Estimating Drilling Cost and Duration Using Copulas Dependencies Models

M. Al Kindi*, M. Al-Lawati and N. Al-Azri

Department of Mechanical and Industrial Engineering, College of Engineering, Sultan Qaboos University, Oman.

Abstract: Estimation of drilling budget and duration is a high-level challenge for oil and gas industry. This is due to the many uncertain activities in the drilling procedure such as material prices, overhead cost, inflation, oil prices, well type, and depth of drilling. Therefore, it is essential to consider all these uncertain variables and the nature of relationships between them. This eventually leads into the minimization of the level of uncertainty and yet makes a "good" estimation points for budget and duration given the well type. In this paper, the copula probability theory is used in order to model the dependencies between cost/duration and MRI (mechanical risk index). The MRI is a mathematical computation, which relates various drilling factors such as: water depth, measured depth, true vertical depth in addition to mud weight and horizontal displacement. In general, the value of MRI is utilized as an input for the drilling cost and duration estimations. Therefore, modeling the uncertain dependencies between MRI and both cost and duration using copulas is important. The cost and duration estimates for each well were extracted from the copula dependency model where research study simulate over 10,000 scenarios. These new estimates were later compared to the actual data in order to validate the performance of the procedure. Most of the wells show moderate - weak relationship of MRI dependence, which means that the variation in these wells can be related to MRI but to the extent that it is not the primary source.

Keywords: Archimedean copula, Monte Carlo simulation, Mechanical risk index (MRI).

<div dir="rtl">

تقدير تكلفة ومدة الحفر باستخدام نماذج تبعيات الروابط

م. الكندي*، م. اللواتي , ن. العزري

الملخص: يعتبر تقدير ميزانية ومدة الحفر من التحديات العالية المستوي في صناعة النفط والغاز. هذا بسبب العديد من الأنشطة الغير مؤكدة في إجراءات الحفر مثل سعر الخامات، التكاليف العامة، التضخم، سعر النفط، نوع البئر وعمق الحفر. فمن الضروري اعتبار كل هذه المتغيرات الغير مؤكدة وطبيعة العلاقة بينهما. في نهاية المطاف لأن هذا يؤدي إلى تقليل مستوى عدم اليقين وبالتالي تقدير موفق للتكلفة والمدة لنوعية بئر بعينه. في هذه المقالة نحن نقدم نظرية احتمال الرابط من أجل نمذجة الاعتمادية بين التكلفة والمدة ومؤشر الخطر الميكانيكي (MRI). مؤشر الخطر الميكانيكي هو حساب رياضي يربط العلاقة بين عوامل الحفر المختلفة مثل عمق المياه، العمق الذي تم قياسه، العمق الرأسي الحقيقي بالإضافة إلى وزن الطمي والإزاحة الأفقية. بشكل عام فإن قيمة مؤشر الخطر الميكانيكي تستخدم كمدخل لتقدير تكلفة الحفر والمدة. ومن هذا المنطلق نجد أهمية نمذجة التبعيات الغير مؤكدة بين مؤشر الخطر الميكانيكي وكل من التكلفة والمدة باستخدام نظرية احتمال Copulas. تقدير التكلفة والمدة لكل بئر تم استخراجها من نمذجة اعتمادية Copula حيث سنقوم بمحاكاة أكثر من ١٠٠٠٠ سيناريو. هذه التقديرات الجديدة سوف يتم مقارنتها مؤخراً مع المعلومات الحقيقية للتحقق من أداء هذا الإجراء. معظم الآبار أظهرت العلاقة مع مؤشر الخطر الميكانيكي من معتدل إلى ضعيف وهذا يعني أن التباين في هذه الآبار يمكن أن يكون ذو صلة ولكن إلى حد كونه ليس المصدر الابتدائي.

الكلمات المفتاحية: رابط أرخيميدس، محاكاة مونت كارلو، مؤشر الخطر الميكانيكي (MRI).

</div>

* Corresponding author's e-mail: kindim@squ.edu.om

1. Introduction

Quantifying dependencies between random variables has been used recently in many finance and insurance risk analysis. Copula dependency model is a superior technique to model dependencies and yet it has rarely been used in petroleum applications (Al-Harthy *et. al.* 2007). In general, copula is a function that links the marginal distributions of the random variables and generate the joint distribution given their dependence level. A copula can model the dependence structure given any type of marginal distribution, which is not possible with other correlation measures because it is able to separate the marginal distribution from the correlation. The two random variables are the MRI, and cost and duration. Cost and duration are two economically detrimental factors to be considered in decision making before venturing well drilling projects. Developing reliable procedures and methods for estimating cost and time in well drilling has been an active field of research. Early attempts started with regression analysis and adaptive methods (Noerager *et al.* 1987; Zoller *et al.* 2003). However, these basic methods overlooked the possibilities of the alternative scenarios that could happen and render the estimates far from reality. Another very intuitive flaw with those models is their dependence on linear regression which is only valid for linearly dependent variables (Aas 2006). Hence, the linear correlation coefficient is a meaningful measure of dependence but may induce misleading conclusions. An attractive alternative method of capturing the dependence can be the use of copula theory which can separate the marginal distribution from the correlation.

The use of copula theory in financial applications was introduced by Embrechts *et al.* (2002). They introduced copula theory for two reasons: first, because it is a principal tool in indicating the drawbacks of linear regression; second, it is an approach for understanding the general concept of dependency. Copula theory has also been very common in the oil and gas industry. Chiyoshi (2004) applied Archimedean copulas in order to estimate the time for drilling wells. Archimedean copula has been used because it has many families that are capable of representing different types of dependencies. Another research conducted by Al-Harthy *et. al.* (2007) found that the two common methods used to model the dependency in the petroleum industry *i.e.* envelop method and the Iman-conover method, failed to address the dependence structure. From this point of view, they illustrated and discussed the benefits of using Archimedean copulas to model dependencies on the reserves problem. Therefore, addressing the dependencies between two economical random variables with the MRI value demonstrates a new approach to give managerial insights to decision makers.

2. Copulas

Copula is a multivariate distribution whose one-dimensional margins are uniform on the interval [0, 1] (Pradier 2011). The definition of a d-dimensional copula is a multivariate distribution, C, with uniformly distributed marginal U (0; 1) on [0, 1].

Sklar's theorem states that every multivariate distribution F with marginal $F_1, F_2 \ldots \ldots F_d$ can be expressed in the form:

$$F(x_1, \ldots \ldots, x_d) = C\big(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)\big) \quad (1)$$

The popular families of copulas are called Archimedean copulas which is one of the two main families of copula along with the copulas of normal mixture distributions.

Archimedean copulas

The Archimedean copulas is defined as:

$$C(u, v) = \phi^{[-1]}\big(\phi(u) + \phi(v)\big) \quad (2)$$

For $u, v \in [0,1]$, $C(u, v)$ is the copula function with u and v represents the uniform distributions, ϕ is the generator and $\phi^{[-1]}$ is the inverse generator.

The main input for the generator for the different families of Archimedean copulas function is known as Kendall's tau correlation, τ:

$$\tau = \frac{c - d}{\binom{n}{2}} \quad (3)$$

where n is the sample size , c is the number of pairs that are moving in the same direction and d is the number of pairs that are moving in the opposite direction of each other.

The considered Archimedean copulas (Clayton, Gumbel, and Frank copulas) fall into the class of the so-called Laplace transform Archimedean copulas. Each family type of these

copulas has its own parameters. Table 1 and Table 2 show the values of these parameters.

Substituting the values of the generators in Eqn. (2) will give a different copula equation for each family.

Clayton copula: The Clayton copula is a systematic copula that exhibits greater dependence in the negative tail than it does in the positive one. This copula is mostly used to study correlated risks because of its ability in capturing lower tail dependence (Mahfoud 2012). Clayton copula is expressed as:

$$C_\theta(u,v) = \max([u^{-\theta} + v^{-\theta} - 1]^{-1/\theta}, 0) \qquad (4)$$

where θ, belongs to $[-1, \infty) \setminus \{0\}$, is a parameter that controls the dependence where zero is not included. Perfect dependence is obtained when θ tends to infinity. In contrast, when θ approaches zero then independence is implied. The main drawback associated with this type of copulas is the fact that it cannot account for negative dependence.

Gumbel copula: The Gumbel copula is also an asymmetric copula that exhibits greater dependence in the positive tail than it does in the negative one. This copula is expressed as:

Table 1. Archimedean copulas with their generators and ranges.

Family	Generators $\phi(t)$	Ranges of the Tau
Clayton (1978)	$\frac{1}{\theta}(t^{-\theta} - 1)$	$[-1, \infty) \setminus \{0\}$
Gubmel (1960)	$(-\ln t)^\theta$	$[1, \infty)$
Frank (1979)	$-\ln(\frac{e^{-t\theta}-1}{e^{-\theta}-1})$	$(-\infty, \infty) \setminus \{0\}$

Table 2. Kendall tau values as a function of theta for each family and their ranges.

Family	Ranges	Tau
Clayton (1978)	$[-1, \infty) \setminus \{0\}$	$\frac{\theta}{\theta + 2}$
Gubmel (1960)	$[1, \infty)$	$1 - \frac{1}{\theta}$
Frank (1979)	$(-\infty, \infty) \setminus \{0\}$	$1 - \frac{4}{\theta}(1 - D_1(\theta))$

$$C_\theta(u,v) = \exp(-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}) \qquad (5)$$

where $1 \leq \theta < \infty$ is a parameter that governs the dependence relationship.

Frank copula: The Frank copula is suitable for modeling data characterized by weak tail dependence. This copula is given by:

$$C_\theta(u,v) = -\theta^{-1}ln \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{(e^{-\theta} - 1)}\right) \qquad (6)$$

where $-\infty < \theta < \infty$ is a parameter that governs the dependence relationship.

2.1 Estimation

Several methods can be used for selecting the best copula family that fits the dependence relation. These methods are categorized as non-parametric and semi parametric. Chiyoshi (2004) presented the procedure in order to implement these methods. The copula selection method in this work will depend on the semi parametric estimate that is known as the likelihood function. This function is extended to include the Akaikes Information Criteria (AIC) which was presented by Chiyoshi (2004). Given a random sample, $\{(X_{1k,...,}X_{pk}): k = 1, ..., n\}$ from distribution $F_\alpha(x_1, ..., x_p) = C_\alpha\left(F_1(x_1), ..., F_p(x_p)\right)$, the usual procedure is to select the parameter α that maximizes the pseudo log- likelihood,

$$L(\alpha) = \sum_{k=1}^{n} \log[C_\alpha\{F_{1n}(X_{1k}), , F_{pn}(X_{pk})\}]. \qquad (7)$$

where C_α is the copula density and F_{in} is the rescaled empirical distribution function for each $1 \leq i \leq p$. and F_{in} is given by:

$$F_{in}(x) = \frac{1}{n+1} \sum_{j=1}^{n} 1(X_{ij} \leq x) \qquad (8)$$

After obtaining the likelihood, AIC will be utilized such that:

$$AIC = -2\log(\text{likelihood}) + 2K \qquad (9)$$

where K is the number of the parameters which in our case is one parameter. The best model is the one with the lowest AIC.

3. Case Study

Cost and time estimates in well drilling are very crucial in deciding the feasibility of a drilling project. Financial crisis has become more probable than before, which is a fact that rendered organizations more vulnerable to bankruptcy because of faulty decision making. Hence, organizations have started to invest more on optimal budgeting and tracking their consumption patterns, which will help in their future decision making. There are different methodologies proposed in the literature for optimal budgeting. In a survey on drilling cost and the complexity of estimation models, Kaiser and Pulsipher 2007, described the development of cost and complexity metrics in well drilling and studied several methodologies like, Joint Association Survey (JAS), Energy Information Administration (EIA), Mechanical Risk Index (MRI), Directional Difficulty Index (DDI), and Difficulty Index (DI). According to their review, the Joint Association Surveys (JAS) and Mechanical Risk Indexes (MRI) are the most popular methods used in evaluating the cost and complexity of drilling in the Gulf of Mexico. MRI was developed in the late 1980s when Conoco Inc. engineers were tasked to compare offset drilling data for a collection of offshore wells in the Gulf of Mexico (Kaiser and Pulsipher 2007). MRI takes into consideration, water depth (WD), measured depth (MD), kick off point (for sidetracks), true vertical depth (TVD) in addition to mud weight (MW) and horizontal displacement (HD) at MD (Williams *et al*. 2001).

3.1 Drilling Data

The historical data used in this work represents more than 300 different wells drilled in different parts of Oman between 2008 and 2012. The data contained information on well drilling direction, well function, various rigs operating in the well, well actual cost, well estimated cost, time to completion, estimated time to complete, and the MRI log for each well. Therefore, the first step was to sort the data according to well type names.

3.2 Copulas and Simulation

Mechanical Risk Index (MRI) has been used as the key input parameter in order to estimate the cost and time of drilling. In order to understand the relationship between actual cost/duration and MRI, coefficient of determination method was utilized. However, it was found that this coefficient cannot explain the variation within the different groups of MRI and the actual cost/duration data as shown in Table 3 and Table 4.

Therefore, copula was used in order to have better view on the nature of the relationship between actual cost/duration and MRI. Therefore, it is very essential to input the random MRI and the actual cost/duration data in a way that fits copula input variables such that values of $u = CDF^{-1}(MRI)$ and $v = CDF^{-1}(cost/duration)$. These values are the input to the generator equation of the best copula family. The flow chart of the model shows how copula has been implemented in order to reach to new estimates for each well type, which is shown in Fig. 1.

Table 3. Coefficient of determination values between actual cost and MRI per category.

MRI Range	R^2
0-300	11.8%
300-600	28.6%
600-900	1.3%
900-1200	0.0%
1200-1500	0.0%
1500-1800	23.3%
1800-2100	26.0%
2100-2400	8.7%
2400-2700	0.0%
2700-3000	7.5%
3000-3300	0.0%
3300-3600	0.0%

Table 4. Coefficient of determination values between duration and MRI per category.

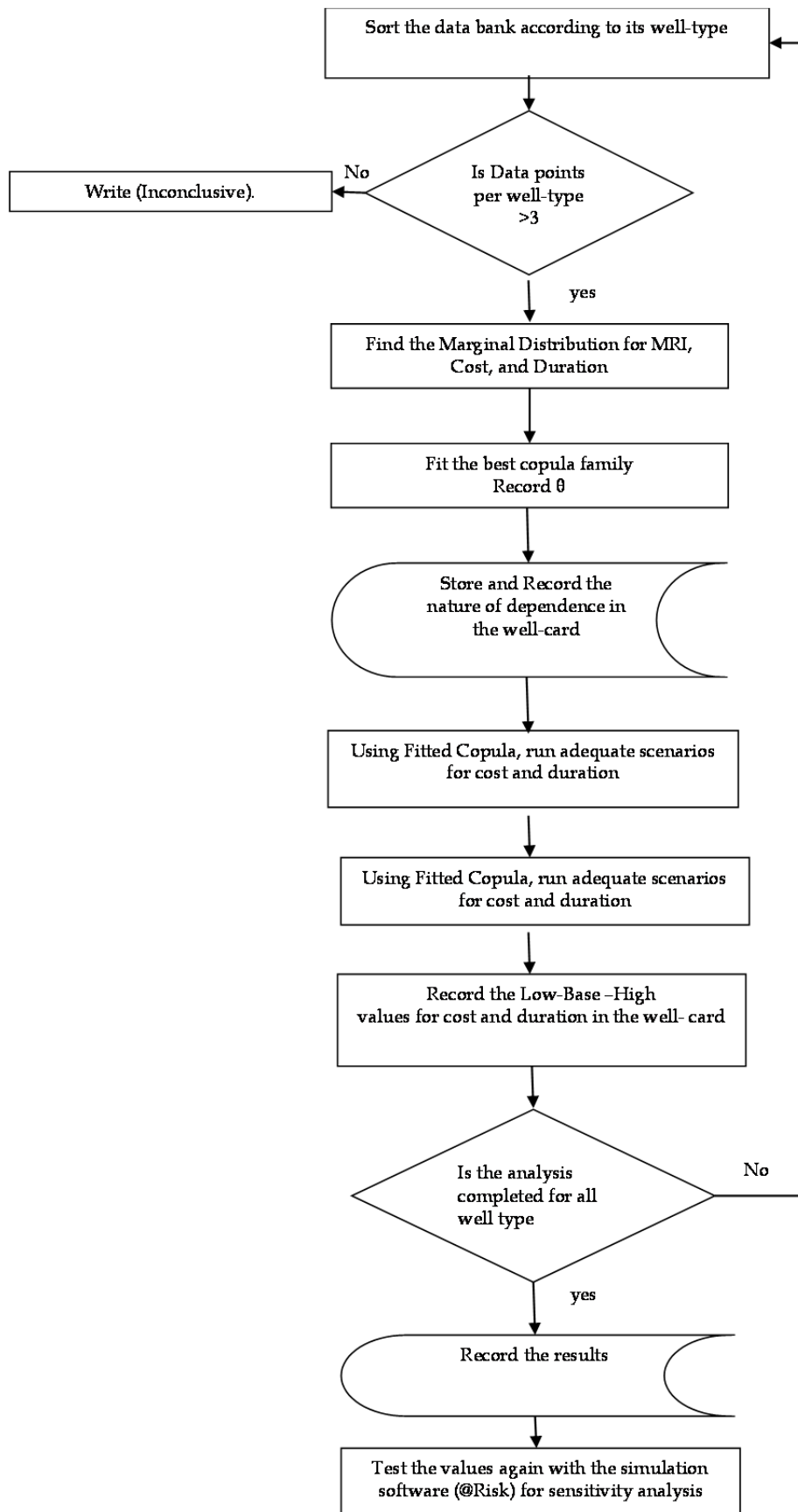| MRI Range	R^2
0-300	17.4%
300-600	24.4%
600-900	0.7%
900-1200	0.3%
1200-1500	0.0%
1500-1800	22.6%
1800-2100	13.6%
2100-2400	12.1%
2400-2700	0.0%
2700-3000	0.0%
3000-3300	0.0%
3300-3600	0.0%

Figure 1. Copula model flow chart.

The initial step is to know the type of copula family that represents the dependence relationship between MRI with actual cost and MRI with duration. It was found that Clayton copula represents the relation between MRI and the actual cost, while Gumbel copula represents the relationship between MRI and duration as shown in Table 5 and Table 6. The tables show the best fit for the copulas models using @risk software. Different goodness of fit criteria was compared such as: Akaike information criterion (AIC), Hannan–Quinn information criterion (HQIC), and Schwarz information criterion (SIC).

After learning/understanding the best copula family that fits the nature of the dependence, a 15000- scenario simulation was run. Each well type data had different value for the Kendall's tau that represents the relationship between MRI and actual cost/duration. Therefore, each will had its own scenarios based on this dependency structure. Figure 2 shows some examples of the simulation outcomes.

Table 5. Copula best fit between MRI and actual cost.

Bivariate Copulas	CopulaBiClayton	CopulaBiGumbel	CopulaBiFrank	CopulaBiT	CopulaBiNormal
MLE fits	Clayton(3.22,4)	Gumbel(2.61,1)	Frank(8.41,1)	T(7,0.8)	Normal(0.88)
Goodness of fit					
AIC	-2498.650074	-2484.865354	-2156.425193	-1938.887777	-1675.124018
AIC ranking	1	2	3	4	5
SIC	-2487.331731	-2473.547012	-2145.10685	-1927.569434	-1669.463904
SIC ranking	1	2	3	4	5
HQIC	-2494.510635	-2480.725916	-2152.285755	-1934.748339	-1673.053356
HQIC ranking	1	2	3	4	5
Log likelihood	1251.327863	1244.435503	1080.215423	971.4467147	838.5629507
Log likelihood ranking	1	2	3	4	5

Table 6. Copula best fit between MRI and duration.

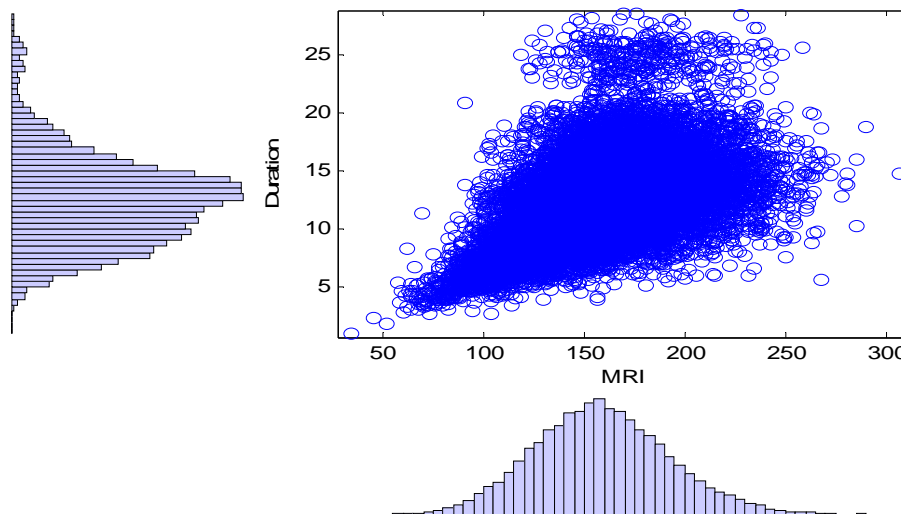| Bivariate Copulas	CopulaBiGumbel	CopulaBiClayton	CopulaBiFrank	CopulaBiT	CopulaBiNormal
MLE fits	Gumbel(2.608,1)	Clayton(3.21,4)	Frank(8.39,1)	T(7,0.869)	Normal(0.86)
Goodness of fit					
AIC	-2375.643145	-2365.262746	-2167.439404	-1905.495511	-1711.04269
AIC ranking	1	2	3	4	5
SIC	-2364.324803	-2353.944403	-2156.121062	-1894.177168	-1705.382576
SIC ranking	1	2	3	4	5
HQIC	-2371.503707	-2361.123307	-2163.299966	-1901.356072	-1708.972028
HQIC ranking	1	2	3	4	5
Log likelihood	1189.824399	1184.634199	1085.722528	954.7505817	856.5222866
Log likelihood ranking	1	2	3	4	5

Figure 2. MRI with cost/duration copula simulations examples.

From this simulation, three values were extracted for each well type. These values represent the low (10th percentile), base (50th percentile) and high values (90th percentile) of the expected estimates of drilling cost and duration. Then, sensitivity analysis was conducted using Monte Carlo simulation in 100 iterations where the probability distribution of the historical data for each well was fitted. The fitting was based on Anderson Darling test. This test depends on calculating the p-value for the goodness of fit test, which leads towards determining the best distribution that fits the data. The selection of the distribution depended on the one with the lowest Anderson Darling statistic value. The selected distribution was utilized with the three output values from copula model in order to obtain the optimum value for the estimates. After running the simulation, it was found that the optimal value is the median. This is supported by the output of Monte Carlo sensitivity analysis for well A as shown in Fig. 3 and Fig. 4.

The next step was to verify the estimates with the actual data. Additional data points were
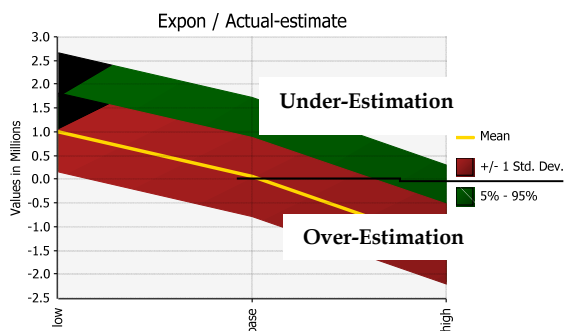

Figure 3. Monte Carlo sensitivity analysis for actual cost estimates for a given well.
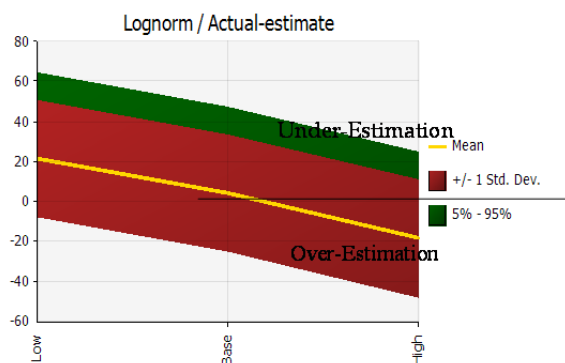

Figure 4. Monte Carlo sensitivity analysis for duration estimates for a given well.

provided in order to validate the calculated estimates. Four different indicators were used to test the efficiency of the presented model. The indicators are: (1) whether the actual cost/duration is within the range of the three predicted estimates, (2) the number of data points that show a lower deviation from the actual values compared to the company forecasted estimates when the base value obtained from copula model is used to compare with the actual values, (3) the number of data points that has a lower deviation from the actual values when the budgeting team picks the best value from the three estimates assuming that the team would pick the closest value to the actual one from the three values provided, and (4) the number of points that show a lower deviation if the team would choose the closest value to the company forecasted estimates from these three values.

4. Simulation Results

The historical data for the years 2008-2012 were used during the implementation of the model in order to get the new estimates. These new estimates were representing the upcoming actual cost and time required when the same well is drilled. The final step was to compare our model output with the real data in order to validate our model performance. New data was provided by the company for this purpose. However, the new dataset included few wells that are common with those recorded from 2008 to 2012 which limited the validation procedure to only 110 points and almost 20% of the number wells from the previous dataset. Table 7 below shows the results of the comparison through utilizing the four different indications

Table 7. The performance of copula model estimates.

Copula results	Indicator			
	(1)	(2)	(3)	(4)
Cost results	94 out of 110 (85%)	35 out of 110 (31.8%)	52 out of 110 (47.3%)	37 out of 110 (33.6%)
Duration results	65 out of 110 (59%)	34 out of 110 (31%)	83 out of 110 (75.5%)	75 out of 110 (68%)

In order to validate the accuracy of the model, we choose four performance indicators. Indicator (1) is to test whether the actual new data point is within the predicted range for both cost and duration. The second indicator (2) is to show if the predicted value using the model is better than the deterministic model used by the expert opinion in the company. The third indicator (3) is to let the expert choose one point (i.e. either low, base, or high value) and compare with the actual result. The last indicator is to choose the nearest prediction to the new actual data (*i.e.* if low prediction value is close to the new actual then it will be selected). The model presented in the paper shows a better prediction compared to previous work by Valdes *et al.* (2013) for a different data set. However, no previous literature is available to benchmark our results for the other remaining indictors.

5. Conclusion

In this paper, copula probability theory was implemented to model the dependency between MRI and drilling cost and duration. Typically, the Archimedean copulas was selected due to its ability to represent the different types of dependencies. The model selection depended on using the Akaikes Information Criteria method. This method was the expansion for the likelihood function. It was found that Clayton copula represents the dependence relationship between MRI and Cost with Gumbel copula represents the dependence relationship between MRI and duration. The model was extended to provide new estimates for the cost and time required to drill wells. At the end, validation of the estimates was done by utilizing several indicators. The results from the validation showed that the model was able to predict the new values through utilizing the historical data.

Acknowledgment

References

Aas K (2006), Technical report on "Modeling the dependence structure of financial assets: A survey of four copulas". *Norwegian Computing Center*, Oslo, Norway.

Al-Harthy M, Begg S, Bratvold RB (2007), Copulas: A new technique to model dependence in petroleum decision making. *Journal of Petroleum Science and Engineering* 57(1): 195-208.

Chiyoshi FY (2004), Modeling dependence with copulas: a useful tool for field development decision process. *Journal of Petroleum Science and Engineering* 44(1): 83-91.

Clayton D G (1978), A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 65(1): 141-151.

Embrechts P, McNeil A, Straumann D (2002), Correlation and dependence in risk management: properties and pitfalls. *Risk Management: value at risk and beyond* 176-223.

Frank MJ (1979), On the simultaneous associativity of F(x,y) and x+y-F(x,y). *Aequationes Math*, 1: 194-226.

Genest C, Ghoudi K, Rivest LP (1995), A semi-parametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82(3): 543-552.

Gumbel EJ (1960), Bivariate exponential distributions. *Journal of the American Statistical Association* 55(292): 698-707.

Jaworski P, Durante F, Hardle WK, Rychlik T (2010), Copula theory and its applications. New York: Springer.

Kaiser MJ, Pulsipher AG (2005), Rigs-to-reef programs in the Gulf of Mexico. *Ocean Development and International Law*, *36*(2): 119-134.

Mahfoud M (2012), Bivariate archimedean copulas: an application to two stock market indices. *Vrije Universiteit Amsterdam*, BMI Paper, Amsterdam.

Nelsen RB (1998), An Introduction to Copulas (Lecture Notes in Statistics).

Noerager JA, Norge E, White JP, Floetra A, Dawson R (1987), Drilling time predictions from statistical analysis. *Society of Petroleum Engineers* doi:10.2118/16164-MS.

Patton AJ (2012), A review of copula models for economic time series. *Journal of Multivariate Analysis* 110: 4-18.

Pradier E (2011), Copula theory: an application to risk modeling. Technical report, *Grenoble INP-Ensimag*.

Valdes A, McVay DA, Noynaert SF (2013), Uncertainty quantification improves well

construction cost estimation in uncon-ventional reservoirs. *Proceedings of SPE Unconventional Resources Conference Canada. 5-7 November, Calgary, Canada.*

Williams C, Mason JS, Spaar J (2001), Operational efficiency on eight-well sidetrack program saves $7.3 million vs historical offsets in MP 299 / 144 GOM. *Society of Petroleum Engineers* doi: 10.2118/67826-MS.

Zoller SL, Graulier JR, Paterson AW (2003), How probabilistic methods were used to generate accurate campaign costs for Enterprise's Bijupira and Salema development. In the proceedings of the SPE/IADC Drilling Conference, Amsterdam, The Netherlands. SPE eLibrary 79902.