AC 2010-3: INCORPORATING UNCERTAINTY INTO LEARNING CURVES: A CASE STUDY IN OIL DRILLING ESTIMATES

Christopher Jablonowski, University of Texas, Austin

Christopher Jablonowski is Assistant Professor of Petroleum and Geosystems Engineering at the University of Texas at Austin where he performs research on decision-making under uncertainty, industrial organization, and safety management systems. Prior to joining the University of Texas at Austin, he worked as an upstream project analyst with IPA, Inc., an economist with the US Government, and as a drilling engineer with Shell Offshore Inc. He holds a B.S. in Civil Engineering from Virginia Tech, a M.B.A. from Tulane University, and a Ph.D. in Energy, Environmental, and Mineral Economics from the Pennsylvania State University.

Incorporating Uncertainty Into Learning Curves: A Case Study in Oil Drilling Estimates

Abstract

In capital projects that contain repeated activities, and in manufacturing processes, it is common for cost and schedule performance to improve over time. This trend in improvement is commonly referred to as a "learning curve." When learning is anticipated, cost and schedule estimates may be reduced dramatically relative to an assumption of no learning. Therefore, it is a best practice to account for a learning effect.

In cases where comparison projects exist, estimating a learning curve for a prospective project can be done with some certainty. This form of deterministic learning is a well established topic in engineering education. However, in cases where the sample of comparison projects is small, there may be significant uncertainty in the rate and magnitude of learning over time, and some form of probabilistic learning is more appropriate. This form of learning is not well established in the research literature nor in the educational domain.

This paper employs a case study approach to investigate options for systematic integration of learning curves in cost and schedule estimates. Deterministic and probabilistic learning are discussed and demonstrated. All computations are made using off-the-shelf spreadsheet software. The results provide engineers and decision-makers with a refined representation of uncertainty, and can improve capital investment valuation and decision-making.

This case study is intended to be used in an undergraduate course in engineering economy or project economics and addresses several educational objectives: it introduces the basic concept of a learning curve, it provides an opportunity to reinforce basic curve fitting methods, and it highlights the value of a probabilistic approach to engineering and economic problems.

Introduction

In capital projects that contain repeated activities, and in manufacturing processes, it is common for cost and schedule performance to improve over time. This trend in improvement is commonly referred to as a "learning curve." In the oil drilling industry many projects contain multiple well drilling campaigns, and learning curves are prevalent.[3] When learning is anticipated, cost and schedule estimates may be reduced dramatically relative to an assumption of no learning. Many operators consider the use of learning curves a best practice, and provide procedures for estimation and implementation in their estimating guidelines.

In cases where comparison projects exist, estimating a learning curve for a prospective project can be done with some certainty. This form of deterministic learning is a well established topic in engineering education. However, in cases where the sample of comparison projects is small, there may be significant uncertainty in the rate and magnitude of learning over time, and some form of probabilistic learning is more appropriate. This form of learning is not well established in the research literature nor in the educational domain.

This paper employs a case study approach to investigate options for systematic integration of learning curves in cost and schedule estimates. Brief reviews of probabilistic estimating methods and learning curves are provided. Deterministic and probabilistic learning are discussed and demonstrated. For each method, the key assumptions are itemized and discussed and a demonstration is provided. While no single procedure will fit every situation, it is concluded that the general method is straightforward, transparent, and can be implemented using off-the-shelf spreadsheet software. The proposed procedures generate results that provide engineers and decision-makers with a refined representation of uncertainty, and can improve capital investment valuation and decision-making.

This case study is intended to be used in an undergraduate course in engineering economy or project economics and addresses several educational objectives: it introduces the basic concept of a learning curve, it provides an opportunity to reinforce basic curve fitting methods, and it highlights the value of a probabilistic approach to engineering and economic problems.

Probabilistic Estimating

Before introducing the concept of learning curves, it is helpful to review current practice in probabilistic estimating. Probabilistic cost and schedule estimating is often employed in industries where uncertainty in outcomes is large and where the variance in estimates can influence decision-making. The oil well drilling engineering community is familiar with probabilistic analysis in well construction estimating, and in some drilling organizations, probabilistic estimating is a required practice. There are numerous papers that focus on theory, methods, and implementation.[2,9,14] Other research reports on empirical investigations.[1,4,8,11,12,15]

Conventional Methods. The conventional approach for probabilistic estimating is to first define a dependent variable of interest (typically a cost or time metric). The dependent variable is defined based on the purpose of the estimate. In oil drilling, estimates can be made for specific well activities, for an entire well interval, or for a whole well. Second, independent variables are specified as random variables based on an analysis of past performance. Third, the independent variables are sampled in a repetitive fashion and then used to develop simulated observations on the dependent variable. For example, the engineer may specify the rate of penetration (ROP) as a random variable, and then use a simulated result for ROP and the footage for a specific interval to compute interval drilling time. The procedure yields a simulated distribution of possible outcomes. The distribution of outcomes provides the engineer and other decision-makers with information not available from a deterministic estimate. In some decision settings, the range of outcomes can be more important than the expected value of the distribution.

The main challenges in implementing probabilistic estimating are data collection and specification of the independent variable probability distributions (shapes and parameters). Several of the aforementioned papers address this issue and provide guidance. There are also other hurdles to implementation of probabilistic estimating. A survey of the global drilling community conducted in 2004 indicated that a large majority of respondents believed that probabilistic analysis has value. But the survey also revealed that respondents felt the analysis took more time, that support tools were lacking, and that additional training was required.[4] If the use of probabilistic analysis is to increase, it is important that workflows and methods remain simple in structure, easy to use, and transparent regarding the assumptions. One means to address these concerns is to perform the analysis using desktop spreadsheet tools. The advantage of this approach is that all of the assumptions and computations are transparent and subject to review and discussion.

Other Methods. There are other methods that can be used to make probabilistic estimates. One of these alternatives is regression analysis. There is a limited literature on regression analysis as applied to drilling operations.[7,10,13] But regression analysis can also be used to make probabilistic estimates.[6] The advantage of a regression based approach is that less information and fewer assumptions are typically required. For example, there is no need to specify independent variable distribution shapes and parameters. When a more abstract specification of the problem is acceptable, regression analysis may be the preferred alternative.

***Statistical Dependence*.** One of the more complicated aspects of implementing a probabilistic analysis is the issue of statistical dependence. For example, when simulating sequential intervals on an individual well, one must specify whether the outcome for a deeper interval is conditional on the outcome of a shallower interval, or if they are independent of each other. In most empirical studies, the simulated variables are often assumed to be statistically independent to simplify the analysis, that is, $Prob(A|B)=Prob(A)$. In many cases, this assumption is reasonable. That is, even in cases where one interval has significant problems, the next interval can be assumed to be independent after the previous interval is secured (with steel casing and cement). However, this is not always an appropriate assumption. For example, problems in a shallow interval may lead to an early casing setting depth and compromise performance in subsequent intervals. The same issue arises when modeling a drilling campaign, that is, explicit specification of inter-well dependence.

The challenges in implementing an analysis with dependence originate in data collection and estimating relationships, and deciding how to systematically incorporate these assumptions into the analysis. The data requirements are significant, relationships must be estimated, and assumptions must be specified for *how* new information will be incorporated into subsequent well plans. Researchers are actively investigating methods for incorporating statistical dependence into probabilistic modeling.[2]

Learning Curves

For the purposes of this discussion, it is helpful to distinguish between the aforementioned type of statistical dependence and more general time dependence. Drilling performance tends to improve over time as engineers and operators acquire additional information about the subsurface and streamline operational procedures. For example, a second well may be more likely to have a poor outcome if the first well had a poor outcome, but this does not preclude the likelihood that the second well will be more efficient in many ways than the first well. This type of improvement is commonly referred to as a learning curve. Learning can result in a dramatic reduction in drilling time and cost. Therefore, when developing estimates for drilling campaigns, it is a best practice to investigate the implications of this effect. Methods for making these estimates are varied, and the choice of method depends on the available resources and the decision that the estimate is being used for. All that is typically required is historical data on drilling performance from relevant offset campaigns. Most learning curve models abstract from the technical attributes of individual wells and therefore provide a high level means to model improvements in drilling performance. In the remainder of this section, a summary of the most common methods is provided.

A Basic Model. It is straightforward to fit a curve to historical drilling data. Ikoku suggested a parsimonious specification as follows:[5]

$$y_n = an^b \qquad (1)$$

y_n = cost or duration metric of well n, n = order of well in drilling sequence, and a, b = parameters to be estimated. The parameters a and b can be estimated using non-linear regression. In some applications, this simple form of learning may be adequate.

Brett and Millheim Model. Brett and Millheim propose an alternate learning curve specification in their influential paper on the subject.[3] Their model is attractive because of its simplicity, and for the intuitive interpretation of the parameters. The Brett and Millheim specification is as follows:

$$y_n = C_1 e^{(1-n)C_2} + C_3 \qquad (2)$$

y_n = cost or duration metric of well n, n = order of well in drilling sequence, and C*'s* = parameters to be estimated. The C_2 parameter indicates the curvature or rate of learning. C_3 is the asymptote for large n, and can be interpreted as a technical limit (for an otherwise static system). The sum $C_1 + C_3$ represents the fitted value for the first well; thus, C_1 is the quantity to be "saved" via learning between the first well and the technical limit. This model has gained wide acceptance and has been used extensively in practice.[10,15] It is conceivable that more complex models could be specified. The incremental benefit of doing so, however, is probably small. The Brett and Millheim model appears to be adequate for drilling applications.

Proposed Methods for Incorporating Learning Curves In Probabilistic Estimates

In this section, a general framework and specific procedures are proposed for incorporating learning curves in probabilistic estimates. This proposal is, in part, a response to Zoller, Graulier, and Paterson, who suggest that "more research is required from the well engineering community on the application of learning curves to probabilistic data."[15] In this section, a three-step procedure is proposed and demonstrated.

Step 1: Probabilistic Estimate Without Learning (Base Case). In this step, the engineer makes a probabilistic estimate for each well, ignoring the learning effect. This can be done using a conventional probabilistic approach or a regression based approach described above. In conventional practice, an underlying performance metric is often simulated (e.g. ROP, running times, etc.) and then applied to wells whose individual characteristics vary. If instead a regression approach is used, the distribution of the forecast error is sampled repeatedly.

For this demonstration, a simplified conventional probabilistic approach is used. It is assumed that the decision-maker's objective is to model total campaign duration for a 9-well campaign of identical wells. It is assumed the wells are statistically independent, and that analysis of past performance suggests that the duration of a well is a normally distributed random variable with a mean of 45 days and standard deviation of 3 days (this is synthetic data, but is a realistic representation of drilling performance). Results from a 1000 trial Monte Carlo simulation are presented in Figures 1 and 2. Each trial results in one potential 9-well campaign. The sample mean equals 405 days and has converged to the specified mean as expected (9 wells x 45 days/well = 405 days). The P10 and P90 values are 394 days and 416 days. This scenario defines the "no learning" base case.

Figure 1: Sample of 50 Simulated Campaigns (out of 1000 trials) Without Learning

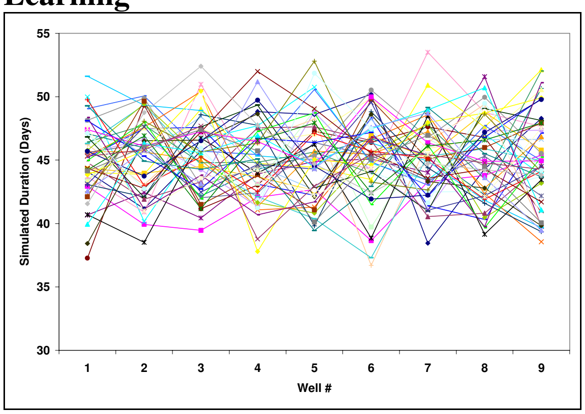
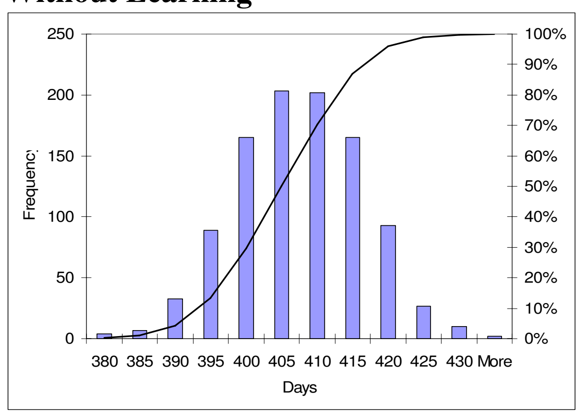


Figure 2: Histogram of Total Duration of Simulated Campaigns (1000 trials) Without Learning



***Step 2: Incorporating Learning*.** For this demonstration, the Brett and Millheim model of Equation (2) is employed. The learning curve can be applied to the previous result in a deterministic or probabilistic manner. The choice depends on the level of confidence that the engineer has in estimating the learning parameters.

*Deterministic Learning***.** If there is small uncertainty in the estimates of the learning equation's parameters, then deterministic learning is appropriate. For example, if the engineer is confident, based on analysis (learning curve estimation) of past performance of similar projects, that the parameters are $C_1 = 7$, $C_2 = 0.5$, and $C_3 = 38$ then a vector of deterministic learning factors can be computed using Equation (2). These factors are then applied to each of the simulated campaigns from the previous step. Note, these parameter values are consistent with a first well duration of 45 days from Step 1. The results of this step are presented in Figures 3 and 4. Visual inspection reveals the learning trend, but considerable scatter remains in each campaign because of the underlying uncertainty in drilling duration. The sample mean is decreased from 405 days to 360 days because of the learning effect. The P10 and P90 values are 350 days and 370 days. The impacts are significant in this example, and in practice, such a result is likely to affect economic analysis and decision-making.

Figure 3: Sample of 50 Simulated Campaigns (out of 1000 trials With Deterministic Learning

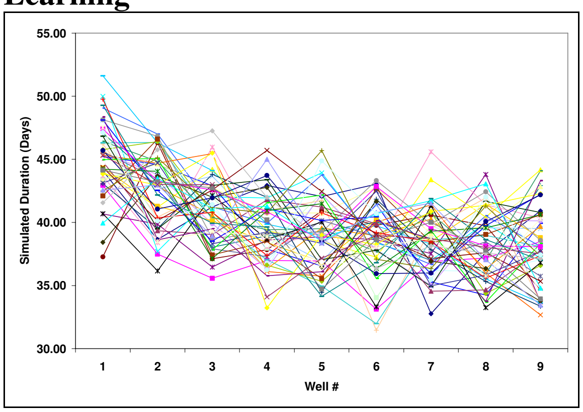
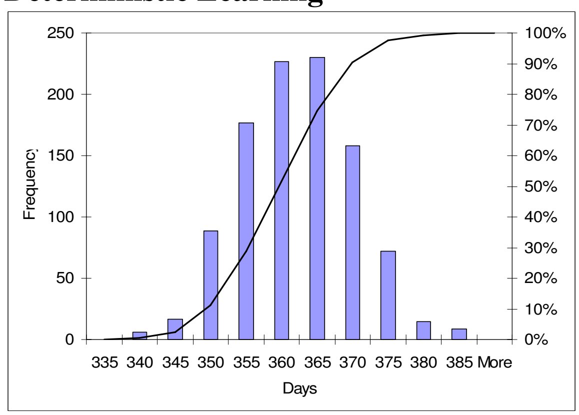


Figure 4: Histogram of Total Duration of Simulated Campaigns (1000 trials) With Deterministic Learning



*Probabilistic Learning***.** If there is large uncertainty in one or more estimates of the learning equation's parameters, then probabilistic learning is appropriate. For example, if the engineer is uncertain about the estimate of the parameter, say C_2, then this parameter can be defined in the simulation as a random variable. For example, assume that C_2 is a normally distributed random variable with a mean equal to 0.5 and standard deviation equal to 0.1. A new set of learning factors can be computed and applied to each of the simulated campaigns (holding the other parameters constant at their previous values). The results of this step are presented in Figures 5 and 6. A visual inspection of Figure 5 reveals more dispersion in the simulated campaigns as a result of the differential rates of learning in each iteration (relative to Figure 3). To be clear, there are two sources of variation in these simulations. The first source is the variation in the underlying drilling process, which exists regardless of how learning is occurring. The second source is the variation in the rate of learning. The sample mean is changed slightly to 362 days because of sample variation. The P10 and P90 values are also slightly changed to 349 days and 376 days, and the difference between these two values has increased relative to the deterministic case.

Figure 5: Sample of 50 (out of 1000) Simulated Campaigns With Probabilistic Learning

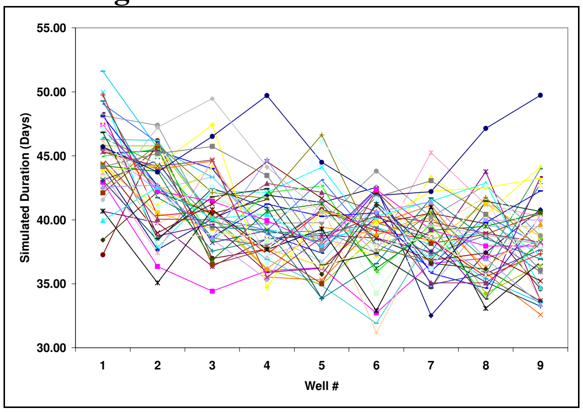
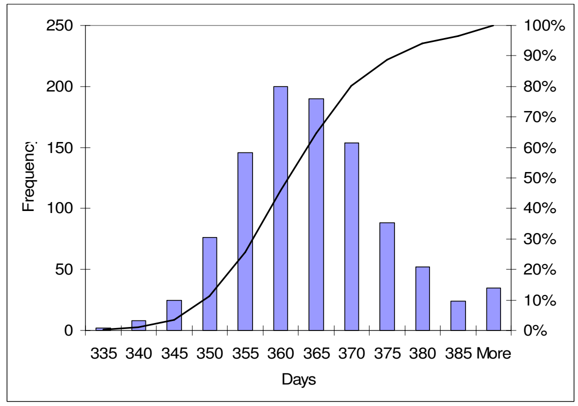


Figure 6: Histogram of Total Duration of Simulated Campaigns (1000 trials) With Probabilistic Learning



In this demonstration, the comparison of deterministic and probabilistic learning has revealed much about the problem. For example, it is observed that the specification of the C_2 parameter as a deterministic or random variable has an almost negligible impact on the probability distribution of the simulated campaigns. Therefore, if the engineer is confident in the estimate of the mean and standard deviation of C_2, then proceeding with the simpler deterministic approach would be reasonable in this case. Note that this is not a general result and in some cases the differences may be significant (e.g. when C_2 is specified with a large variance). Other specifications that investigate the other learning equation parameters, in isolation or jointly, are also possible but not presented here. For example, one could define the sum $\eta = C_1 + C_3$ as a random variable, specify the dimensionless ratio $\delta = C_1/(C_1 + C_3)$ as a random variable, use realizations on δ and η to solve for C_1 and C_3, then compute the learning factors.

Step 3: Post-Simulation Adjustments**.** Zoller, Graulier, and Paterson make the astute observation that when applying learning factors as described here, it is possible to predict a value for an individual well outside of the original range specified in the probabilistic analysis of Step 1.[15] The authors then propose and implement a novel correction for this effect. But it is arguable whether or not this is a significant problem. In practice, the variance of the base probabilistic estimate does not typically include future learning effects. Also, one could argue that the original probabilistic estimate should be updated as wells are executed (mean and variance), although appropriate procedures in this regard will be case-specific. Other specifications of the correction are possible, for example:

$$y_n = Max\left[\overline{P}, P_n f_n\right] \qquad (3)$$

y_n = cost or duration metric of well n, $\overline{P}$ = a user-specified cutoff for minimum values, P_n = the probabilistic prediction for well n, and f_n = the learning factor for well n. In this case the engineer specifies the minimum value, or absolute technical limit, to be expected during the campaign. In the end, whether to adjust values is up to the engineer and will depend on the purpose of the estimate.

To demonstrate the effect of post-simulation adjustments, the rule in Equation (3) was applied to the previous probabilistic results with $\overline{P}$ = 38 days. The results of the adjustment are presented in Figures 7 and 8. A visual inspection clearly shows the baseline duration. The sample mean is increased to slightly to 367 days as expected. The P10 and P90 values are also slightly changed to 356 days and 378 days.

Figure 7: Sample of 50 (out of 1000) Simulated Campaigns With Probabilistic Learning and Cutoff

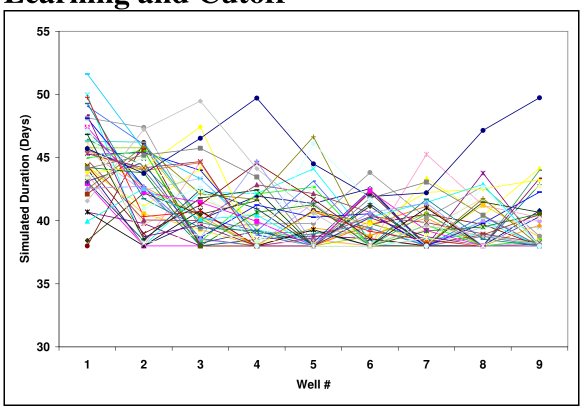
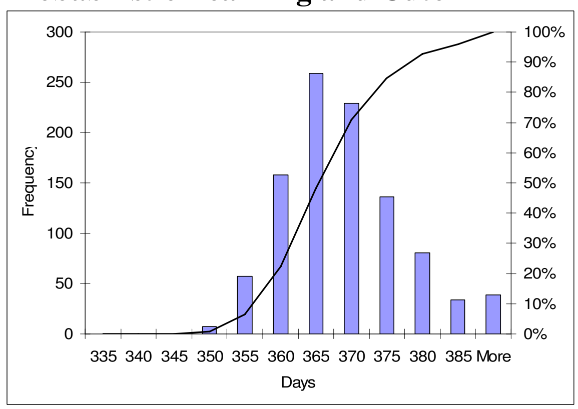


Figure 8: Histogram of Total Duration of Simulated Campaigns (1000 trials) With Probabilistic Learning and Cutoff



The results from all of the preceding computations are summarized in Table 1. This format for the results provides the engineer and decision-maker with a snapshot of the analysis and how the assumptions affect the variables of interest. As is the case in many decision-making settings, especially in drilling cost estimating, the analysis does not yield a recommendation. It only provides information that can be used to support a decision. Note, if an econometric analysis was also completed, the results could be added to this table in an identical manner to the probabilistic results.

Table 1: Summary of Campaign Estimates

Method for Individual Well Estimate	Learning Assumptions	P10 (days)	EV (days)	P90 (days)
Deterministic	None		405	
Probabilistic	None	394	405	416
	Deterministic	350	360	370
	Probabilistic	349	362	376
	Probabilistic with Cutoff	356	367	378

Conclusions

Under certain conditions (e.g. where the sample of comparison projects is small) there may be significant uncertainty in the rate and magnitude of learning over time. When developing a cost or schedule estimate in this setting, a probabilistic analysis of learning may be more informative than a deterministic one. While deterministic learning is a well established topic, probabilistic learning is not well established in the research literature nor in the educational domain.

This paper develops procedures for systematic integration of learning curves in cost estimates. Both deterministic and probabilistic learning are discussed and demonstrated. The results provide engineers and decision-makers with a refined representation of uncertainty, and can improve capital investment valuation and decision-making. For implementation of these methods to be successful, they should remain simple in structure, easy to use, and transparent regarding the assumptions.

This case study is intended to be used in an undergraduate course in engineering economy or project economics. It contains several teaching points: it introduces the basic concept of a learning curve, it provides an opportunity to reinforce basic curve fitting methods, and it highlights the value of a probabilistic approach to engineering and economic problems.

References

1. Adams, A.J., Gibson, C., Smith, R. 2009. Probabilistic Well Time Estimation Revisited. Paper SPE 119287 presented at the *SPE/IADC Drilling Conference and Exhibition* in Amsterdam, The Netherlands, March 17-19.
2. Akins, W., Abell, M., Diggins, E. 2005. Enhancing Drilling Risk & Performance Management Through the Use of Probabilistic Time & Cost Estimating. Paper 92340 presented at the *SPE/IADC Drilling Conference* in Amsterdam, The Netherlands, February 23-25.
3. Brett, J., Millheim, K. 1986. The Drilling Performance Curve: A Yardstick for Judging Drilling Performance. Paper 15362 presented at the *SPE Annual Technical Conference and Exhibition* in New Orleans, LA, October 5-8.
4. Hariharan, P., Judge, R., Nguyen, D. 2006. The Use of Probabilistic Analysis for Estimating Drilling Time and Costs While Evaluating Economic Benefits of New Technologies. Paper 98695 presented at the *IADC/SPE Drilling Conference* in Miami, FL, February 21-23.
5. Ikoku, C.U. 1978. Application of Learning Curve Models to Oil and Gas Well Drilling. Paper SPE 7119 presented at the *SPE California Regional Meeting*, San Francisco, April 12-14.
6. Jablonowski, C., MacEachern, D. 2009. Using Regression Analysis to Develop Probabilistic Well Construction Estimates. Presented at the *AADE National Technical Conference and Exhibition*, New Orleans, LA, March 31-April 2.
7. Kaiser, M., Pulsipher, A. 2007. Generalized Functional Models for Drilling Cost Estimation. *SPE Drilling and Completion*, June: 67-73.
8. Kitchel, B., Moore, S., Banks, W., Borland, B. 1997. Probabilistic Drilling-Cost Estimating. *SPE Computer Applications*, August: 121-125.
9. Murtha, J. 1997. Monte Carlo Simulation: Its Status and Future. *JPT*, April: 361-373.

10. Noerager, J., White, J., Floetra, A., Dawson, R. 1987. Drilling Time Predictions From Statistical Analysis. Paper 16164 presented at the *SPE/IADC Drilling Conference* in New Orleans, LA, March 15-18.
11. Peterson, S., Murtha, J., Roberts, R. 1995. Drilling Performance Predictions: Case Studies Illustrating the Use of Risk Analysis. Paper 29364 presented at the *SPE/IADC Drilling Conference* in Amsterdam, The Netherlands, February 28-March 2.
12. Peterson, S., Murtha, J., Schneider, F. 1993. Risk Analysis and Monte Carlo Simulation Applied to the Generation of Drilling AFE Estimates. Paper 26339 presented at the *SPE Annual Technical Conference and Exhibition* in Houston, TX, October 3-6.
13. Whelehan, O.P., Thorogood, J.L. 1994. An Automated System for Predicting Drilling Performance. Paper SPE 27487 presented at the *IADC/SPE Drilling Conference* in Dallas, TX, February 15-18.
14. Williamson, H., Sawaryn, S., Morrison, J. 2004. Some Pitfalls in Well Forecasting. Paper 89984 presented at the *SPE Annual Technical Conference and Exhibition* in Houston, TX, October 26-29.
15. Zoller, S., Graulier, J., Paterson, A. 2003. How Probabilistic Methods Were Used to Generate Accurate Campaign Costs for Enterprise's Bijupira & Salema Development. Paper 79902 presented at the *SPE/IADC Drilling Conference* in Amsterdam, The Netherlands, February 19-21.