Exponentiation Bias in Production and Reserves Estimates Made by Use of Exponential-Decline Models

Scott Morris, Earth Science Associates; **Gordon M. Kaufman**, Massachusetts Institute of Technology; and **John D. Grace**, Earth Science Associates

Summary

The standard method to evaluate an oil- or gas-production-decline curve estimated with an exponential function—taking the logarithms of both sides of the equation, estimating the parameters of the transformed function through linear regression, and exponentiating—leads to biased estimates of future production. The bias arises in the process of exponentiation.

The direction and magnitude of exponentiation bias depend on three driver variables: the variance of the post-peak-production history; the number of post-peak observations on production; and the estimated rates of production during the forecast period. A correction factor, dependent on the confluent-hypergeometric-limit function, applied to the biased estimators produces unbiased estimates of future production. The correction factor can be quickly evaluated and introduced into the work flow for use in evaluating exponential-decline curves.

The net bias in estimates of future production is more likely to be negative than positive. Negative bias understates remaining resources and reserves. The probability of negative, rather than positive net bias, is an increasing function of the maturity of production at the point of evaluation. The absolute magnitude of the bias is a direct function of both the variance of the empirical post-peak-production history and the forecasted rates of future production. It is an inverse function of the length of the post-peak-production history.

A data set of 54,254 completion-level monthly production histories from the Gulf of Mexico (GOM) was used to quantify the bias and show the characteristics of production that determine its direction and magnitude. In this data set, exponentiation bias in estimates of remaining resources usually results in small absolute errors. Holding out varying fractions of the production histories of the completions analyzed, the interquartile range for errors in the estimated remaining resources (relative to unbiased estimates) extends from an underestimate of 886 to an overestimate of 2,105 BOE.

However, at the extreme ends of the distribution of errors, maximum underestimates of 8.3 million BOE and overestimates as large as 22.5 million BOE were found. More than 14% of the completions analyzed had forecast errors of more than 30%. Extreme biases are predictably associated with specific ranges and combinations of values of the three driver variables. Therefore, exponentiation bias can have very large and predictable effects on the economic value of estimated remaining resources, but they can be reliably corrected.

Introduction

Forecasting future production rates, remaining resources, and reserves in reservoirs and wells by use of production history has served as a basic petroleum-engineering tool for more than a century. Although there are several mathematical functions used to relate past to future production, the model of exponential decline is perhaps the most broadly applied. Unfortunately, the standard method to quantitatively evaluate the exponential model produces biased estimates of future production, the volume of remaining resources, and reserves derived from them.

Arps (1945) published the first rigorous and systematic exposition of decline curves in the middle of the 20th century. By use of the technology of the day, log-linear (i.e., semilog) graph paper, Arps (1945) plotted production on the logarithmic vertical axis and time on the linear horizontal axis. In hand-fitting a straight line through the data, its slope estimates the constant-decline rate of the exponential model. Eventually, graphical methods gave way to computer-automated statistical estimates of the parameters of the log-linear specification. In either event, estimates of the log-transformed intercept and slope parameters of the line are exponentiated. The area under the forecast curve from the last production observation to the point of abandonment is typically taken as an estimate of remaining producible resources.

The source of error identified here is not in the statistical estimation of the log-linear model of production vs. time. The bias occurs when the estimated parameters from the log-linear specification are reconverted by exponentiation back into linear space. We call this exponentiation bias. Finney (1941) originally identified this error and showed how to correct the associated bias. We focus first on the effect of the bias on estimates of future production and then discuss the implications for estimates of remaining resources and reserves.

Given the broad use of exponential models, recognition and correction of exponentiation bias are not as widespread in the scientific literature as might be expected. The papers that do exist range widely, including economics (Goldberger 1968), biogeochemistry (Middelburg et al. 1997), and limnology (Sobek et al. 2011). However, the problem is most often noted and rectified in the biological sciences (Sprugel 1983). Despite the centrality of exponential models in petroleum-engineering practice, there is no identification of the problem in the literature or evidence of correction in professional practice. To be fair, there are many other sources of error in decline-curve analysis that may overshadow exponentiation bias. Nevertheless, it is important for petroleum engineers to be aware of exponentiation bias and have tools to correct it. In this paper, we set the problem and solution in context by use of a data set from the GOM of 54,254 completion-level exponential-decline models (Earth Science Associates 2016).

Exponentiation Bias in Estimates of Future Production

Bias in Log-Linear Parameter Estimates. Arps (1945) proposed a deterministic model of which the decline in the production rate, q, at time, t, from a reservoir, well, or completion is proportional to the production rate. This is known as exponential decline and can be expressed as a differential equation,

$$\frac{1}{q}\frac{\mathrm{d}q}{\mathrm{d}t} = b. \qquad (1)$$

In Eq. 1, b is the decline rate and therefore $b < 0$. Solving Eq. 1 yields

$$y = a + bt, \qquad (2)$$

Original SPE manuscript received for review 8 August 2016. Revised manuscript received for review 21 November 2016. Paper (SPE 185943) peer approved 16 December 2016.

where $y = \ln(q)$ and a is a constant of integration. Typically, the production of a well is thought of as observations on y and t (time, measured in months on line). The regression model adopted here is

$$y_i = a + bt_i + \varepsilon_i. \quad (3)$$

If we assume that residuals, ε_i, are independently and identically distributed normally with zero mean and constant variance, σ^2, the standard-regression technique applied is ordinary least squares to estimate values for a ($\hat{a}$) and b ($\hat{b}$). This in turn provides an estimator for y called $\hat{y}$:

$$\hat{y} = \hat{a} + \hat{b}t. \quad (4)$$

Eq. 4 is an unbiased estimator for $y = \ln(q)$; however, an unbiased estimator of q is desired. It would seem that a reasonable estimator for $q = e^y$ is $\hat{Y} = e\hat{y}$:

$$\hat{Y} = \hat{A}e^{\hat{b}t}, \quad (5)$$

where $\hat{A} = e\hat{a}$. Eq. 5 is the usual equation used to model exponential decline of a completion, well, or reservoir. However, this model produces a biased estimate for future production. Because $y_i \approx N(a + bt_i, \sigma^2)$, then e^{y_i} is log-normally distributed, and it follows that

$$E(A^{y_i}) = e^{a + bt_i + \frac{1}{2}\sigma^2}. \quad (6)$$

Comparing Eq. 6 with Eq. 5 shows the multiplicative bias of $e^{-\frac{1}{2}\sigma^2}$. If a, b, and σ^2 are known with certainty, then Eq. 6 would be a suitable estimator for q. However, because a, b, and σ^2 are being estimated, to develop an unbiased estimator for production, a function, $f(\hat{\sigma}^2)$, must be found such that

$$E[e^{\hat{a} + \hat{b}t}f(\hat{\sigma}^2)] = e^{a + bt + \frac{1}{2}\sigma^2}. \quad (7)$$

This holds because $\hat{a}$, $\hat{b}$, and $\hat{\sigma}^2$ are complete sufficient statistics for a, b, and σ^2 (Lehmann and Casella 1998) and the distribution of $\hat{\sigma}^2$ is independent of the distribution of $\hat{a}$ and $\hat{b}$.

Our analysis focuses exclusively on identification and correction of exponential-decline-curve-estimation bias. We recognize, however, that the bias issues addressed here extend to generalized exponential- and hyperbolic-decline-curve models as well as to other even-more-general functional forms. We conjecture that logarithmic backtransformation of decline-curve-parameter estimates to estimate production per time period suffers from related forms of bias. Proof of this and computation of corrections are likely to be different, but related to the structure of proof and corrections for the classical exponential-decline-curve model presented here. Although perhaps important in both theory and engineering practice, a study of this conjecture exceeds the bounds of the present paper and awaits attention by others.

Correcting the Bias. Finney (1941) first recognized this bias in log-linear models and proposed a minimum-variance unbiased estimator (MVUE). Unfortunately, the initial solutions from Finney (1941) were written in terms of a function that converged slowly, which made it difficult to use in practice. Aitchison and Brown (1957) numerically approximated the Finney (1941) correction. Heien (1968) extended those ideas to log-linear regression, and Bradu and Mundlak (1970) further expanded them to general log-normal regression and also derived expressions for the variance of the MVUE. In his original work, Finney (1941) noted that his estimator could be written in terms of a Bessel function, but this did not prove useful until Seaborn (1991) rewrote it as a hypergeometric function. The hypergeometric function can be rapidly approximated numerically, making the bias correction easy to use in practice.

The MVUE for the expected value of production, $E(q)$, is the biased estimator (Eq. 5) multiplied by a correction factor, G_t, to correct for intercept-term bias. This estimator, as given by Shen and Zhu (2008), is

$$E(q) = \hat{q} = \hat{Y}(G_t) = (\hat{A}e^{\hat{b}t})G_t. \quad (8)$$

The correction factor is written in terms of the confluent-hypergeometric-limit function:

$${}_0F_1(\alpha; z) = \sum_{n=0}^{\infty} \frac{z^n}{n!(\alpha)_n}, \quad (9)$$

where $(\alpha)_n = \alpha(\alpha+1)(\alpha+2)...(\alpha+n-1)$ (Seaborn 1991),

$$G_t = {}_0F_1\left\{\frac{m}{2}; \frac{m[1-f(t)]}{4}\hat{\sigma}^2\right\}, \quad (10)$$

where $m = n - (p+1)$ is the number of degrees of freedom; p is the number of regressors; and n is the number of observations (here, $p = 1$, and therefore $m = n - 2$); where $f(t) = \frac{1}{n} + \frac{(t-\bar{t})^2}{\sum_{i=1}^{n}(t_i - \bar{t})^2}$, $\bar{t}$ is the midpoint (at $n/2$) of the observations on production from the month of peak production to the month of last production (n); and where $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}$ is an MVUE of the variance of y.

Further Exploring Exponentiation Bias. The value of the correction factor, G_t, determines if the bias of an uncorrected estimate of production (Eq. 5) is positive or negative. Because the value of G_t is maximized at $t = \bar{t}$ and is greater than unity, the bias at that point is always negative **(Fig. 1).** Wherever the bias is negative, the uncorrected regression underestimates future production.

From its maximum at $t = \bar{t}$, G_t declines symmetrically in both directions away from $\bar{t}$. It equals unity at two points in time equally spaced before and after $\bar{t}$, hereafter referred to as t_{C-} and t_{C+}, respectively. The crossing point preceding $\bar{t}$, t_{C-} is of no interest for prediction. However, the crossing point after $\bar{t}$, t_{C+} is important because $G_t = 1$ and at that point there is no bias. Therefore, from $\bar{t}$ to t_{C+}, exponentiation bias is always negative but decreasing in absolute magnitude until disappearing at t_{C+}. At $t > t_{C+}$, $G_t < 1$ and the bias becomes positive and increases in magnitude as G_t becomes progressively smaller with increasing t. We refer to t_{C+} as the bias-crossover point (from negative to positive).

There are two related reasons that the value of G_t reaches its maximum at the midpoint of the production time series ($\bar{t}$). First, in ordinary-least-squares regression, the variance of the estimator of the dependent variable [$\mathrm{Var}(\hat{y})$ in Eq. 4] is minimized at the midpoint of the independent variable's domain and symmetrically increases in both directions away from it. Second, for $G_t > 1$ to occur in general, the second argument of the confluent-hypergeometric-limit function must be positive. Expanding this term from Eq. 10, we obtain

$$\frac{m[1-f(t)]}{4}\hat{\sigma}^2 = \frac{m}{4}[\hat{\sigma}^2 - f(t)\hat{\sigma}^2] = \frac{m}{4}[\hat{\sigma}^2 - \mathrm{Var}(\hat{y})]. \quad (11)$$

Therefore, $\frac{m[1-f(t)]}{4}\hat{\sigma}^2$ is positive only when $\mathrm{Var}(\hat{y}) < \hat{\sigma}^2$. Because $\mathrm{Var}(\hat{y})$ at $t = \bar{t}$ is the minimal value for $\mathrm{Var}(\hat{y}_t)$ and by definition, $\mathrm{Var}(\hat{y})$ at $t = \bar{t}$ is $\hat{\sigma}^2/n$, then necessarily

$$\mathrm{Var}(\hat{y}_{t=\bar{t}}) = \frac{\hat{\sigma}^2}{n} < \hat{\sigma}^2, \quad (12)$$

and the second argument of the confluent-hypergeometric-limit function in Eq. 10 is maximized, as is G_t.

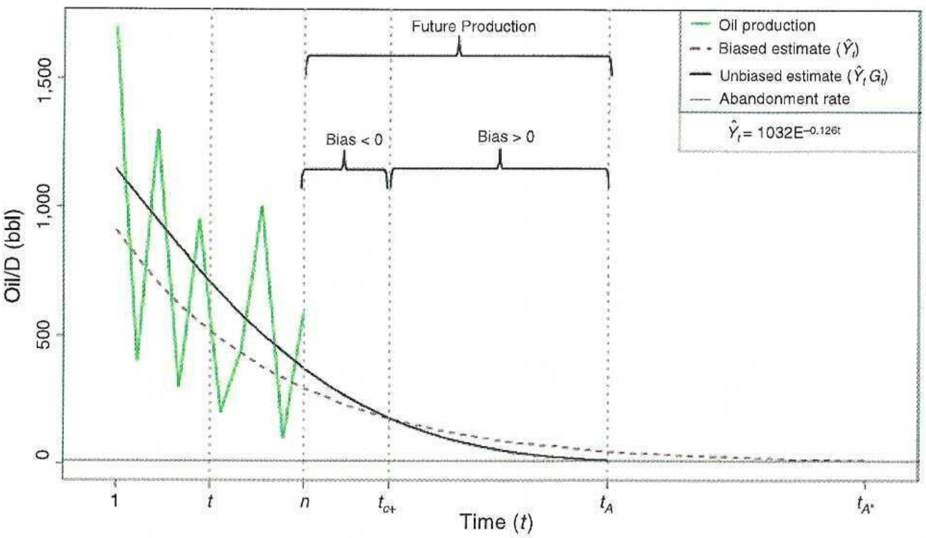

Fig. 1—The operation of exponentiation bias in a decline curve. Monthly oil production is shown from its peak production ($t = 1$) to its final observation ($t = n$). At the median observation, $\bar{t}$, the correction factor ($G\bar{t}$=1.37) is maximized and the absolute magnitude of negative bias reaches its maximum value and then declines to zero at t_{C+}. Between the final observation of production ($t = n$) and the time of abandonment (dependent on the unbiased estimator, t_A), estimates of future production by use of the usual estimator ($\hat{Y}_t$) are negatively biased between t_n and t_{C+} and positively biased between t_{C+} and t_A. The estimated time of abandonment, dependent on the biased estimator, is also shown at t_{A*}.

To solve for t_{C+}, the second parameter of the confluent-hypergeometric-limit function must equal zero. This occurs when

$$\hat{\sigma}^2 = \mathrm{Var}(\hat{y}) = \hat{\sigma}^2\left[\frac{1}{n} + \frac{(t_C - \bar{t})^2}{\sum_{t=1}^{N}(t-\bar{t})^2}\right] \Rightarrow t_{C+}$$

$$= \bar{t} + \sqrt{\frac{n-1}{n}\left[\sum_{i=1}^{n}(t_i - \bar{t})^2\right]}. \qquad (13)$$

Expanding the summation gives

$$\sum_{i=1}^{n}(t-\bar{t})^2 = \sum_{i=1}^{n} t_i^2 - 2\bar{t}\sum_{i=1}^{n} t_i + n\bar{t}^2. \qquad (14)$$

The solution of t_{C+} becomes simpler when the independent variable (time) is denominated as the number of months past peak production; the month of peak production is $t = 1$ and the independent variable becomes $t = 1, 2, 3, ..., n$. The right-hand side of Eq. 14 will further simplify, by use of the summation formulas, to

$$\frac{n(n+1)(n-1)}{12}. \qquad (15)$$

Substituting $\sum_{i=1}^{n} t_i^2 - 2\bar{t}\sum_{i=1}^{n} t_i + n\bar{t}^2$ from Eq. 15 into Eq. 13 yields

$$t_{C+} = \frac{n+1}{2} + \frac{n-1}{2}\sqrt{\frac{n+1}{3}}. \qquad (16)$$

Then t_{C+}, where bias crosses from negative to positive, is expressed entirely in terms of n. By use of this expression, it is obvious that t_{C+} always occurs on or after the last observation (t_n) because

$$t_{C+} = \frac{n+1}{2} + \frac{n-1}{2}\sqrt{\frac{n+1}{3}} \geq n, \qquad (17)$$

for all $n \geq 2$ with equivalence only when $n = 2$.

Exponentiation Bias in Estimated Remaining Resources

Estimating monthly production in the future is the fundamental purpose of decline-curve analysis. By use of forecasted future production, estimates are also made of remaining resources and of reserves. The former is the sum of estimated production from the month after the last empirical observation on production until the abandonment rate is reached. Reserves are an economic function of estimated remaining resources. Therefore, estimates of both remaining resources and reserves can inherit bias if they are built on biased estimates of future production. The bias in calculating remaining resources inherited from use of a decline curve generated from a biased exponential model (Eq. 5) will be referred to as net bias.

Exponential-decline-curve estimates of future production, without correcting for exponentiation bias, can be consistently negative or both negative and positive between the last observed production and abandonment. Therefore, net bias can be negative, positive, or zero depending on two factors.

The first factor is the maximum value of the bias-correction factor G_t at $t = \bar{t}$. The higher the variance and the larger G_t at $t = \bar{t}$, the more the line representing the biased estimate of future production is rotated counterclockwise through a fulcrum at t_{C+} relative to the fixed line representing the unbiased estimate of production (Fig. 1).

The second factor is where t_{C+} falls relative to the point of abandonment for the unbiased decline curve (t_A). There are two cases.

The first case is that if $t_{C+} \geq t_A$, then the bias for all months of future production is unambiguously negative, as well as estimates of remaining resources and reserves derived from them. Moreover, the biased estimate reaches the time of abandonment (t_{A*}) before the unbiased estimate (so $t_{A*} < t_A$) and the unbiased estimate forecasts a longer period of future production. Therefore, the absolute magnitude of the underestimation in remaining resources is the sum of the difference between the two estimators over the time period they are both forecasting production plus the sum of additional resources forecasted by the unbiased estimate during the extra time of production:

$$\sum_{t=(n+1)}^{t_{A*}}(|\hat{Y}_t G_t - \hat{Y}_t|) + \sum_{t=t_{A*}}^{t_A}(\hat{Y}_t G_t). \qquad (18)$$

The second case is that if $t_{C+} < t_A$, the sign of net bias depends on the difference between the magnitude of negative bias between t_{n+1} and t_{C+} and the magnitude of the positive bias between t_{C+} and t_A. Added to this is the quantity of additional resources to be produced in the forecast from $t = t_A$ to $t = t_{A*}$. As in the case of $t_{C+} < t_A$, the unbiased estimate reaches abandonment before the biased estimate, so $t_A < t_{A*}$. Within the two subperiods of negative and positive bias, the absolute magnitudes of the under/overestimates are functions of the production rate and the correction factor in each month. In the rare case where the magnitude of the negative bias (between t_{n+1} and t_{C+}) is equal to the sum of the positive bias (between t_{C+} and t_A) plus the volume of output produced between t_A to t_{A*}, the total net bias is zero.

These characteristics of exponentiation bias lead to several practical conclusions applicable to use of an exponential-decline model, estimated in semilog space, and then exponentiated to forecast future production, remaining producible oil and gas, and reserves.

1. The uncorrected estimates of future production from very-mature wells are most likely to be negatively biased (because large n postpones t_{C+} beyond t_A) and small in magnitude (because the physical volume of production is low).
2. Completions evaluated shortly after peak production have small n, so postponement of t_{C+} beyond t_n is relatively short and more likely to fall before t_A. It will, therefore, have a negative bias between t_n and t_{C+} and will have positive bias between t_{C+} and t_A. If the production rate is high, then the time between t_{C+} and t_A will be large and the biased estimator will overestimate the remaining resources. The magnitude of the overestimation will also be determined by $\hat{\sigma}^2$.
3. If $\hat{\sigma}^2$ is small, then the difference in remaining resources between the biased and unbiased estimators is likely to be insignificant.
4. If $t_{C+} \geq t_A$, then the larger the $\hat{\sigma}^2$ and the greater the underestimation of remaining resources. With other conditions remaining the same, the lower the $\hat{\sigma}^2$, then the smaller the absolute magnitude of bias will be, but it is still negative.

Figs. 2a through 2d show examples of completion-level production histories from the GOM. In Figs. 2a and 2b, the values of $\hat{\sigma}^2$ are high, making $G_{t=\bar{t}}$ high (both also illustrate the decline in variance as a function of post-peak time on line). However, because both also have long production histories, the bias-crossover points are postponed beyond t_A. Therefore, the bias is unambiguously negative over the forecast period but relatively small in volume because the final observations on production are so close to the abandonment rates for both completions.

Fig. 2c shows a much-lower-variance post-peak-production history, therefore $G_{t=\bar{t}}$ is nearly unity and biased and unbiased estimates of production beyond the last observation are approximately equal. In Fig. 2d, the situation is even more extreme in the positive direction. The production history is so short that the bias-crossover point occurs slightly after the final observation on production and the bias remains positive until abandonment, making the uncorrected estimate a small overestimate of unbiased estimate of remaining resources.

Confidence intervals (CIs) can also be constructed for the unbiased estimator by applying the formula for variance (Eq. 19) as given by Shen and Zhu (2008):

$$\mathrm{Var}[\hat{q}(t)] = [q(t)]^2 \left(e^{f(t)\sigma^2} {}_0F_1 \left\{ \frac{m}{2}; \frac{m\,[1-f(t)]}{4} \sigma^4 \right\} - 1 \right). \qquad (19)$$

Substituting $\hat{q}(t)$ for $q(t)$ and $\hat{\sigma}$ for σ yields an estimate of the variance of the unbiased estimator. Assuming normality and following standard techniques lead to a CI for $\hat{q}(t)$ by use of

$$CI = \hat{q}(t) \pm z\sqrt{\mathrm{Var}[\hat{q}(t)]}, \qquad (20)$$

with z chosen to correspond to confidence-level choice. Figs. 2e through 2h show the same forecasts as in Figs. 2a through 2d, with CIs added.

Results for the GOM

To quantify the effect of exponentiation bias on a very large data set, we used 54,254 completion-level monthly production histories from the GOM. Of the 54,254 completions, 42,707 had production histories suitable for decline-curve analysis on the oil production and 46,971 had suitable gas-production histories. For each completion, corrected- and uncorrected-exponential-decline curves were applied to post-peak-production histories to estimate remaining resources (see Appendix A for the method used to determine peak production). Abandonment rates of 10 BOPD and 50 Mcf/D of gas were used. Subsequently, the post-peak-production history of each completion was re-evaluated by use of the first 10% of the production history. This was repeated, by 10% increments, up to 90% of the production history. These holdouts reflect the position an engineer would face in estimating future production from the early through late stages of production.

Figs. 3a through 3i show the distribution of errors, defined by subtracting the unbiased from the biased estimate of remaining oil and gas for each completion. Because of the similarity of results between oil and gas, errors are expressed in BOE, with gas converted at 5.62 Mcf of gas per BOE. The errors range, over all holdout cohorts, from a maximum underestimate of 8.3 million BOE to maximum overestimate of 22.5 million BOE (**Table 1**). However, most errors are much smaller; the interquartile range of errors over all holdouts was from an underestimate of 886 to an overestimate of 2,105 BOE.

That the distribution of error systematically changes over the maturity of a completion is clear from Figs. 3a through 3i. Most importantly, the number of completions with negatively biased errors increases with the maturity of production. Completions very early in their production history, when only 10–20% of the post-peak-production history is complete (Figs. 3a and 3b), show the highest frequency of positively biased estimates (68% of the estimates for the 10% cohort and 60% of the 20% cohort). However, by the midpoint of the production history (Fig. 3e), positively biased estimates drop to 30% of the total, and by the 90% cohort, only 7% are positive. The share of errors reflecting negatively biased estimates correspondingly increase as a function of maturity to dominate the distributions of mature-production histories.

Dividing each of the full distributions shown in Figs. 3a through 3i at zero into distributions of negative and positive errors provides further insights (Table 1). Although the share of positive vs. negative errors shifts, the absolute value of the means and medians are both relatively stable functions of maturity. The medians and means of absolute negative errors increase by a factor of approximately 3–4 between the 10 and 90% cohorts; the same statistics for positive errors decrease by approximately the same factors over the same cohorts (Table 1). Because of error in both directions (the mean ranges from –6,058 to 50,960 BOE and the median from –1,146 to 3,741 BOE over maturity cohorts) in most instances, exponentiation bias does not create either large overestimation or underestimation of remaining resources and, therefore, reserves.

The fact that, on average, the volumetric effect of exponentiation bias is small relative to the ultimate recovery does not mean that it can be ignored. In the extremes of the distributions, the absolute magnitude of errors for remaining resource runs to millions of BOE. Because most errors are negative, reserves estimates inherit underestimation and the magnitudes of error in the extreme quantiles of the distribution translate into millions to hundreds of millions of dollars of understatement of economic value. The extreme positive errors, leading to overestimation, are an order larger and occur more frequently in the holdouts that use fewer data (Table 1).

The extreme positive and negative errors in estimates of remaining resources are systematically associated with variance ($\hat{\sigma}^2$). The top 0.1% of the negative errors over the entire distribution of errors over all the holdouts has a median $\hat{\sigma}^2 = 0.60$ and mean $\hat{\sigma}^2 = 0.64$. The top 0.1% of the positive errors has a median $\hat{\sigma}^2 = 0.067$ and mean $\hat{\sigma}^2 = 0.095$ These compare with a median $\hat{\sigma}^2 = 0.22$ and mean $\hat{\sigma}^2 = 0.29$ for the entire sample. Therefore, the differences between the central tendencies of variances in the

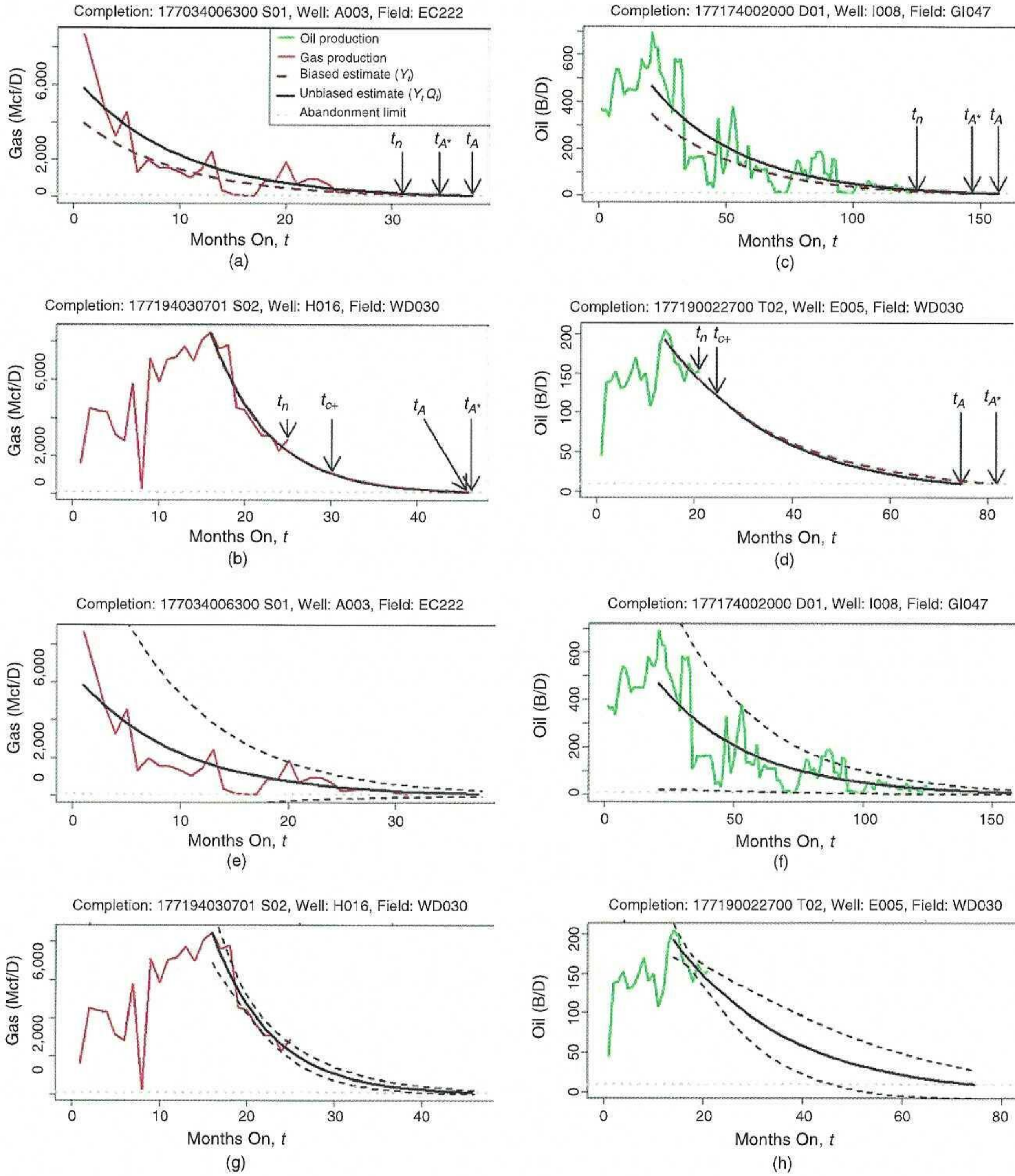

Fig. 2—(a through d) Four completion-level decline curves from the GOM. (a) and (b) Cases of exponentiation-bias underestimation of future production. (c) Case of nearly zero bias, because variance is very low. In (d), the combination of low variance and very-short production history leads to positive bias. (e through h) These are the same as (a through d), but with 95% CIs (shown in black dashed lines). (e, f) CIs are wide; those for (e) extend below zero. (g, h) There are much-narrower intervals caused by low variability in the production data. Completions are identified by the well API number and the completion interval number. Field names are abbreviated: EC = East Cameron, GI = Grand Isle, and WD = West Delta.

0.1% quantiles differ by a factor of approximately three. The difference in the degrees of freedom between the extreme positive and negative errors is also significant. The top 0.1% of negative errors has a mean degree of freedom of 61, whereas the top 0.1% of positive errors has a mean of 10. The mean degree of freedom for the entire sample is 17.

The largest negative and positive errors in estimating remaining resources occur when outlying values of $\hat{\sigma}^2$ and n are combined with high production rates between the last production and abandonment because the absolute magnitude of the bias is the product of the correction factor (G_t) and the uncorrected estimated-future-production rate ($\hat{Y}_t$), as shown in Eq. 8.

Finally, our analysis of the absolute magnitudes of bias and the frequency with which high error estimates occur are conservative because of two methodological decisions in the analysis of the GOM data. First, we used months on line as the measure of time, which eliminates months of zero production. Use of calendar months would necessarily increase the variance of records of

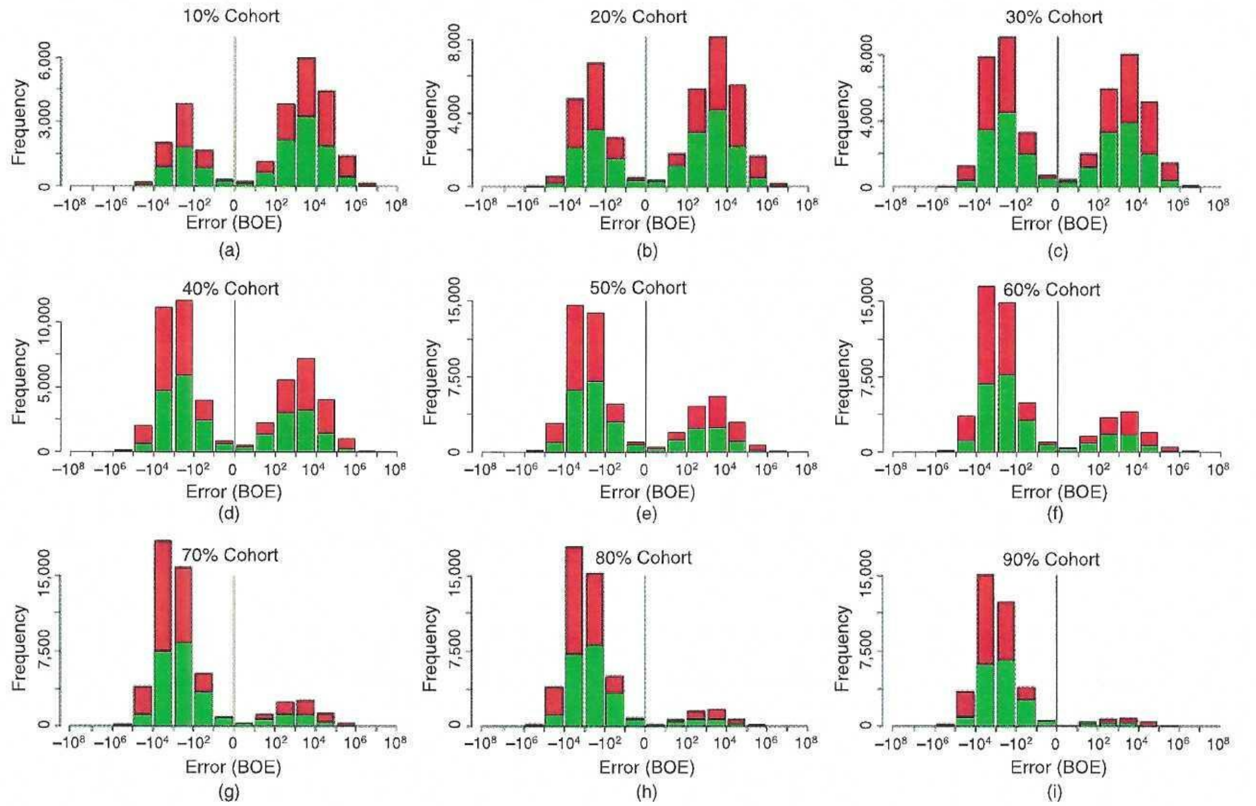

Fig. 3—(a through i) The distribution of errors from exponentiation bias in estimating the remaining resources of 54,254 completion-level production histories from the GOM (green is oil; red is gas). The numbers above each graph show the percentage of the production history used in the analysis of each cohort. The error on the *x*-axis is the biased estimate of remaining resources minus the unbiased estimates in BOE.

post-peak-production histories and therefore the sizes of biases in estimates of future production. Second, because we identified peak production as the maximum output after the final break in regime of the production history (Appendix A), the variance used in the analysis was lower than that obtained if peak production was defined as the global maximum rate of production.

Conclusions

Fitting an exponential-decline-curve model to post-peak-production history, by use of the usual method of semilog transformation, produces biased estimates of future production. The cost of failure to correct this bias is that volumes of oil and gas are wrongly added to or removed from the remaining volumes estimated to be technically and economically recoverable.

Empirical results for the sample of 54,254 completion-level production histories from the GOM show that the net direction of the bias is typically negative, creating an underestimation of future production, remaining resources, and reserves. The absolute magnitude of errors in both directions is usually very small relative to ultimate recovery. However, in rare cases, it can cause errors of hundreds of thousands to millions of BOE of estimated future production.

Although the magnitude of the error is usually low, its correction by use of the confluent-hypergeometric-limit function is straightforward and computationally inexpensive to apply. Therefore, there is no reason not to correct the bias, even when the error is small. There are very compelling economic reasons to correct the bias under those empirical conditions that lead to extreme magnitudes of error.

There are many technical and economic factors that may affect the production rate of a completion beyond the performance of the reservoir. These include mechanical problems from the completion to the field-level processing facilities and weather and

	Negative Errors (Underestimate)			Positive Errors (Overestimate)		
Holdout Cohort	Absolute Median	Absolute Mean	Absolute Maximum	Median	Mean	Maximum
10%	341	1,719	338,600	3,741	50,960	22,480,000
20%	497	2,412	1,129,000	3,130	44,960	12,910,000
30%	664	3,732	3,845,000	2,432	36,190	22,340,000
40%	770	4,435	5,455,000	1,891	26,430	5,862,000
50%	870	4,708	5,483,000	1,543	26,340	10,080,000
60%	967	4,990	4,869,000	1,323	21,930	4,302,000
70%	1,019	5,202	8,332,000	1,109	17,760	3,636,000
80%	1,055	5,433	4,588,000	1,095	19,080	3,514,000
90%	1,146	6,058	4,034,000	1,160	19,240	2,884,000

Table 1—Characteristics of errors from exponentiation bias in the GOM (shown in BOE).

transportation disruptions; there are also economic considerations that may lead to reduction or shut in of production. The effects of these exogenous sources of variance may swamp errors in production forecasts arising directly from exponentiation bias, but they do not eliminate or offset errors caused by this bias.

In contrast, in other branches of science where this problem has been identified and corrected, the exponential function has usually been used for modeling interpolation of values within the bounds of existing data. In that case, the bias is always negative, because t_{C+} always falls after the end of the data series and its magnitude is larger because G_t is larger the closer t is to $\bar{t}$. For decline curves, the use of the exponential function is for extrapolation to estimate values beyond the data series. Therefore, net exponentiation bias can be positive, negative, or zero, and the absolute magnitude is likely to be smaller because the correction factor is multiplied by low production at the end of the completion's producing life. In addition, positive and negative components offset in the calculation of net bias.

Nevertheless, given the ease of correction and the extremely high economic consequences of (admittedly low-probability) high-magnitude errors, failure to correct exponentiation bias is unjustified as a matter of standard petroleum-engineering practice. Given the multitude of factors that degrade the accuracy of decline-curve analysis, its overall accuracy can be unambiguously improved by recognizing and treating this source of error that arises for purely a mathematical reason.

Nomenclature

a, b = two parameters of linear-equation model
$\hat{a}, \hat{b}$ = estimates for a and b
$\hat{A}$ = $e\hat{a}$
e = base of natural logarithms
E = expected value operator
f = function operator
${}_0F_1$ = confluent hypergeometric-limit-function operator
G_t = hypergeometric-limit-function correction factor at time t
m = number of degrees of freedom
n = number of observations between peak and final observed production
N = normal distribution operator
p = number of regressors
q = production per unit of time
$\hat{q}$ = estimated production per unit of time by use of the corrected estimator
t = unit of time
$\bar{t}$ = midpoint of production-time series between peak and last observed production
t_A = time of abandonment for the corrected estimator
t_{A*} = time of abandonment for the uncorrected estimator
t_{C-} = crossover point (zero bias) before $\bar{t}$
t_{C+} = crossover point (zero bias) after $\bar{t}$
Var = variance operator
y = natural logarithm of the production rate
$\hat{y}$ = estimator of the natural logarithm of the production rate
$\hat{Y}$ = $e\hat{y}$
z = the second argument of the confluent-hypergeometric-limit function
α = the first argument of the confluent-hypergeometric-limit function
ε_i = stochastic error term of linear regression
σ^2 = variance of y

Acknowledgments

The authors wish to express their appreciation to M. Watson, M. Wigwe, and E. Attansi for their very helpful suggestions for improving this paper, and to M. Spater for editorial assistance. We also thank the editors and reviewers of this journal for their helpful comments and suggestions, which resulted in substantial improvements in this paper.

References

Aitchison, J. and Brown, J. A. C. 1957. *The Lognormal Distribution With Special Reference to Its Use in Economics*. Cambridge, UK: Cambridge University Press.

Arps, J. J. 1945. Analysis of Decline Curves. In *Transactions of the Society of Petroleum Engineers*, Vol. 160, Part 1, 228–247. SPE-945228-G. Richardson, Texas: SPE. http://dx.doi.org/10.2118/945228-G.

Brown, R. L., Durbin, J., and Evans, J. M. 1975. Techniques for Testing the Consistency of Regression Relationships over Time. *J. Roy. Stat. Soc. B. Met.* **37** (2): 149–192.

Bradu, D. and Mundlak, Y. 1970. Estimation in Lognormal Linear Models. *J. Am. Stat. Assoc.* **65** (329): 198–211. http://dx.doi.org/10.1080/01621459.1970.10481074.

Earth Science Associates. 2016. GOM^3, Version 7. Long Beach, California: Earth Science Associates.

Finney, D. J. 1941. On the Distribution of a Variate Whose Logarithm is Normally Distributed. *Suppl. J. Roy. Stat. Soc.* **7** (2): 155–161. http://dx.doi.org/10.2307/2983663.

Goldberger, A. S. 1968. The Interpretation and Estimation of Cobb-Douglas Functions. *Econometrica* **35** (3–4): 464–472. http://dx.doi.org/10.2307/1909517.

Heien, D. M. 1968. A Note on Log-Linear Regression. *J. Am. Stat. Assoc.* **63** (323): 1034–1038. http://dx.doi.org/10.2307/2283895

Lehmann, E. L. and Casella, G. 1998. *Theory of Point Estimation*, second edition. New York City: Springer-Verlag.

Middelburg, J. J, Soetaert, K., and Herman, P. M. J. 1997. Empirical Relationships for Use in Global Diagenetic Models. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **44** (2): 327–344. http://dx.doi.org/10.1016/S0967-0637(96)00101-X.

Seaborn, J. B. 1991. *Hypergeometric Functions and Their Applications*. New York City: Springer-Verlag.

Shen, H. and Zhu, Z. 2008. Efficient Mean Estimation in Log-Normal Linear Models. *J. Stat. Plan. Infer.* **138** (3): 552–567. http://dx.doi.org/10.1016/j.jspi.2006.10.016.

Sobek, S., Nisell, J., and Fölster, J. 2011. Predicting the Depth and Volume of Lakes from Map- Derived Parameters. *Inland Waters* **1** (3): 177–184.

Sprugel, D. G. 1983. Correcting for Bias in Log-Transformed Allometric Equations. *Ecology* **64** (1) 209–210. http://dx.doi.org/10.2307/1937343.

Zeileis, A. 2006. Implementing a Class of Structural Change Tests: An Econometric Computing Approach. *Comput. Stat. Data An.* **50** (11): 2987–3008. http://dx.doi.org/10.1016/j.csda.2005.07.001.

Zeileis, A., Lesich, F., Hornik, K. et al. 2002. strucchange: An R Package for Testing for Structural Change in Linear Regression Models. *J. Stat. Soft.* **7** (2): 1–38.

Appendix A—Determining Peak Production

In the ideal presentation of decline-curve analysis, peak production is taken as the global maximum rate of production from the completion, well, or reservoir under study. However, many things can occur between the global maximum and the last observation on production. Shutdowns for repairs, weather, and transportation problems are among the frequently encountered reasons to stop or severely reduce production that are not strictly related to performance of the well and/or reservoir. In all these cases, reducing or stopping production adds to the variance of the production, which affects exponentiation bias. To fit the decline curve over the part of the production history most relevant to forecasting future performance, engineers often change the start of the analysis to a point beyond the global maximum, thus eliminating irrelevant variations in production.

In this study, decline curves were estimated in batches of tens of thousands, so detailed examination of individual production histories and adjustments of starting points was infeasible. To assure that decline-regression techniques are properly applied to each production history, a cumulative sum (CUSUM) test designed to identify structural breaks in a regression regime was applied. This test was developed in the 1950s for production-process-quality control (Brown et al. 1975).

Here, the (exponential) rate of decline of a completion's production from a global maximum is assumed constant. Specifically, the efp function in the strucchange library of R was used to find and test the significance of changes in all production histories

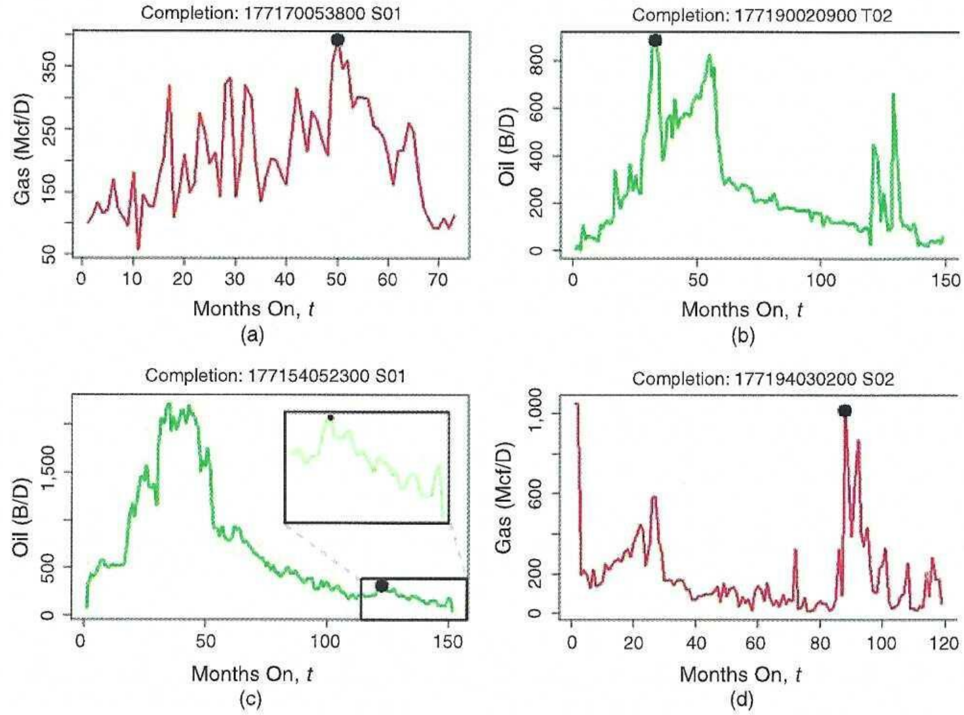

Fig. A-1—(a through d) Black dots show where the start of the final decline regime identified by CUSUM testing begins. Oil is shown in green and gas in red. (a, b) These are examples in which no significant break in regime is found, so the global maximum was used as the start of the decline regime. (c, d) These are examples of a break in regime beyond the global maximum.

(Zeileis et al. 2002; Zeileis 2006). Once all significant structural breaks in a completion's production history are identified, the maximum production after the final break in the production history is taken as the maximum production rate from which the decline curve is estimated.

Of the 54,254 completion-level production histories examined here, 27,227 (or 50.2% oil-production histories) and 25,804 (or 47.6% gas-production histories) contained significant breaks in regime. In **Figs. A-1a through A-1d,** examples of four production histories are shown. Two of them contain no statistically significant breaks in regime (Figs. A-1a and A-1b) and two do (Figs. A-1c and A-1d). In the latter two cases, the maximum production rate used in the decline-curve analysis was shifted from the global maximum production to the maximum associated with the point of the structural break (the black dots).

The decision to limit the decline-curve analysis to time after the final structural break and the decision to use months on line to measure time minimized the empirical variance relative to use of a global production maximum and calendar time as the independent variable. Both decisions supported focusing on the effect of exponentiation bias, but they also made assessment of its empirical effect conservative.

SI Metric Conversion Factors

bbl × 1.589873	$E-01 = m^3$
ft^3 × 2.831685	$E-02 = m^3$

Scott Morris joined Earth Science Associates as a research associate in 2012. He holds a bachelor's degree in statistics from the University of California, Riverside, and a master's degree in applied mathematics from California State University, Fullerton.

Gordon M. Kaufman is the Morris A. Adelman Professor of Management Emeritus at the Massachusetts Institute of Technology's Sloan School of Management. He is a fellow of the American Association for the Advancement of Science, the American Statistical Association (ASA), and the International Society for Bayesian Analysis, and was the 2015 International Association for Mathematical Geosciences Distinguished Lecturer. During the past 55 years, Kaufman has consulted for many oil and gas companies as well as for several branches of the Department of the Interior. He has authored or coauthored 60 articles and three books. Kaufman is a past chair of the ASA Advisory Committee to the US Energy Information Administration.

John D. Grace is the president of Earth Science Associates, which produces analytic software for the oil and gas industry. His research has focused on basin analysis/hydrocarbon-resource assessment, energy economics, and spatial and multivariate statistical analysis. Grace has served as an adviser to companies, government, and international agencies and has previously taught geology and mathematics at Louisiana State University, the University of Southern California, and California State University, Fullerton. He has authored or coauthored dozens of articles and one book, and he holds two patents. Grace is a member of SPE and has published in the *Journal of Petroleum Technology*. He holds a PhD degree from Louisiana State University.