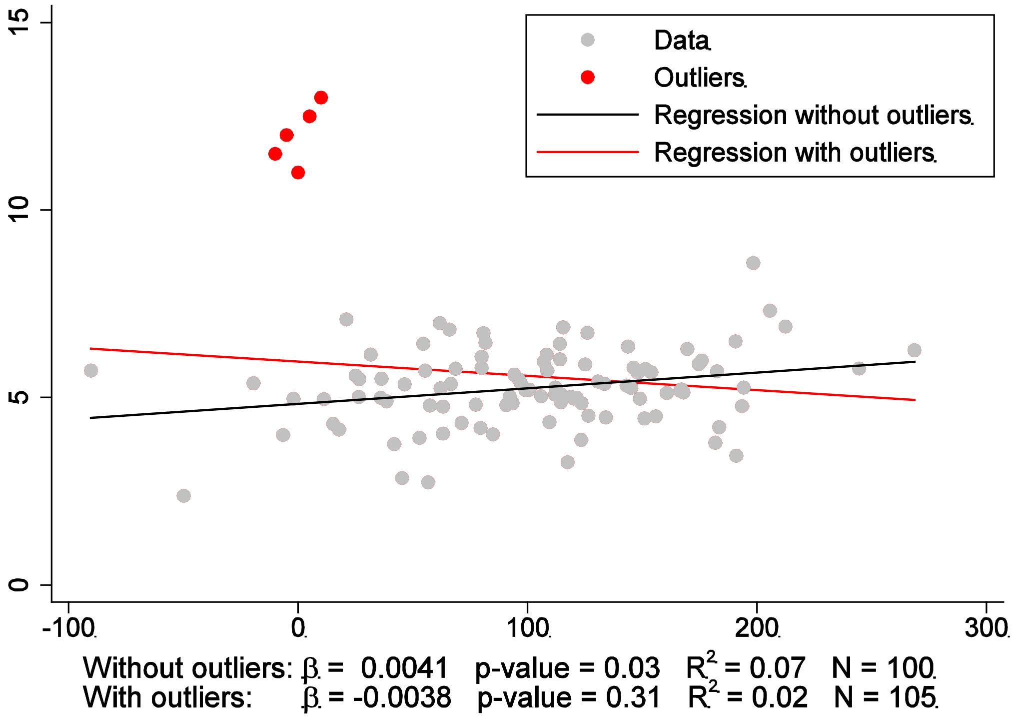
4.0 Analysis of relationship between project size and cost overruns

When analysing a sample of data – such as offshore oil and gas development projects on the NCS, for instance – the statistical mean can be a fairly misleading statistic in the presence of outliers. A better alternative could be to use the median (50th percentile), since this is far more robust to outliers. The choice between these two statistics is rooted in the characteristics of the variable of interest. As a guideline, consider the following question: is it probable that one outlier could significantly move the average? For our main variable of interest, cost overruns, the average value is arguably sensitive to outliers. While using the relative rather than the absolute cost overrun is a step towards making the measure more robust to outliers, it might not be sufficient. According to the PDO guidelines, the budget at the time the project is sanctioned should be accurate to the extent that 95 per cent of projects are realised at a cost within $\pm$ 20 per cent of the budget. As demonstrated in section 3, this is clearly not the case. That has implications for the choice of econometric methodology when assessing the relationship between cost overruns and project size.

OLS regression is considered to be the workhorse of econometric modelling. It can be a fairly powerful tool and is consequently frequently utilised as an econometric approach to quantifying the relationship between cost overruns and project size. While OLS comes with the caveat of a handful of rigid assumptions, I would primarily contend that OLS is essentially an estimate of the expected conditional mean, $E(CostOverrun_i|ProjectSize_i)$, and as such is sensitive to outliers in the same way as the unconditional mean, as argued in the preceding discussion. Likewise, in the presence of outliers, it could be more feasible to estimate the conditional median, $med(CostOverrun_i|ProjectSize_i)$, using quantile regression with LAD estimation. Specifically, I suggest that using OLS estimation instead of LAD is more likely to provide either an insignificant or a significantly negative estimate of the β coefficient when the true population parameter is positive.

This follows from two aspects: cost overrun is a constrained variable, and smaller projects outnumber larger projects. Combining these two aspects, it can be argued that outliers of cost overruns will tend to be both positive and confined to the lower percentiles of project size. If project size really has a positive, albeit weak, effect on cost overruns, our OLS estimate will be negatively biased. See figure 3 for an illustration. This shows a scatter plot for a sample of simulated data and a fitted OLS regression line, which is compared with a regression line based on the same data with a few additional outliers. As shown, positive outliers at the lower end of the x axis will bias the regression downwards so that the initial positive and significant β coefficient becomes negative but statistically insignificant from zero. Since the mean cost overrun is likely to be biased by outliers, it would hence be more efficient to use LAD estimation. To corroborate the assertion that LAD is more efficient than OLS, I will provide evidence below through both a Monte Carlo simulation and an empirical example – which is further validated through bootstrapping – using offshore oil and gas development projects from the NCS.

Figure 3: Illustration of the impact of outliers



The figure illustrates how the OLS regression line is potentially influenced by the inclusion of outliers. Data are simulated.

4.1 Monte Carlo simulation

The following econometric approach is applied to assess the empirical relationship between cost overruns and project size. Inflation-adjusted relative cost overruns from offshore oil and gas development projects on the NCS are used as the dependent variable and inflation-adjusted project investment size in NOK billion is used as the independent variable. The model to be estimated is given in equation (2),

$$CostOverrun_i = \alpha + \beta ProjectSize_i + \varepsilon_i. \tag{2}$$

Two estimation techniques are applied, OLS and LAD, where the former is based on the conditional mean function which minimises $\sum_i \hat{\varepsilon}_i^2$ and the latter is based on the conditional median function which minimises $\sum_i |\varepsilon_i|$. The β estimators under OLS and LAD are given in equations (3) and (4) respectively,

$$\beta_{OLS} = (x'x)^{-1}x'y \tag{4}$$

and

$$Q(\beta_q) = \sum_{i:y_i \geq x_i'\beta}^{N} q\left|y_i - x_i'\beta_q\right| + \sum_{i:y_i < x_i'\beta}^{N} (1-q)\left|y_i - x_i'\beta_q\right|, \tag{4}$$

where x is the input data matrix, y the output vector and q the quantile.

The Monte Carlo simulation is implemented by executing a thousand iterations, where a hundred pairs of project size and cost overruns are generated. Project size is randomly drawn from a lognormal distribution and cost overrun is calculated as specified in equation (2). The error term is similar to project size generated from a lognormal distribution and the β coefficient is set to a low but positive value (0.0001). The results from the Monte Carlo simulation are shown in figure 4, which displays the distribution of the estimated β coefficient from the OLS regressions compared with the distribution of the coefficient from the LAD estimation. Some noteworthy findings are made. First, the mode of the LAD β coefficient distribution is closer to the true value, but the OLS distribution is closer in mean. Second, LAD has a smaller standard deviation. This implies that LAD is more efficient. Finally, OLS has a higher kurtosis and longer tails. This means the OLS estimator is more likely to give extreme values for the β coefficient in either direction — in other words, negative and significant results are more likely with OLS than the LAD estimator.

Figure 4: Simulation outcome for β coefficients (1 000 iterations with a sample size of 100)

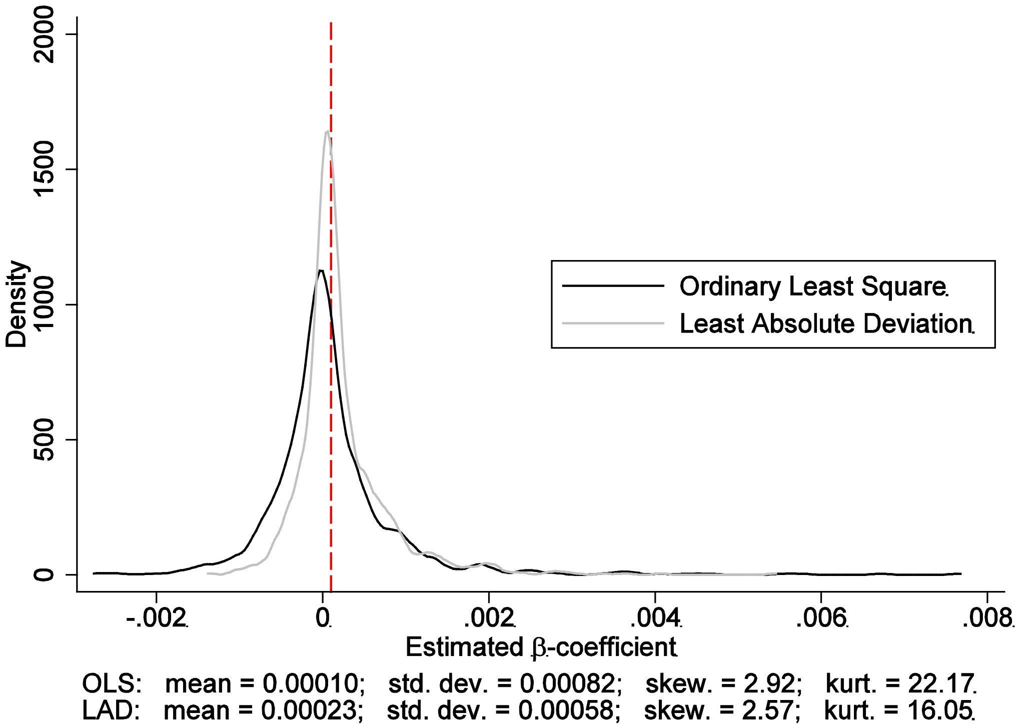


Table 2 shows the outcome of the Monte Carlo simulation in terms of whether the β coefficient is found to be significant with OLS and LAD estimations. The true value of the β coefficient is positive, albeit small, and as demonstrated in the table, LAD estimation obtains the correct result more often than OLS.

Table 2: Simulation outcome (1000 iterations of with a sample size of 100)

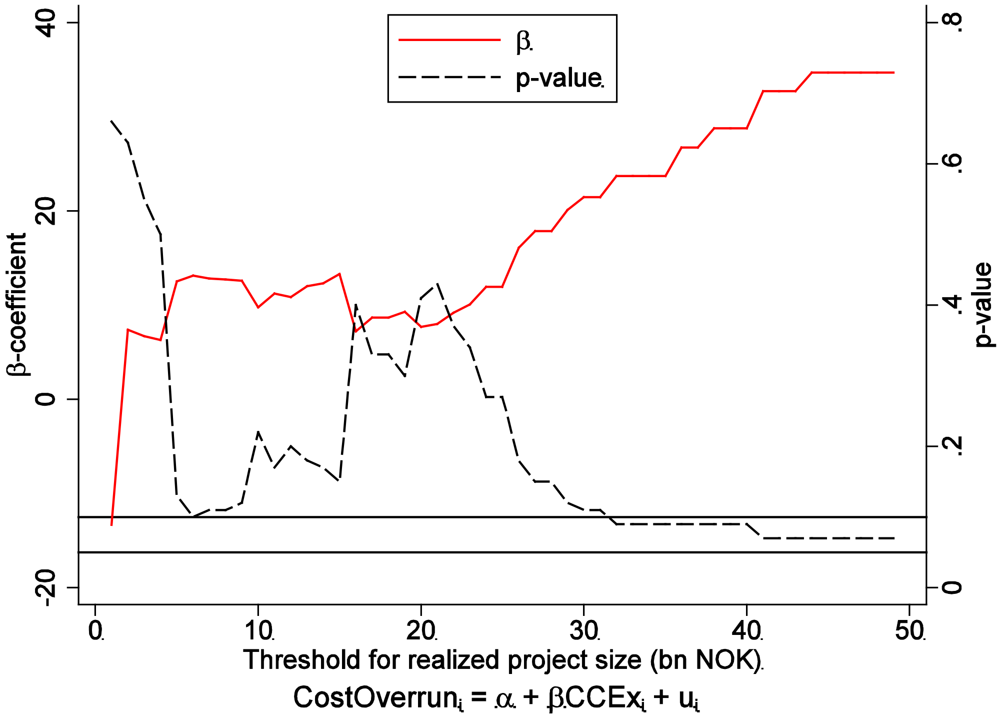
		LAD		
		Significant	Insignificant	Sum
OLS	Significant	61	21	82
	Insignificant	106	812	918
	Sum	167	833	1000

Number of iterations yielding either significant β coefficients from regressing cost overrun on project size under both OLS and LAD estimation.

4.2 Empirical results from the NCS

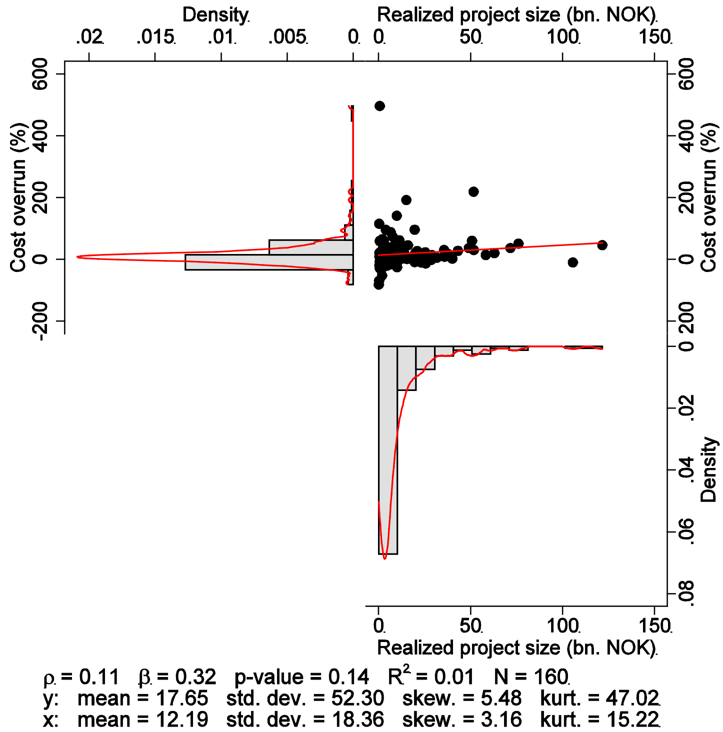
With insight from the simulation study, let us now return to the real data – the 160 oil and gas projects on the NCS in 1988-2016. Two approaches can primarily be taken to modelling the relationship between cost overrun and project size. The latter can be treated either as a continuous variable or as a dichotomous variable. In the case of a dichotomous project size variable, we have a variable which can take on the value of either one or zero. Typically, we assign the value of one to "large" projects. Operationalising project size in this way is certainly a valid approach, but no inherently intuitive way exists for defining the threshold between "large" and "small" projects. Figure 5 shows the results from regressing cost overrun on a dichotomous project size variable. The problem of an arbitrary dummy is solved by considering the whole spectrum of thresholds. As observed, the estimated β coefficients tend to increase with the threshold. However, the coefficients only become marginally significant at a 10 per cent level at the higher thresholds (> NOK 30 billion).

Figure 5: Regressing cost overrun on project size as a dummy variable with different thresholds



Results from regressing cost overrun on a continuous project size variable is shown in figure 6. The OLS regression line is shown here together with a scatter plot and the distributions of both dependent and independent variables. As with the previous regression, results are inconclusive in the sense that no significant relationship can be obtained. A positive relationship appears to exist – in other words, larger projects have more cost overruns – but the β coefficient cannot be said to be significantly different from zero.

Figure 6: Regressing cost overrun on project size as a continuous variable



Having demonstrated that OLS is incapable of providing a significant relationship between cost overruns and project size for this particular sample, I now turn to the LAD estimator. Table 2 shows the regression result as specified in equation (2) for both LAD and OLS estimations. As shown in table 3, regressing cost overrun on a continuous project size with LAD yields a positive and highly significant β coefficient – unlike the OLS coefficient, which is insignificant even at a 10 per cent level.

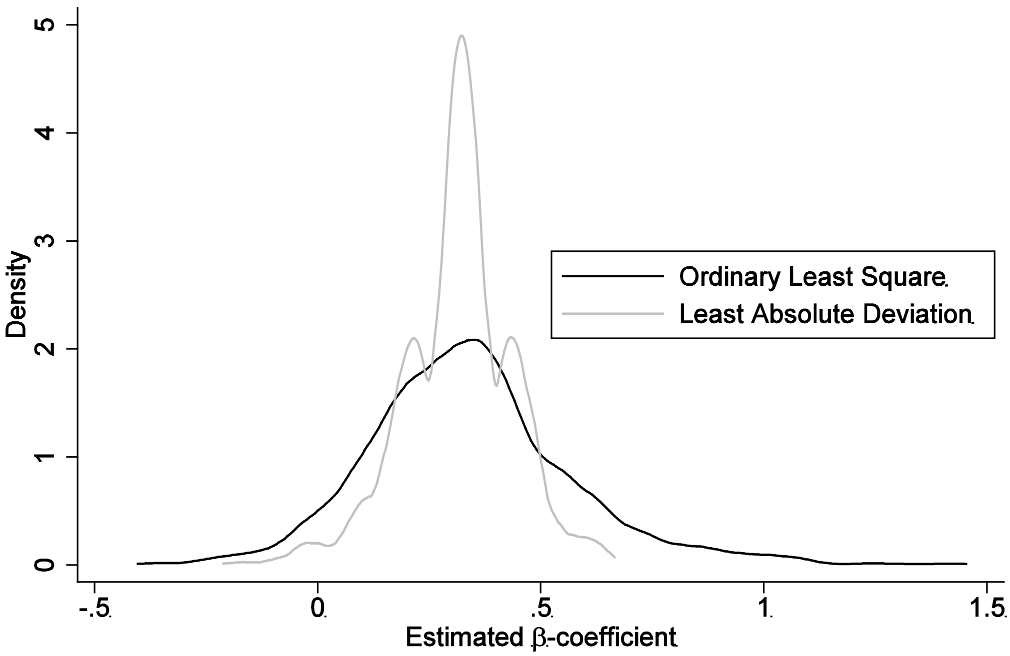
Table 3: Empirical relationship between cost overruns and project size on the NCS

	Dependent variable: cost overrun	
	(1)	(2)
Independent variable	LAD	OLS
Project size	0.329***	0.320
	(0.000)	(0.137)
Constant	5.877***	13.745**
	(0.004)	(0.011)
N	160	160
R^2		0.013
Pseudo R^2	0.0236	

Model (1) is implemented using OLS estimation and model (2) using LAD. The parentheses contain p values and the asterisk furthermore denotes the level of significance: *** p value < 0.01, ** p value < 0.05, * p -value < 0.1

To illustrate further how quantile regression is more appropriate than OLS estimation, I bootstrap the sample and plot the statistical distributions of the LAD and OLS β coefficients against each other. See figure 7. As observed, the bootstrap results predominantly conform with the results from the Monte Carlo simulation. Specifically, while the mode of the two distributions largely coincide, the OLS distribution exhibits a larger standard deviation and kurtosis. By comparison, since the OLS β covers a larger interval, it is far more likely to yield conflicting results across different samples.

Figure 7: Comparison of bootstrapped β coefficients from OLS and LAD



OLS: mean = 0.34203; std. dev. = 0.23125; skew. = 0.67; kurt. = 4.69
LAD: mean = 0.31518; std. dev. = 0.12161; skew. = -0.40; kurt. = 4.07

5.0 Conclusion

This article investigates the relationship between cost overruns and project investment size in offshore oil and gas projects on the NCS. The empirical literature appears to be nonconforming in the sense that both significant – either positive or negative – or insignificant results are frequently found. I suggest that this instability is a combined result of the statistical characteristics of cost overruns coupled with an inadequate econometric approach. Projects are unique by definition – they are inherently extensively idiosyncratic. This idiosyncratic property has implications for the behaviour of cost overruns. Specifically, we should expect cost overruns to be both heteroscedastic and riddled with outliers. It is these outliers which make the typical econometric approach of OLS inadequate at estimating the relationship between overruns and project size. The statistical mean is considered an unsatisfactory description of a variable when the data contain outliers which can potentially move the average significantly. In such cases, it is better to use the median as a statistic. OLS regression is an estimation of the conditional mean function, and is thus prone to suffer from the danger of outliers.

I propose that it would be better to use more robust estimation techniques such as quantile regression using LAD, which estimates the conditional median function. If a positive relationship really exists between cost overruns and project size, as suggested by the theoretical literature, it is likely that outliers will tend to bias the β coefficient downwards so that it either becomes insignificant or even significantly negative when using OLS. This follows from the idiosyncratic behaviour of projects, which means that the data will probably contain outliers and that these are likely to feature a large cost overrun and a small project size. Cost overrun is a constrained variable. It can only take one value between -100 per cent and positive infinity. As such, we can have outliers with a 500 per cent cost overrun but not an underrun of 500 per cent. Furthermore, since smaller projects are more abundant than megaprojects, there are more chances for getting an outlier at the lower end of the project size scale.

Through a Monte Carlo simulation and an empirical example with data from the NCS, I find support for this notion that OLS estimation is less adequate than LAD estimation. Through simulated data, we observe that the statistical distribution of the β coefficient tends to be narrower for LAD compared with OLS. This makes LAD a more efficient estimator than OLS. Since the true β is likely to be positive, albeit close to zero, the implication is that OLS more frequently yields a negative or insignificant β coefficient. Using real data, regressing cost overrun on project size results in an insignificant β coefficient when using OLS estimation. Utilising LAD estimation, however, the coefficient for project size becomes highly significant at the one per cent level. Further validation by bootstrapping the OLS and LAD β coefficient yields similar results to the Monte Carlo simulation. The distribution of the LAD coefficient is narrower than the OLS, thus making the former more efficient.

Literature

Avotos, I (1983). “Cost-relevance analysis for overrun control”. *International Journal of Project Management*, 1(3): 142-148.

Ansar, A, Flyvbjerg, B, Budzier, A and Lunn, D (2016). *Big is Fragile: an attempt at theorizing scale*.

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

Dantata, Nasiru A, Touran, Ali, and Schneck, Donald C (2006). "Trends in US rail transit project cost overrun”. *Transportation Research Board*.

Dahl, Roy Endré, Lorentzen, Sindre, Oglend, Atle, and Osmundsen, Petter (2017). "Pro-cyclical petroleum investments and cost overruns in Norway”. *Energy Policy* 100: 68-78.

Flyvbjerg, Bent, Skamris Holm, Mette K, and Buhl, Søren L (2004). "What causes cost overrun in transport infrastructure projects?" *Transport reviews* 24, no 1: 3-18.

Emhjellen, K, and Osmundsen, P (2003). “Cost estimation overruns in the North Sea”. *Project Management Journal*, 34(1): 23-29.

Gray, Andrew R, MacDonell, Stephen G, and Shepperd, Martin J (1999). "Factors systematically associated with errors in subjective estimates of software development effort: the stability of expert judgment". *Software Metrics Symposium, 1999. Proceedings. Sixth International*: 216-227. IEEE.

Hatton, Les (2007). "How accurately do engineers predict software maintenance tasks?". *IEEE Computer*, 40, no 2: 64-69.

Heemstra, F J, and Kusters, R J (1991). "Function point analysis: Evaluation of a software cost estimation model". *European Journal of Information Systems* 1, no 4: 229-237.

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, no 1: 13-21.

Jørgensen, M, Halkjelsvik, T, and Kitchenham, B (2012). “How does project size affect cost estimation error? Statistical artefacts and methodological challenges”. *International Journal of Project Management*, 30(7): 839-849.

Lorentzen, S, Oglend, A, and Osmundsen, P (2017a). “Cost overruns on the Norwegian continental shelf: the element of surprise”. *Energy*.

Lorentzen, S, Osmundsen, P, and Sandberg, F H (2017b). “Kostnader på norsk sokkel”. *Samfunnsøkonomen*, no 2: 77-91.

Lorentzen, S, and Osmundsen, P (2017). “Cost overruns on petroleum projects. The effect of project size”. *15th IAEE European Conference*.

Merrow, E W (2011). I*ndustrial Megaprojects: Concepts, Strategies, and Practices for Success*. Chichester, UK. John Wiley and Sons Ltd.

Moløkken-Østvold, Kjetil, Jørgensen, Magne, Tanilkan, Sinan S, Gallis, Hans, Lien, Anette C, and Hove, S W (2004). "A survey on software estimation in the Norwegian industry". *Software Metrics, 2004. Proceedings. 10th International Symposium on*: 208-219. IEEE.

Odeck, James (2004). "Cost overruns in road construction — what are their sizes and determinants?" *Transport policy* 11, no 1: 43-53.

Pickrell, Don H (1989). *Urban rail transit projects: forecast versus actual ridership and costs*. Final Report. No PB-90-148693/XAB. Transportation Systems Center, Cambridge, MA (USA).

Sauer, Chris, Gemino, Andrew, and Reich, Blaize Horner (2007). "The impact of size and volatility on IT project performance." *Communications of the ACM* 50, no 11: 79-84.

Skitmore, M and Drew, D (2003). “The analysis of pre-tender building price forecasting performance: a case study”. *Engineering, Construction and Architectural Management*, 10(1): 36-42.

Shehu, Z, Endut, I R, Akintoye, A, and Holt, G D (2014). “Cost overrun in the Malaysian construction industry projects: A deeper insight”. *International Journal of Project Management*, 32(8): 1471-1480.

Taleb, N N (2007). *The Black Swan: The Impact of the Highly Improbable*. Random House.

Van Oorschot, K E, Bertrand, J W M, and Rutte, Christel G (2005). "Field studies into the dynamics of product development tasks". *International Journal of Operations & Production Management*, 25, no 8: 720-739.

Yang, Da, Wang, Qing, Li, Mingshu, Yang, Ye, Ye, Kai, and Du, Jing. (2008) "A survey on software cost estimation in the Chinese software industry". *Proceedings of the Second ACM-IEEE international symposium on empirical software engineering and measurement*: 253-262. ACM.