WHEN UNCERTAINTY IS BENEFICIAL: INTERESTING IMPLICATIONS FOR THE HYDROCARBON DISCOVERY PROCESS

YIGAL GERCHAK

Department of Industrial Engineering, Tel Aviv University, Tel Aviv, 69978, Israel, ygerchak@eng.tau.ac.il

QI-MING HE

Department of Industrial Engineering, DalTech, Halifax, Nova Scotia, Canada, qi-ming.he@dal.ca

J. DAVID FULLER

Department of Management Sciences, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1, dfuller@engmail.uwaterloo.ca

EKACHIDD CHUNGCHAROEN

Department of Industrial and Operations Management, Thammasat University, Bangkok, Thailand, echungch@alpha.tu.ac.th

(Received October 1997; revisions received June 1999, February 2000, April 2001; accepted May 2001)

Consider a basin with remaining undiscovered fields, in which both field sizes and their number are uncertain. Assuming that the probability of finding a field is increasing in its size, we show that the expected size of the *first* remaining field discovered, which is of particular importance, is *increasing* in the variability of field sizes, and results from a simulation model of exploration illustrate this trend for the first several fields discovered. We also provide simple bounds on the expected size of the first field discovered. It seems that more information (less variability) leads to less value—something unexpected. We explore and explain this seemingly counterintuitive result.

Suppose that the probability of selecting an item from a set is increasing in the item's size. A case in point is the hydrocarbon discovery process, where it has been observed that the probability of discovering a particular field within a basin is increasing in the field's size (e.g., Arps and Roberts 1958, Barouch and Kaufman 1976, Power and Fuller 1991). This is akin to other common observations, like finding a grownup's socks in the wash before the baby's or quarters in your pockets before the pennies. How would then the expected size of an item selected (first) of those remaining be influenced by the variability of sizes? We prove that even for a population with a fixed mean the expected size of the item selected is *increasing* in the variability. Thus the more uncertain the sizes of remaining fields, the larger the expected size of the field found first. This observation has a seemingly counterintuitive implication for hydrocarbon exploration, since it seems to imply that improved geological knowledge of a certain type, i.e. which reduces field size uncertainty and leaves mean field size unchanged, will reduce the expected size of the first field found. Loosely speaking, it seems that *more information leads to less value*—something rather unexpected if we are accustomed to prefer less variability, e.g., in financial decisions. In this paper, we explore and explain this seemingly counterintuitive result.

We also derive simple upper and lower bounds on the expected size of the next field found.

The expected size of the first, or first few, remaining fields discovered in a basin is particularly important because although all, or most, of the economically viable fields will eventually be discovered with enough exploratory effort, discounting of potential values of discoveries makes the early ones of greatest interest. Faced with a choice between drilling expensive exploratory wells in one basin or another, a company's decision may depend on the expected sizes of the first few discoveries in the basins. The actual size(s) of the first field(s) discovered may not follow this pattern but the observation that their expected size(s) do is informative for planning purposes.

THE MODEL

Let $X_1, X_2, \ldots$ be identically distributed nonnegative random variables, and let N be a positive integer-valued random variable, independent of the X_i's. In the context of

Subject classifications: Industries, petroleum/natural gas: exploration. Natural resources, energy: discovery process. Probability models.
Area of review: ENVIRONMENT, ENERGY, AND NATURAL RESOURCES.

0030-364X/02/5006-1068 $05.00
1526-5463 electronic ISSN

the hydrocarbon discovery process the X_i's are the field sizes, and N the random number of remaining fields, in an unexplored basin. Note that we do not require that the field sizes be statistically independent. In the geological literature on statistical analysis of hydrocarbon accumulations, the notion of field size distribution is widely used (e.g., Arps and Roberts 1958, O'Carroll and Smith 1980, Smith and Ward 1981, Schuenemeyer and Drew 1983, Lee and Wang 1985, and Power 1992). For example, a log-normal distribution is frequently employed to represent the probability that any given undiscovered field in a basin is less than or equal to a certain size. The field size distribution summarizes geological knowledge about potential field sizes in a basin, but it does not represent the number of undiscovered fields that might exist. The number of undiscovered fields is itself uncertain, in the sense of being unknown to us, and may be represented by a random variable.

For each realization $x_1, \ldots, x_n$ of the above variables, we sample one of these realized values, in the hydrocarbon discovery process that corresponds to the first field discovered. It has been documented in geological literature (e.g., Cozzolino 1972, O'Carroll and Smith 1980, Smith and Ward 1981, Lee and Wang 1983a,b, Attanasi and Drew 1985, Drew et al. 1988, Power 1992) that the probability of choosing a particular x_i tends to be proportional to x_i^β, $\beta > 0$, and thus increasing in x_i. Thus, for a given realization $x_1, \ldots, x_n$, the expected value of the variable sampled is

$$\sum_{i=1}^{n} x_i \frac{x_i^\beta}{\sum_{j=1}^{n} x_j^\beta} = \frac{\sum_{i=1}^{n} x_i^{\beta+1}}{\sum_{j=1}^{n} x_j^\beta}. \qquad (1)$$

Thus prior to the realization, the expected value (size) of the variable (field) sampled is

$$\begin{aligned} &E_{N,X_1,X_2,\ldots}\left(\frac{\sum_{i=1}^{N} X_i^{\beta+1}}{\sum_{j=1}^{N} X_j^\beta}\right) \\ &\quad = E_N\left\{E_{X_1,X_2,\ldots}\left[\frac{\sum_{i=1}^{N} X_i^{\beta+1}}{\sum_{j=1}^{N} X_j^\beta}\middle| N\right]\right\} \\ &\quad = E_N\left\{N E_{X_1,X_2,\ldots}\left(\frac{X_1^{\beta+1}}{\sum_{j=1}^{N} X_j^\beta}\middle| N\right)\right\}, \qquad (2) \end{aligned}$$

where the first equality is due to the independence of the X_i's of N and the second equality follows from the fact that the X_j's are identically distributed. (We selected X_1 arbitrarily as a random field size.) Note that despite the multiplication by N on the RHS of (2), the quantity corresponds to the expected size of a *single* field, as it evolved from the LHS expression.

The distribution of the random number of remaining fields N will remain fixed throughout. The common distribution of the X_i's, denoted by F, will be varied. We are interested in how the quantities

$$E\left(\frac{X_1^{\beta+1}}{\sum_{j=1}^{n} X_j^\beta}\right) \equiv E_F(n)$$

and

$$E\left\{N E\left(\frac{X_1^{\beta+1}}{\sum_{j=1}^{N} X_j^\beta}\middle| N\right)\right\} \equiv E_F$$

are influenced by the variability of X_i, i.e. by making the distribution F more (or less) variable. That corresponds to the implications for first discovered field size of the uncertainty in field sizes. Note that while E_F is the quantity in which we are ultimately interested (expected size of first field discovered), $nE_F(n)$ would be the expected size if the number of fields was known to equal n.

While our work was motivated by hydrocarbon discovery issues, the expected size of the first "target" found could also be relevant in other mining applications, as well as search and detection problems.

UNCERTAINTY VS. CERTAINTY

First, we shall compare a situation with (arbitrarily) uncertain field size to that of known field size. The known size will be assumed to be equal to the expected size of the uncertain case. While the result here is implied by the more general ones given later, we chose to present this case first since it is easier to relate to.

PROPOSITION 1. $E_F(n) \geqslant E(X_1)/n$ *for any* $\beta \geqslant 0$.

PROOF. For any $i = 1, \ldots, n$,

$$\begin{aligned} E\left(\frac{X_1 X_i^\beta}{\sum_{j=1}^{n} X_j^\beta}\right) &= E\left\{\frac{X_1}{(X_1^\beta + \cdots + X_n^\beta)^{\frac{1}{\beta+1}}} \cdot \frac{X_i^\beta}{(X_1^\beta + \cdots + X_n^\beta)^{\frac{\beta}{\beta+1}}}\right\} \\ &\leqslant \left\{E\frac{X_1^{\beta+1}}{(X_1^\beta + \cdots + X_n^\beta)^{\frac{\beta+1}{\beta+1}}}\right\}^{\frac{1}{\beta+1}} \\ &\quad \cdot \left\{E\frac{X_i^{\beta\cdot\frac{\beta+1}{\beta}}}{(X_1^\beta + \cdots + X_n^\beta)^{\frac{\beta}{\beta+1}\cdot\frac{\beta+1}{\beta}}}\right\}^{\frac{\beta}{\beta+1}} \\ &= E\left(\frac{X_1^{\beta+1}}{X_1^\beta + \cdots + X_n^\beta}\right), \end{aligned}$$

where the inequality above is an application of Hölder's Inequality (e.g., Rudin 1976, p. 139), and the last equality follows from the fact that $X_1, \ldots, X_n$ are identically distributed.

Thus

$$\begin{aligned} E_F(n) &= E\left(\frac{X_1^{\beta+1}}{X_1^\beta + \cdots + X_n^\beta}\right) \\ &\geqslant \frac{1}{n}\left\{E\left(\frac{X_1 X_1^\beta}{X_1^\beta + \cdots + X_n^\beta}\right) + \cdots + E\left(\frac{X_1 X_n^\beta}{X_1^\beta + \cdots + X_n^\beta}\right)\right\} \\ &= \frac{1}{n}E(X_1). \qquad \square \end{aligned}$$

It follows from above that

$$E_F = E_N\left(N E\frac{X_1^{\beta+1}}{\sum_{j=1}^{N} X_j^\beta}\right) \geqslant E_N\left\{N \cdot \frac{E(X_1)}{N}\right\} = E(X_1). \qquad (3)$$

If all fields had been of the same known size $E(X_1)$, the expected size of a field sampled would, of course, be $E(X_1)$. Thus inequality (3) asserts that this certain case has the *lowest* expected size sampled. Uncertainty in field size increases (or at least does not decrease) the expected size of the field found first.

Note that the above result does *not* impose any restrictions on β beyond nonnegativity; some later results do.

BOUNDS

The above discussion provided a lower bound on the expected size sampled. One can also provide an upper bound:

PROPOSITION 2. $E_F \leqslant E\{\max(X_1, \dots, X_N)\}$.

PROOF. By (1)

$$E_F = E\left(\sum_{i=1}^{N} X_i \frac{X_i^\beta}{\sum_{j=1}^N X_j^\beta}\right).$$

But

$$\sum_{i=1}^{N} X_i \frac{X_i^\beta}{\sum_{j=1}^N X_j^\beta} \leqslant \max(X_1, \dots, X_N) \sum_{i=1}^{N} \frac{X_i^\beta}{\sum_{j=1}^N X_j^\beta} = \max(X_1, \dots, X_N).$$

Thus $E_F \leqslant E\{\max(X_1, \dots, X_N)\}$. □

As for computing the expectation of the maximum of random variables, the task is complex unless the random variables (field sizes) can be assumed to be independent. In that case

$$E\{\max(X_1, \dots, X_n)\} = \int_0^\infty \{1 - [F(x)]^n\}\, dx.$$

Approximations of this expectation are available for many distributions (e.g., Arnold et al. 1994). In particular, if n is large, then for normal and exponential distributions the expectation grows logarithmically in n. For the exponential distribution with mean $1/\lambda$, we have $E\{\max(X_1, \dots, X_n)\} = (\sum_{i=1}^n 1/i)/\lambda \approx (\log n)/\lambda$, and the convergence is thus quite fast. For the normal, the log approximation requires a sample to be of about 100. There are no such simple approximations for lognormal and Pareto variables, but some recursive relations have been derived (Arnold et al. 1992, Johnson et al. 1994).

After $E\{\max(X_1, \dots, X_n)\}$ has been calculated/approximated, as a function of n, we can take its expectation with respect to N.

INCREASED UNCERTAINTY: CONVEX ORDER

We now wish to examine the implication for E_F of increasing the variability of F. We consider the convex order $F \leqslant_{cx} G$. Let $X \sim F$ and $Y \sim G$ be two random variables such that $E[\phi(X)] \leqslant E[\phi(Y)]$ for all convex functions ϕ. Then X is said to be smaller than Y in the convex order, denoted as $X \leqslant_{cx} Y$ or as $F \leqslant_{cx} G$. The random variable Y is then "more variable" than the random variable X (e.g., Shaked and Shanthikumar 1994). Note that random variables which are convex ordered may have the same mean.

If the X_j's are independent (which we shall now assume),

$$E_F(n) = \int\!\!\int_{\dots} \frac{x_1^{\beta+1}}{x_1^\beta + \cdots + x_n^\beta}\, dF(x_1) \dots dF(x_n).$$

To prove that $E_F(n) \leqslant E_G(n)$ when $F \leqslant_{cx} G$, it is thus sufficient to prove that the function

$$\frac{x_1^{\beta+1}}{x_1^\beta + \cdots + x_n^\beta} \equiv g(x_1, \dots, x_n)$$

is convex in $x_1, \dots, x_n$ (individually). We prove this for $\beta \leqslant 1$. The range $0 \leqslant \beta \leqslant 1$ is common in many studies, e.g., Chungcharoen (1997) uses $\beta = 0.82$ in a study of exploration for natural gas off Canada's east coast.

LEMMA 1. *If* $0 \leqslant \beta \leqslant 1$ *then* $g(x_1, \dots, x_n)$ *is convex in* $x_1, \dots, x_n$ *(individually).*

PROOF. Let $\Delta = x_1^\beta + \cdots + x_n^\beta$. Then

$$\begin{aligned}\frac{\partial g}{\partial x_1} &= \frac{1}{\Delta^2}\left[(\beta+1)x_1^\beta \Delta - x_1^{\beta+1} \beta x_1^{\beta-1}\right]\\ &= \frac{1}{\Delta^2}\left[x_1^{2\beta} + (\beta+1)x_1^\beta x_2^\beta + \cdots + (\beta+1)x_1^\beta x_n^\beta\right],\\ \frac{\partial^2 g}{\partial x_1^2} &= \frac{1}{\Delta^2}\left[2\beta x_1^{2\beta-1} + (\beta+1)\beta x_1^{\beta-1}(x_2^\beta + \cdots + x_n^\beta)\right]\\ &\quad - \frac{2}{\Delta^3}\left[x_1^{2\beta} + (\beta+1)x_1^\beta(x_2^\beta + \cdots + x_n^\beta)\right]\beta x_1^{\beta-1}\\ &= \frac{\beta x_1^{\beta-1}(x_2^\beta + \cdots + x_n^\beta)}{\Delta^3}\{(1+\beta)(x_2^\beta + \cdots + x_n^\beta)\\ &\quad + (1-\beta)x_1^\beta\},\end{aligned}$$

which is positive if $0 \leqslant \beta \leqslant 1$. Now, for $j \neq 1$,

$$\frac{\partial g}{\partial x_j} = \frac{-x_1^{\beta+1}}{\Delta^2}(\beta x_j^{\beta-1}),$$

$$\frac{\partial^2 g}{\partial x_j^2} = \frac{\beta x_1^{\beta+1} x_j^{\beta-2}}{\Delta^3}\left[-(\beta-1)\Delta + 2\beta x_j^\beta\right],$$

which is also positive if $0 \leqslant \beta \leqslant 1$. □

Let $W_1(x_2, \dots, x_n) = \int_0^\infty g(x_1, \dots, x_n)\, dF(x_1)$. Since g is convex in each x_i, W_1 is convex in each of $x_2, \dots, x_n$. Now let

$$W_2(x_3, \dots, x_n) = \int_0^\infty W_1(x_2, \dots, x_n)\, dF(x_2).$$

Then W_2 is convex in each of $x_3, \dots, x_n$. Continuing in that manner, we finally get

$$E_F(n) = \int_0^\infty W_{n-1}(x_n)\, dF(x_n).$$

When $F \leqslant_{cx} G$ we have $W_i^F \leqslant W_i^G$ for $i = 1, \ldots, n$. Thus we obtain

PROPOSITION 3. *If* $0 \leqslant \beta \leqslant 1$ *and* $X_1, \ldots, X_n$ *are independent, then* $F \leqslant_{cx} G \Rightarrow E_F \leqslant E_G$. □

Thus, for $0 \leqslant \beta \leqslant 1$, a more uncertain field-size distribution, in the sense of convex order, results in a higher expected size of the first field discovered. If the field sizes are not independent, one needs to work with multivariate convex order, which is not standard, so we shall not pursue this.

The function g is not convex for $\beta > 1$. To explore the effect of increased variability in that case, a different approach is needed.

STANDARDIZED SIZES

Let $\mathrm{Var}(X_i) = \sigma^2, i = 1, \ldots, N$. The X_i's need not be independent. Define

$$X_i^* = X_i/\sigma, \quad i = 1, \ldots, N.$$

Then $\mathrm{Var}(X_i^*) = 1$ regardless of the variance of X_i. Now

$$\frac{X_1^{\beta+1}}{\sum_{i=1}^N X_i^\beta} = \sigma\left(\frac{X_1^{*\beta+1}}{\sum_{i=1}^N X_i^{*\beta}}\right).$$

Thus

$$E\left(\frac{X_1^{\beta+1}}{\sum_{i=1}^N X_i^\beta}\right) = \sigma E\left(\frac{X_1^{*\beta+1}}{\sum_{i=1}^N X_i^{*\beta}}\right).$$

Since the expectation on the right does not depend on σ, and is positive, its product with σ is clearly increasing in σ. So we proved

PROPOSITION 4. *If* $\mathrm{Var}\, X_i = \sigma^2, i = 1, \ldots, n$, *then* $E_F(n)$, *and hence* E_F, *are increasing in* σ. □

While this result holds for any β, the convex ordering result is stronger in the range where it applies ($0 \leqslant \beta \leqslant 1$), since convex ordering is considerably more general than ordering of variances. We thus included it.

CONCLUDING REMARKS

While our analysis focused on the first discovered field, our qualitative observations will apply to the first *several* discovered fields. Chungcharoen (1997) developed a simulation model of the hydrocarbon discovery process. The model samples from a field size distribution and a number of fields distribution, then follows a size-biased, sampling-without-replacement representation of the discovery process in a manner consistent with the expected value (1). The model produces distributions of the total volume and net present value of discoveries, for any specified number and timing of exploratory wells. Figures 1 and 2 show the calculated density functions for the case of 10 exploratory wells seeking natural gas fields off Canada's east coast. Each diagram shows two density functions: the one with the mode to the left is for the best estimates of all data, including the standard deviation of the field size distribution; the one further to the right is for the same data except for a doubling of the standard deviation of the field size distribution. Figures 1 and 2 clearly illustrate that an increase in the variability of the field size distribution leads not only to an increase in the variability of the volume or value distributions, but also an increase in their expected values. Indeed, it was this observation which motivated this Note.

There are two possible meanings for a measure of variability of field sizes. The first is that the variability is the real dispersion of sizes, i.e., that there are in fact fields of various known sizes, in unknown locations. The second is that the standard deviation represents, in part, the real dispersion, but also, to some extent, our ignorance about the dispersion of sizes. Regardless of the meaning, it makes perfectly good sense for a risk-neutral decision maker to prefer to explore in a region that has a larger variability but the same mean value of field sizes. If the first meaning of variability is valid, then the size-biased search process will tend to find early the more prevalent large fields that in fact exist in the higher-variability area. If, on the other hand, the second meaning is more valid, then it is still reasonable to prefer the more variable area because many such decisions will, on average, lead to larger first discoveries

Figure 1. Comparison of the distributions of the total volume between the cases of the regular standard deviation and doubled standard deviation of the field size distribution for 10 exploratory wells.

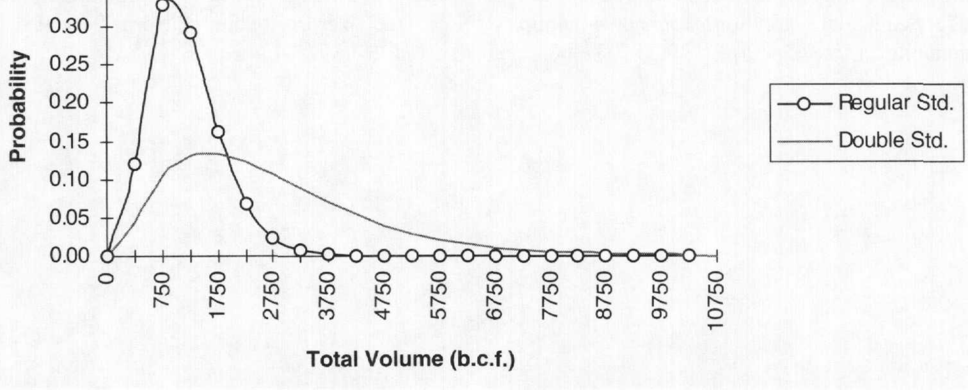
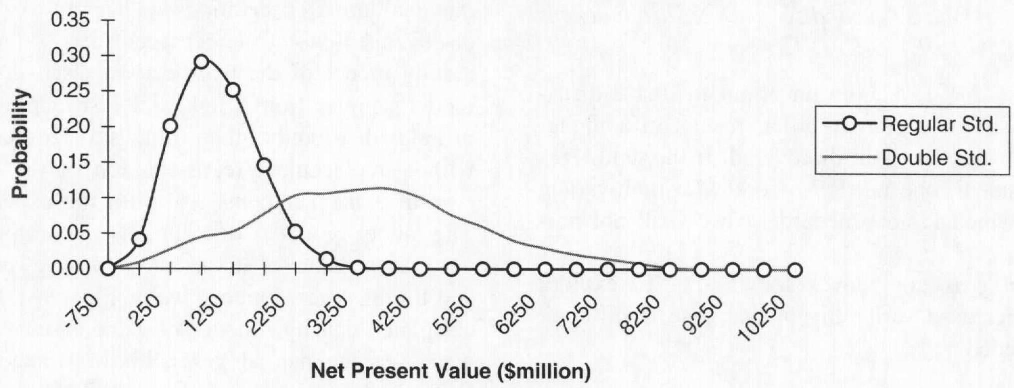


Figure 2. Comparison of the distributions of the net present value between the cases of the regular standard deviation and doubled standard deviation of the field size distribution for 10 exploratory wells.



than would a policy of choosing the less variable area. We cannot directly transfer our usual preference for less variability from, e.g., financial decisions, to the choice of areas to explore via a size-biased search process.

ACKNOWLEDGMENTS

The authors thank Benjamin Hobbs, Area Editor, the Associate Editor, and the referees for comments that led to improvements in this note. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Arnold, B. C., N. Balakrishnan, H. N. Nagaraja. 1992. *A First Course in Order Statistics*. Wiley, New York.

Arps, J. J., T. G. Roberts. 1958. Economics of drilling for cretaceous oil on east flank of Denver-Julesburg Basin. *AAPG Bull.* **42** 2549–2566.

Attanasi, E. D., L. J. Drew. 1985. Lognormal field size distributions as a consequence of economic truncation. *Math. Geology* **17**(4) 335–351.

Barouch, E., G. M. Kaufman. 1976. Probabilistic modeling of oil and gas discovery. F. S. Roberts, ed. *Energy: Mathematical models. Proceedings of a SIMS Conference on Energy.* SIAM, Philadelphia, PA, 248–260.

Chungcharoen, E. 1997. Economic analysis of hydrocarbon explorations by simulation with geological uncertainties. Ph.D. thesis, University of Waterloo, Waterloo, Ontario, Canada.

Cozzolino, J. M. 1972. Sequential search for an unknown number of objects of nonuniform size. *Oper. Res.* **20**(2) 293–308.

Drew, L. J., E. D. Attanasi, J. H. Schuenemeyer. 1988. Observed oil and gas field size distributions: A consequence of the discovery process and prices of oil and gas. *Math. Geology* **20**(8) 939–953.

Johnson, N. L., S. Kotz, N. Balakrishnan. 1994. *Continuous Univarite Distributions*. Vol. 1, 2nd ed. Wiley, New York.

Lee, P. J., P. C. C. Wang. 1983a. Probabilistic formulation of a method for the evaluation of petroleum resources. *J. Internat. Soc. Math. Geology* **15** 163–181.

——, ——. 1983b. Conditional analysis for petroleum resource evaluations. *J. Internat. Soc. Math. Geology* **15** 349–361.

——, ——. 1985. Prediction of oil or gas pool sizes when discovery record is available. *J. Internat. Soc. Math. Geology* **17**(2) 95–113.

O'Caroll, F. M., J. L. Smith. 1980. Probabilistic methods for estimating undiscovered petroleum resources. *Adv. Econom. Energy Resources* **3** 31–63.

Power, M. 1992. The effects of technology and basin specific learning on the discovery rate. *J. Canadian Petroleum Tech.* **31**(3) 49–52.

——, J. D. Fuller. 1991. Predicting the discoveries and finding costs of natural gas: The example of the Scotian shelf. *Energy J.* **12**(3) 77–93.

Rudin, W. 1976. *Principles of Mathematical Analysis*, 3rd ed. McGraw Hill, New York.

Schuenemeyer, J. H., L. J. Drew. 1983. A procedure to estimate the parent population of the size of oil and gas fields as revealed by a study of economic truncation. *Math. Geology* **15** 145–161.

Shaked, M., J. G. Shanthikumar. 1994. *Stochastic Orders and Their Applications.* Academic Press, New York.

Smith, J. L., G. L. Ward. 1981. Maximum likelihood estimates of the size distribution of North Sea oil fields. *Math. Geology* **13**(5) 399–413.