*World Oil Resources: A Statistical Perspective**

DeVerle P. Harris

College of Mines
University of Arizona
Tucson, Arizona

Donald E. Myers

Department of Mathematics
University of Arizona
Tucson, Arizona

* Originally prepared for Pan Heuristics, a division of R & D Associates, Marina Del Rey, California 90291.

ISBN 0-12-014904-4

I. INTRODUCTION

A. Motivation

This study was motivated by two scholarly publications: Bernardo Grossling's book *Window on Oil* (Grossling, 1976) and Richard

Nehring's study *Giant Oil Fields and World Oil Resources* (Nehring, 1978). Grossling's primary objective was the evaluation of the oil potential of lesser developed regions of the world. Even so, Grossling's work is often cited for his estimates of recoverable oil and gas resources of the world, which are more optimistic than other estimates[1]:

Recoverable oil	$1{,}960 \times 10^9$	to	$5{,}600 \times 10^9$ bbl
Recoverable gas	$11{,}200 \times 10^{12}$	to	$28{,}000 \times 10^{12}$ ft^3

Of course, given the deep concern about current and future oil supplies, estimates of the world's recoverable oil or that of any constituent region receives immediate and critical examination. Such examination was intensified by the publication by the Rand Corporation of Nehring's study (1978). Nehring estimates[1] that recoverable oil resources are in the range of $(1700–2300) \times 10^9$ bbl. This range is more compatible with other previous estimates[1] (see Table I) than with Grossling's estimate. The suggestion that Grossling's estimates are optimistic is further supported by the result of a Delphi experiment conducted by the World Energy Conference (1978): $(1920–2420) \times 10^9$ bbl.

Examination of estimates other than Grossling's suggests 2000×10^9 bbl as a representative quantity (see Table I). The essence of the debate on world oil resources can be crudely put by posing the following questions: Does this commonality represent a convergence on truth or a convergence due to a "herding" tendency? Such a tendency has been well documented by psychometricians. Is Grossling's dissenting opinion a reflection of an optimistic bias, or is it the one clear vision of the magnitude and uncertainty of world oil resources? Answers to these questions turn on geological issues, the technologies and sampling efficiency of exploration, the ability of exploration to preferentially select the most productive ground, the relationship of size of total oil resources to minimum deposit size, and the bias in exploration results introduced by political and economic issues.

The judgment could be made that Grossling's high estimate results from the selection of an overly optimistic upper richness factor (barrels of oil per square mile). However, the value used by Grossling is less, on a per-unit-of-area basis, than the sum of cumulative production, known reserves, and expected future reserve additions to known de-

[1] A brief description of the methodologies employed in these estimates is provided in Appendix A.

TABLE I

Recent Estimates of Ultimate World Crude Oil Production[a]

Year	Author	Organization	Quantity (10^9 bbl)	References
1962	L. G. Weeks	Consultant	2000	Cited by W. P. Ryman in Hubbert (1969), Table 8.2
1965	T. A. Hendricks	U.S. Geological Survey	2480	Hendricks (1965), Table 6
1967	W. P. Ryman	Esso	2090	In Hubbert (1969), Table 8.2
1968	Shell	Shell	1800	Warman (1971)
1969	M. King Hubbert	National Academy of Sciences, National Research Council	1350–2000	Hubbert (1969), Fig. 8.23
1971	H. R. Warman	British Petroleum Ltd.	1200–2000	Warman (1971)

1972	J. D. Moody and H. H. Emmerich	Mobil	1800–1900	Moody and Emmerich (1972)
1972	Richard L. Jodry	Sun	1952	Cited by Hubbert (1974), Table 9 and Fig. 6
1974	J. D. Moody and R. W. Esser	Mobil	2000	Moody and Esser (1975)
1975	B. Grossling	U.S. Geological Survey	1960–5600	Grossling (1976), p. 105
1977	H. D. Klemme	Consultant	1000–1500 (undiscovered)	Klemme (1977)
1978	R. Nehring	Rand	1700–2300	Nehring (1978)
1978	World Energy Conference		(2170) 1920–2420	World Energy Conference (1978)

[a] After a table compiled by Attanasi and Root (1981) (from Hubbert, 1974), with some additions by the present authors.

posits for one large resource region (USSR); therefore, such a judgment is really saying that while such a richness has been observed, it is unreasonable to consider that factor as representative of the entire world, even when estimating the quantity of total recoverable oil (including undiscovered). Such a statement is both an evaluation of physical, economic, and technologic issues and a statement about the equivalence of likelihood with the condition of reasonableness.

B. Objective

The objective of this study is to develop two probability distributions, one for recoverable oil plus oil equivalent of gas (OE)[2] resources and one for only recoverable oil resources of the world. This objective includes the examination of existing estimates of world oil resources with respect to these distributions. Preferably the probability distributions would be relatively free of subjective judgments. Even partial achievement of such an objective requires the adoption of a model which is data driven. Such a requirement, given our current states of knowledge and available data, rules out a geologically based model. Remaining options include econometric models and discovery process models. The model employed in this study is a very simplistic form of the latter of these.

C. Structure of This Article

The analysis employed in this study is mathematical in nature; consequently, some of its descriptions employ substantial mathematical notation. A complete appreciation of the analysis in detail requires a thorough understanding of regression analysis and statistical inference, as well as methods of Monte Carlo simulation. However, this article is designed so that the main ideas involved in the analysis can be understood and the results appreciated by a more general reader who reads only Sections I, VII, and VIII. Anyone interested in the details of the analyses also should read the introduction first, however, because this

[2] Hereafter in this report the abbreviation OE is used to refer to the sum of oil and the oil equivalent of gas, based upon the conversion factor of 4440 ft^3/bbl; this is the conversion factor selected by Grossling in his study of world oil resources.

overview of the study will assist him in understanding the remainder of the article.

D. Overview of Approach

1. General

This study considers world resources of oil and oil equivalent (OE), H_T, to be the sum of two separate quantities: H_M, the OE resources of the Middle East, and H_{ROW}, the OE resources of the rest of the world:

$$H_T = H_M + H_{ROW}$$

The major part of this study concerns the objective, as opposed to subjective, estimation of H_{ROW}. A published estimate of H_M is added to the estimate of H_{ROW} to give an estimate of H_T.

The model employed in this study for the estimation of H_{ROW} is empirical in nature. Simply stated, it builds on the correlation observed by Grossling (1976) between number of wells[3,4] per square mile and the barrels of OE[5] per square mile (see Table II). Table II lists the drilling density (wells per square mile) and the density of OE (barrels per square mile) for each of 14 resource regions[3] of the world. OE density

---

[3] This study took as a point of departure the data assembled by Grossling: the definitions of regions and the selection of number of wells, instead of footage drilled, are based upon this decision.

[4] Menard and Sharman (1975) have used drilling densities in a rather different way which in part suggests a possible future improvement. They used drilling data to explore the efficiency of exploration techniques by comparing a random search for oil fields with the historical success rate. Although their results provide an estimate of remaining undiscovered oil, the estimate pertains only to the contiguous 48 states and hence is not comparable to our results. However, the drilling data they used, which included hole depths, were converted to standardized lengths. This suggests that if the data were available in the appropriate form, drilling densities in our model could be given by a count of standardized lengths. This was not possible using Grossling's data. In most other respects the objective of our analysis and that of Menard and Sharman is quite different.

[5] Using the conversion factor of 4440 ft^3 gas/bbl of oil.

TABLE II

Drilling Density and Oil Equivalent Found per Unit Prospective Area[a]

Region	Wells per square mile[b]	Petroleum EVRD[c] (bbl oil equivalent per square mile)
Africa (excluding A below)	0.00094	6,501
Australia and New Zealand	0.0010	9,012
S and SE Asia mainland	0.0049	15,949
Algeria + Egypt + Nigeria + Libya = A	0.0057	117,950
Middle East	0.0059	671,243
Latin America (excluding B & C below)	0.0068	7,184
China, P.R.	0.0078	56,489
Western Pacific	0.0085	32,851
Argentina + Mexico = B	0.033	38,955
Western Europe	0.036	125,030
Canada	0.079	31,333
Colombia + Trinidad and Tobago + Venezuela = C	0.082	144,213
USSR	0.152	176,676
USA, conterminous	0.957	155,896

[a] Source: Grossling (1976), p. 68, Table 29. Reprinted with the permission of Financial Times Business Information, London.
[b] Total wells (exploratory + development).
[c] Gas converted to oil at 4440 ft^3 per barrel.

is Grossling's EVRD (expected value of recoverable discoveries) of oil and oil equivalent of gas per square mile. The meaning of EVRD can be appreciated by examining Fig. 1, which has been reproduced from Grossling: EVRD = cumulative production + extended reserves. Thus, EVRD includes cumulative production, known reserves, and reserve additions that are expected to be made to known deposits. It does not include quantities of OE in undiscovered deposits.

The data of Table II exhibit a very great variation in the OE densities of the 14 regions. Let us assume that this difference, on average, represents variation in the drilling densities of the regions. Then, if we knew

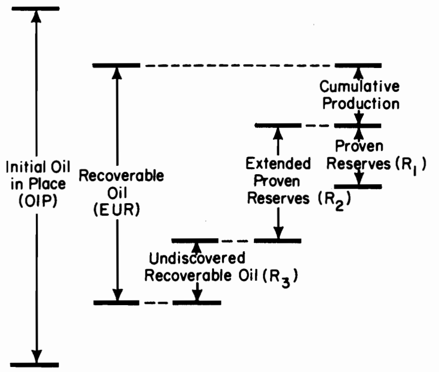

Fig. 1. Span of the various resource and reserve terms for oil. [Source: Grossling (1976), p. 54, Fig. 1. Reprinted with the permission of Financial Times Business Information, London.]

the quantitative relationship between drilling density and OE density, we could estimate what the OE density of a region would be, on average, if it were drilled at some higher density. In this way, we could estimate the quantity of OE that will be discovered in that region in the future as drilling proceeds. Simplistically, this study first seeks to quantify the statistical relationship between these two densities; then this relationship is used to estimate ultimately recoverable OE reserves of the world, given various drilling densities, densities which are thought to be high enough to thoroughly test the region.

Simply stated, $\hat{H}_{\text{ROW}}$ (an estimate of H_{ROW}) is the product of area A and $\hat{V}_{\text{ROW}}$, estimated OE density (bbl/mile^2):

$$\hat{H}_{\text{ROW}} = A \cdot \hat{V}_{\text{ROW}}$$

where $\hat{V}_{\text{ROW}}$ is the number of estimated barrels of OE per square mile, A is the area of the region, and $\hat{H}_{\text{ROW}}$ is the estimated quantity of OE for the region.

The basic approach of this study is to estimate a statistical relationship of OE density to drilling density and then to use this relationship by evaluating it at some specified density, a density high enough to imply thorough exploration. The resulting value for V_{ROW} is multiplied, as in the above equation, by area to give the quantity of OE that would be discovered in the region. For high drilling densities, the quantity discovered approaches the total quantity that exists in the region.

As indicated, an essential step in estimating H_{ROW} is the use of a statistical relationship between V_{ROW} and drilling density. The resulting estimate is no more reliable than is the estimated relationship. Therefore, explicit consideration of uncertainties is necessary; furthermore, these uncertainties (errors) are the bases for the probability dimension used in this analysis.

2. *Uncertainty, Expectation, and Estimation*

The essential problem in making estimates of world OE resources is how to use information about past discoveries to project the estimate to include as yet undiscovered resources. By its very nature there will be uncertainty associated with any such extrapolation. Even if the estimate were given as a range of values, there would be some level of uncertainty. This uncertainty could be quantified in at least two ways. If the range of values is interpreted as an interval for a random variable X with density function $f(x)$, then

$$P(L \leq X \leq U) = \int_L^U f(x)\, dx$$

where L and U are lower and upper values. None of the estimates given in Table I includes identification of such a density. Alternatively, an interval estimate might be interpreted as a confidence interval for the mean; in this case it is necessary to specify the confidence level. When the estimates incorporate subjective appraisals, it is difficult to determine probabilities or confidence levels; in fact, in such cases it is difficult to determine the credibility of such estimates except after the fact.

The most serious consequence of utilizing subjective methods is the possibility of being influenced in such judgments by the results. If particular published estimates are judged to be "optimistic" or alternatively "reliable," there will likely be a tendency to interpret evidence or data accordingly when making new estimates.

One of the advantages of a statistical approach is to avoid such a possibility. Another is the mechanism for updating estimates as additional data become available with a resultant decrease in the uncertainty associated with the new estimates. Some of the appraisals given in Table I are based on unpublished methods and hence it is not possible to compare those with statistical methods; some other estimates incorporated subjective methods.

Of course, there are serious difficulties associated with trying to give a probability characterization to an interval of values and likewise if a confidence interval for the mean is to be obtained. Note that certain difficulties are intrinsic. As $U - L$ decreases, in general we also have $P(L \leq X \leq U)$ decreasing; that is, more specific estimates have greater uncertainties attached. Additional data may improve the estimate of the probability but cannot counteract the above relationship. Similarly, the width of a confidence interval can be decreased with fixed confidence level with additional data, but the mean is only one characteristic and does not preclude the intrinsic variability.

In summary, we see that a statistical approach recognizes the uncertainty that is intrinsic in making resource appraisals and attempts to quantify it but does not remove it. Since the density function is almost certainly not known and it is not possible to sample directly to estimate the mean, we must consider statistical models that incorporate a dependence on observable variables. In Section I,D,3 we will summarize the approach to be followed.

3. *Regression Models and Simulated Random Sampling*

Partly because of the discrepancy between Grossling's estimates and others in Table I and his use of drilling densities, it seemed reasonable to utilize drilling densities as a starting point. One of the most common tools used to estimate or predict one variable in terms of a dependency on one or more other variables is the regression equation. In its simplest form, Y is considered the dependent variable and X the independent variable, and the model is

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where β_0, β_1 are constants and ε is a normal random variable with mean zero and variance σ^2. The mean of Y is a linear function of X. Using data pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$, the parameters $\beta_0, \beta_1, \sigma^2$ can be estimated by regression analysis. Given these estimated parameters, the mean of Y and prediction intervals for Y can also be obtained.

Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be estimators for β_0 and β_1. Then the result of regression analysis is an equation that can be used for predicting the mean value of Y, given a value for X:

$$E[Y_x] = \hat{\beta}_0 + \hat{\beta}_1 X$$

where $E[Y_x]$ is the expected (average) value for Y given X. Suppose that instead of using the regression equation to estimate an expected value of Y given X, we use it to generate the population of values of Y_x represented by $E[Y_x]$. Any value of Y_x within this population can be described as the expected value of Y given X modified by an error term ε:

$$Y_x = E[Y_x] + \varepsilon$$

Let us specify ε to be the product of $t^{(n-2)}$, a t-distributed random variable having a parameter of $n - 2$ degrees of freedom, and γ, the standard error of the estimate of Y_x by $\hat{\beta}_0 + \hat{\beta}_1 X$:

$$Y_x = \underbrace{\hat{\beta}_0 + \hat{\beta}_1 X}_{E[Y_x]} + \underbrace{t^{(n-2)} \cdot \gamma}_{\varepsilon}$$

Since there is an infinite population of values for $t^{(n-2)}$, there is an implied infinite population of values for Y_x, one for every value of $t^{(n-2)}$. Thus, for a given value of X, there is a probability distribution for Y_x; furthermore, the population of values of Y_x can be simulated by combining the expected value (mean) of Y_x with an error term derived by drawing a t-distributed random variable and multiplying it by the standard error of the estimate for Y_x.

The regression model described above allows transformations on Y or X or both, in particular logarithmic. The logarithmic transformation was utilized in this study together with an areal adjustment. The latter was necessary to meet the constant variance assumption implicit in the assumed model. If X or Y is replaced by transformations, such as logarithmic, the confidence interval does not easily transform and may produce only a one-sided confidence interval[6] if the transformation is

---

[6] In particular, if $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimators for β_0 and β_1, S^2 is the estimator for σ^2, and $S^2_{\hat{\beta}_0+\hat{\beta}_1 x}$ is the estimator for $\mathrm{Var}(\hat{\beta}_0 + \hat{\beta}_1 X)$, then the confidence interval for the mean of Y, i.e., $\beta_0 + \beta_1 X$, is given by

$$\hat{\beta}_0 + \hat{\beta}_1 X - t_{\alpha/2} S_{\hat{\beta}_0+\hat{\beta}_1 x} \leq \beta_0 + \beta_1 X \leq \hat{\beta}_0 + \hat{\beta}_1 X + t_{\alpha/2} S_{\hat{\beta}_0+\hat{\beta}_1 x}$$

The value $t_{\alpha/2}$ is obtained from a table for the t-distribution; it is the point such that the area to the right is $\alpha/2$ and it corresponds to a confidence level of $(1 - \alpha)100\%$. For example, if we desire a 99% confidence level, then $\alpha/2 = 0.005$. The degrees of freedom for the t-table are given by $n - 2$, where n is the number of pairs of data points used. The prediction interval is also centered at $\hat{\beta}_0 + \hat{\beta}_1 X$ but incorporates the additional uncertainty represented by "predicting" a random variable from its mean and variance. The prediction interval is given by

$$\hat{\beta}_0 + \hat{\beta}_1 X \pm t_{\alpha/2}(S^2 + S^2_{\hat{\beta}_0+\hat{\beta}_1 x})^{1/2}.$$

nonlinear—that is, if $Y = \ln U$, then in general $E[Y] \neq \ln E[U]$. The prediction interval may be transformed, and hence it is somewhat more useful.

In the model initially employed in this study, drilling density (W) is the independent variable, and OE density (V) is the dependent variable:

$$\ln V = \beta_0 + \beta_1/\ln W + \varepsilon$$

However, it later became apparent that the constant variance assumption of the regression model would not be satisfied. Consequently, the initial model was replaced by the following model:

$$Y = \beta_0 X_1 + \beta_1 X_2 + \varepsilon'$$

where $Y = \ln V \cdot g(A)$, $X_1 = g(A)$, $X_2 = g(A)/\ln W$, $g(A)$ is a function of area A, and $\varepsilon' = g(A) \cdot \varepsilon$.

In this form it is the variance of Y that must be constant (with respect to X_1 and X_2). Of course, the confidence interval and prediction intervals will be for $E[Y]$ and for Y, respectively. It was necessary to consider the validity of this constant variance assumption, as well as several others, and appropriate statistical tests. This "curve fitting" approach for matching OE with drilling densities adjusted for areas has considerable intuitive appeal and simplicity of application, but it should not be assumed that drilling densities and area adjustments are the only variables that might be used.

There was a subjective decision to exclude the Middle Eastern fields from this regression model. In order to produce a world estimate it was then necessary to assimilate the estimates from the regression model with estimates from the Middle East. This was accomplished by considering the resources from the Middle East and the rest of the world (ROW) as two independent random variables. Using Nehring and Grossling's data for the Middle East a rectangular distribution was fitted. After choosing a drilling density, the regression model was used to simulate a population value for H_{ROW} in terms of the t-distribution. Using a Monte Carlo simulation, each random variable was repeatedly sampled and the values added, and thus the distribution for the sum was constructed (see Fig. 2).

While it is not possible to show that the regression model is correct, the application of the model is described in detail and the underlying hypotheses are examined. If additional data become available, these can also be incorporated into the model. Finally, it should not be as-

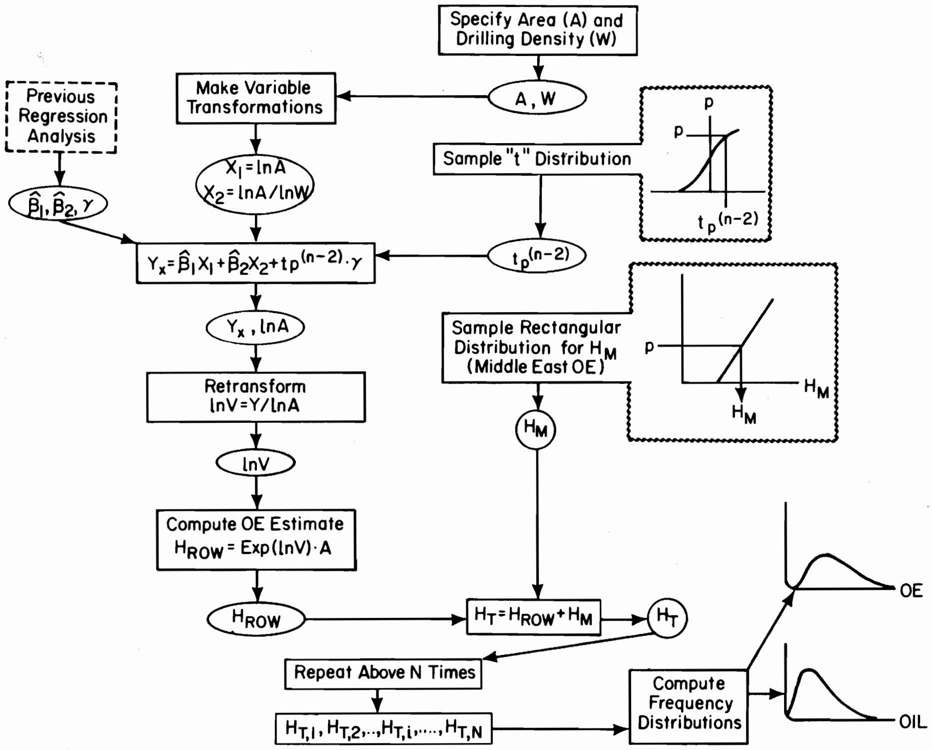

Fig. 2. Schematic diagram of procedure for estimating world resources of OE and oil.

sumed that the use of such a model implies a causal relationship between drilling densities and OE.

E. A Word on Data and Information

This study took as a departure point the basic data compiled by Grossling (1976). The generation of the data of Table II, which is the data base for this study, required Grossling to bring together by region the following items: (1) data on cumulative production, (2) an estimate of known reserves, (3) an estimate of prospective area, (4) an estimate of future reserve additions to known deposits, and (5) data on the number of wells. It should be noted carefully that estimates of the quantities of oil and oil equivalent of gas in *undiscovered* deposits are *not* included in the above items. Even so, since this study used relationships developed from these five items to estimate the recoverable oil and oil equivalent of gas in undiscovered deposits, the nature and source of the information used merits examination.

Only the first and the last of these quantities, cumulative production and number of wells, qualify as data in the usual sense. While they contain errors, the errors are of the usual kind: error of measurement and of reporting. Generally, the potential magnitude of these errors, as a percentage of the quantity recorded, is greater for the LDC countries; this is particularly the case for data on drilling. In fact, it is only through the diligent and impressive efforts of Grossling (1976) that we have a consistent data set on the number of wells for some of the LDC regions.

Items (2) and (4) clearly are estimates. Of these, the estimate of known reserves probably contains the smallest error. Prospective area in a comprehensive sense is a quantity about which there is much uncertainty. To begin with, resource theory is not developed to the point where a set of experts would agree on the amount of truly prospective ground, even if geologic information were uniformly available for all regions. As noted by Grossling (1976, p. 12):

> The outline of prospective areas has changed in the course of time, and also varies from author to author depending upon the criteria used.

He states further (p. 15):

> Not all of the relevant factors can be ascertained beforehand. And, what is more important, not all the relevant factors are known even today after more than a century of petroleum exploration and development.

To compound the problems, there is a great variation in the amount and quality of geologic information about the various resource regions of the world. In view of these problems, Grossling adopted a very simplistic definition of prospective ground: Prospective area is that part of a region which consists of sedimentary rocks having a thickness of at least 2000 ft, including continental shelves to a water depth of 200 m. Such a definition of prospective ground ignores both the thickness dimension (given that the thickness meets the minimum requirement) and the nature of the sedimentary rocks; consequently, it is a poor description of true prospectiveness. Grossling defends the use of this poor measure on the grounds that it is the only measure which was uniformly available. While conceptually the measure suffers, the fact that it is based upon data which are uniformly available is very important to the application of statistical models. While the areas reported may have measurement and reporting errors, these errors are not believed to be large with respect to the units of measure. Furthermore, it seems reasonable to assume that overstated areas may be compensated by understated areas.

An issue that is related to prospective area is that of the delineation of regions, which in this study consisted in aggregating various political entities into resource regions. Again, we took as given the regions identified by Grossling (1976) because data on drilling and OE densities were available to us for only these regions. The aggregation of regions can raise complex problems in statistical analysis. For further discussion on aggregation and its effect on estimates, see Section V,D.

The quantity referred to by Grossling as EVRD and used to get the OE densities is the sum of cumulative production, estimated reserves, and expected future additions to reserves of known deposits. An estimate of the last of these three items, expected future reserve additions, is subject to considerable error. Estimates by two experts may vary considerably. And the estimate employed by Grossling may err considerably for a given region. As long as the effects of the total errors in EVRD compensate across regions, the model will yield unbiased estimates. Lacking information about the nature of these estimates, it seems reasonable to assume that this condition is met.

F. Oil Equivalent

The models employed in this study for statistical inference are predicated on the sum of oil and oil equivalent of gas. This study converted all gas to oil equivalent by using the conversion factor of 4440 ft^3 gas/bbl oil. As with the other data employed in statistical analysis, this factor was taken from Grossling (1976). The authors did not make an independent estimate of this conversion factor; furthermore, the effect of a different factor on the estimates was not examined.

Everything else being equal, OE density is preferred to oil density on empirical and conceptual grounds. First, many deposits contain oil and gas. Furthermore, given the fact that much exploration drilling is directed to structure-based anomalies and that differential entrapment may cause an anomaly to be a gas instead of an oil deposit, a density measure which reflects oil and gas is desirable. On a pragmatic basis, OE density is preferred because it shows less variation across regions with respect to drilling density than does oil density alone.

It is acknowledged by the authors that the selection of the appropriate conversion factor becomes especially problematic in view of the unequal effect of deeper drilling on oil and gas discoveries. Clearly, recent drilling depths, which are significantly greater than those reflected in past drilling and discovery data, would result in a different conversion factor, reflecting the relatively greater frequency of occurrence of gas than oil at greater depths. That being the case, the question arises naturally about the credibility of the estimates of this model when they are based upon a conversion factor which reflects depths of drilling that are shallower than future drilling depths will be. This question cannot be answered by the authors in a quantitative fashion at this time. However, consideration of the effect of drilling depth on the conversion factor prompts the following statement:

> The estimates made in this study assume that drilling depths of the future will fall, on average, within the same intervals as they have in the past. With this assumption, the only thing that changes in the model when drilling density increases is the *number* of wells, i.e., density of drilling.

The fact that average well depths now are greater than they used to be and that these depths will probably increase in the future means that the greater drilling densities of the future may be accompanied by an increase in the conversion ratio and an increase in the density of oil plus oil equivalent of gas. In general, the use of 4440 ft^3/bbl as a conversion factor suggests that estimates by the model will tend to be conservative with respect to the total of oil plus oil equivalent of gas, but optimistic with respect to the fraction of this total that is oil.

G. A Preview of Results

Interpretation of the results of this study must recognize that this study is incomplete at present because it has not considered explicitly the economic implications of the drilling densities used for inference. Consequently, all statements must be considered conditional upon the appropriateness of the drilling density. With this caveat, a few comments are offered as a summary of the study.

Drilling densities of 1.5 and 2.0 $wells/mile^2$ result in modal estimates of the statistical model of this study which, overall, are supportive of recent subjective estimates of world oil resources. Point estimates of world oil resources for these drilling densities are 2029×10^9 and 2154×10^9 bbl, respectively. These quantities lie within the ranges reported by Nehring [(1700–2300) $\times 10^9$ bbl], Grossling [(1960–5600) $\times 10^9$ bbl], and the World Energy Conference [(1920–2420) $\times 10^9$ bbl].

It is contended that the appropriate comparison of the model's estimates to the subjective estimates of Table I uses the model's modal (most likely), not mean, value. On this basis, there is a high degree of conformity. Since a drilling density of 1.5 wells/mile^2 is only 50% greater than the current density of the United States, it seems conceivable that such a density for the ROW region (prospective area of the world exclusive of the Middle East region) is possible and may not be incompatible with the economic reference for what has been reported by most estimators as "ultimately recoverable oil resources." Given current drilling rates, such a density for the United States is not far away. Some subregions of the ROW region may never achieve such a density, but it seems possible that others (such as the United States) may exceed it.

While point estimates by the model seem to be compatible with geological estimates of world oil resources, the 90% confidence ranges produced by the statistical analysis, given drilling densities of 1.5 and 2.0 wells/mile^2, are broader than recently estimated ranges by Grossling, Nehring, and the World Energy Conference:

Nehring	(1700–2300) $\times 10^9$ bbl
Grossling	(1960–5600) $\times 10^9$ bbl
World Energy Conference	(1920–2420) $\times 10^9$ bbl
This statistical model:	
Drilling density of 1.5	(1212–8220) $\times 10^9$ bbl
Drilling density of 2.0	(1234–8335) $\times 10^9$ bbl

Neither Nehring nor Grossling provides probability or confidence statements about his range; consequently, it is not possible to really compare unequivocally the results of the statistical model with their ranges. If these ranges are considered to bound possible states of nature at a high level of confidence, e.g., 90–95%, then the results of the statistical model disagree markedly with Nehring and the World Energy Conference about the breadth of the range (see curve C of Fig. 11, p. 145). Disagreement is less with Grossling.

A criticism of estimates of the statistical model that merits thoughtful consideration is that the breadth of the ranges, and the magnitude of the mean values for selected drilling densities, reflect the fact that the model treats important influences, such as geology, infrastructure, and political economics as a random disturbance, and that if these factors were taken into account, the distribution of possible states would be much narrower. Everything else being equal, it must indeed be true that the variance of the estimate of a statistical model which is extended to include additional factors, where real information exists, would be less than that for a simple drilling model. While conceptually this criticism is appropriate, it loses strength when we are comparing subjective geological estimates to the model estimates, for everything else is not equal in this case. For evaluation and comment on these and other issues of information and inference see Sections VII and VIII.

II. THE INITIAL MODEL AND ASSUMPTIONS

A. Structure of the Model

The basic proposition underlying the initial analysis of this study can be formally stated as follows:

$$H = A \cdot f(W) \cdot e^u \qquad (1)$$

where H represents the barrels of OE (oil plus oil equivalent of gas), W is the number of wells per square mile, A is the prospective area (in square miles) of the region, $f(W)$ is the function which relates W to H, given A, and u is an error term which reflects the host of geologic, economic, technologic, political, terrain, climatic, and infrastructure factors that affect H but are not included in Eq. (1).

If we divide Eq. (1) by A, and let $H/A = V$, we have an alternative expression, one which is normalized on area:

$$H/A = f(W) \cdot e^u \qquad (2)$$

Alternatively,

$$V = f(W) \cdot e^u \qquad (3)$$

where V represents the barrels of OE per square mile. Taking logarithms of both sides of Eq. (3), we have the following:

$$\ln V = \ln[f(W)] + u \qquad (4)$$

In a world in which $\ln V$ is described by Eq. (4) it is futile to speak of a correct estimate of $\ln V$ for a given value of W because of the magnitude of u, the sum effect of all excluded influences on ln V. Instead, we strive for an estimate of ln V which has desirable statistical properties, such as being unbiased, and which is accurate on average.

Let us define $E[\ln V_w]$ as the expected value of ln V, given W; in simple terms this is the average value of ln V for a particular value of W, i.e., the effects of u have been averaged out. If we take the expectation of Eq. (4), we see that $E[\ln V]$ is equal to the value of $\ln[f(W)]$ provided that the expected value of u is zero:

$$E[\ln V_w]^7 = E[\ln[f(W)]] + E[u] = \ln[f(W)] \quad (5)$$

when $E[u] = 0.0$. Thus, if we knew $\ln[f(W)]$, we could compute the average value of ln V. In practice, $f(W)$ is not known. We must specify the form of f and then estimate its parameters. The estimated $\ln[f(W)]$ provides an estimate of $E[\ln V_w]$. Under certain conditions, regression analysis provides an unbiased estimate of $E[\ln V_w]$. These conditions include the following:

1. $\ln[f(W)]$, or some transformation, is linear in W
2. $E[u] = 0.0$
3. $E[u_i \cdot u_j] = 0.0$
4. $E[u_i^2] = \sigma_u^2$

(6)

Thus, the task in this study is to specify a form of $\ln[f(W)]$ that both fits the data and meets the above conditions.

B. Specification of $f(W)$

The specification of any model should be compatible with relevant theory. In this case, however, the very complex and dynamic world in which OE resources are defined is reduced to a drilling relationship and an aggregate error term. Consequently, there is very little relevant theory. In a static world in which economic, political, infrastructure, and climatic factors are constant for all regions and in which any region is equally likely to be selected for exploration, drilling and OE densities for a subset of regions may serve as a basis for specifying $\ln[f(W)]$ and estimating its parameters. The model so quantified could be used to

[7] $E[\ln V_w]$ is shorthand notation for $E[\ln V|W]$, i.e., the expectation of ln V given W.

infer the OE densities for all other regions within the universe of regions. But even for this ideal world, theory indicates only that $f(W)$ should become asymptotic to an upper limit as the number of wells per square mile increases to a large number.

The circumstances under which the data of Table I were generated depart from these ideal circumstances. However, the plot of $\ln(V)$ versus $\ln(W)$ in Fig. 3 suggests that when the Middle East is deleted, the data exhibit a generally asymptotic pattern. Thus, if the exclusion of the Middle East region can be justified, the adoption of an asymptotic relationship may be indicated on empirical grounds, even though the strict conditions for a theoretical selection are violated. But under what conditions can the Middle East be deleted from the sample? Discarding data simply because they don't conform to a prescribed model certainly qualifies as dubious statistical procedure. Clearly, discarding is acceptable only when there is evidence of a mixing of populations or when data are knowingly contaminated in some fashion, for example, the effect of a world war on demand for tungsten or the effect of price controls on a time series of prices.

By virtue of its size and the prolific yields of its oil fields, the Middle East region certainly has no known equal in the sample of producing

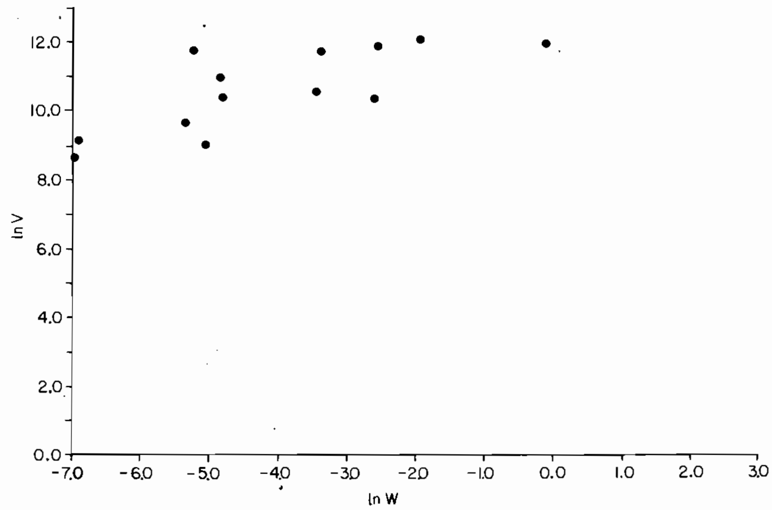

Fig. 3. Data on drilling and OE densities (Middle East has been excluded) [source of data: Grossling (1976)].

regions. If all unknown regions are similar to the regions composing our sample, exclusive of the Middle East, then the argument would be strong for defining two populations: the Middle East and the rest of the world. While we cannot discount totally the possibility that there exist one or more unknown regions similar to the Middle East, the probability that such an accumulation exists and has not yet been discovered is generally considered to be quite small. Thus, by virtue of the anomalous appearance of the Middle East in the bivariate plot of Fig. 4 and the evidence of the uniqueness of the Middle East as an oil resource region, the resource regions of the world here are partitioned into two populations: the Middle East (ME) and the rest of the world (ROW). The model, $f(W)$, will be specified on the data from the producing regions of Table II, exclusive of the Middle East. Once $f(W)$ is quantified, it will be employed to estimate densities of recoverable OE.

The specification of $f(W)$ is

$$f(W) = \exp[\alpha_0 + (\alpha_1/\ln W)] \qquad (7)$$

where α_0 and α_1 are the parameters to be estimated.

The specification of Eq. (1) as multiplicative in the relationship between $A, f(W)$, and e^u, and the specification of Eq. (7) as an exponential accommodates the empirical evidence that quantities of oil and gas

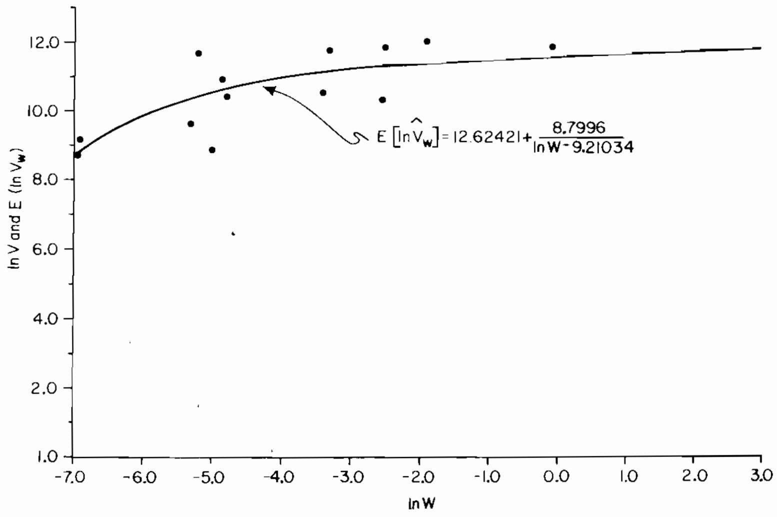

Fig. 4. Fit of the model to the data.

in fields and in regions commonly are distributed log–normally; that is, when frequency is plotted against the logarithm of size, the resulting plot resembles a normal distribution. The significance of this feature will be more apparent later, when the issue of estimation of $f(W)$ is considered. For the present, it serves only to suggest that it is reasonable to express OE density, the quantity described by $f(W)$, as a logarithm, ln V, and that we will be attempting to estimate $E[\ln V_w]$.

Taking the logarithms of both sides of Eq. (7), we have

$$\ln[f(W)] = \alpha_0 + (\alpha_1/\ln W) = E[\ln V_w] \tag{8}$$

C. Estimation of the Model

Suppose that we redefine the variables in Eq. (8):

$$Y = E[\ln V_w] \text{ and } X = 1/\ln W \tag{9}$$

The result is a simple linear equation, the parameters of which can be estimated by straightforward regression analysis:

$$Y = \alpha_0 + \alpha_1 X \tag{10}$$

Clearly, the data in Table II in their present form are not compatible with Eq. (10). These data must be transformed by taking the logarithms of each of the numbers in columns 1 and 2 and by forming the reciprocal of each logarithm in column 1. The resulting data can be considered as observations on the variables X and Y of Eq. (10). Columns 1 and 2 of Table II represent a sample of 13 joint observations on W and V, respectively. Thus, by transforming these data and subjecting them to regression analysis, estimates of α_0 and α_1 can be obtained. With one minor modification, this is basically what was done in this study. For computational purposes, the data in Table II were scaled prior to the transformation; data in column 1 were multiplied by 10,000 and data in column 2 were multiplied by 10.[8]

$$\begin{matrix} & (0.5618) & (2.2530) \\ \hat{Y} = & 14.9268 & -\ 8.7996X \end{matrix} \tag{11}$$

[8] Technically, this corresponds to the model $\ln[10 \times f(W)] = \alpha_0 + [\alpha_1/\ln(W \times 10{,}000)]$. Thus, letting $y = \ln[10 \times f(W)]$ and $x = 1/\ln(W \times 10{,}000)$, we have $y = \alpha_0 + \alpha_1 x$.

With respect to our model, the regression analysis provided the following:

$$E[\ln(\hat{10V_w})] = \underset{(0.5618)}{14.9268} - \underset{(2.2530)}{8.7996}/\ln(W \times 10{,}000), \quad W > 0.0002 \quad (12)$$

Standard error of the estimate = 0.8211; R^2 = 0.58; F = 15.25, and significance level = 0.002; T statistic for constant term = 26.57; T statistic for coefficient = 3.91; Durbin–Watson test = 3.19; Von Neumann ratio = 3.46; thirteen observations: V = barrels of oil plus oil equivalent of gas per square mile, W = number of drill holes per square mile.

As noted, the above equation describes the barrels of oil plus oil equivalent of gas for 10 mile^2 as a function of the number of drill holes per 10,000 mile^2. It can be restated to describe the barrels of oil plus oil equivalent of gas per square mile as a function of the number of drill holes per square mile:

$$\ln(10) + E[\ln \hat{V}_w] = 14.9268 - 8.7996/[\ln W + \ln(10{,}000)] \quad (13)$$

$$E[\ln \hat{V}_w] = 14.9268 - \ln(10) + 8.7996/[\ln W + \ln(10{,}000)] \quad (14)$$

Therefore,

$$E[\ln \hat{V}_w] = 12.62421 - 8.7996/(\ln W + 9.21034) \quad (15)$$

By exponentiating Eq. (15), we have an expression for $f(\hat{W})$:

$$\begin{aligned}\exp(E[\ln \hat{V}_w)] &= f(\hat{W}) \\ &= 303825.86 \exp[-8.7996/(\ln W + 9.21034)] \quad (16)\end{aligned}$$

D. Estimates by the Model

Estimates of $E[\ln V_w]$ for four drilling densities were made using the regression equation of Eq. (15). These estimates and their associated drilling densities are provided in columns 1 and 2 of Table III.

Let us make the assumption that the error term u is normally distributed having a mean of zero and known variance of σ^2:

$$u \sim N(0,\sigma^2) \quad (17)$$

Given this assumption and the fact that σ^2 is not known but must be estimated from the sample data, the $(1 - \alpha)$ confidence interval for the expected value of $\ln V$ for a given drilling density is defined as follows:

$$E[\ln \hat{V}_w] - S_1 \cdot t_{\alpha/2}^{(n-2)} \leq E[\ln V_w] \leq E[\ln \hat{V}_w] + S_1 \cdot t_{\alpha/2}^{(n-2)} \quad (18)$$

TABLE III

Predicted OE and Oil Densities (bbl/mile²)

Drilling density (W) (wells/mile²)	Predicted values		90% Confidence intervals		Lowest expected value of oil density for selected confidence[c]		
	OE[a]	Oil[b]	$\exp(E[\ln V_w])$ (OE)	$0.564 \exp(E[\ln V_w])$ (oil)	95%	90%	50%
1.0	116,774	65,861	62,982–216,858	35,522–122,308	35,522	41,233	65,861
1.5	121,668	68,621	64,670–228,902	36,474–129,101	36,474	42,478	68,621
2.0	124,950	70,472	65,826–237,179	37,126–133,769	37,126	43,329	70,472
10.0	141,477	79,793	71,381–280,405	40,259–158,148	40,259	47,481	79,793

[a] $\exp[12.62421 - 8.7996/(\ln W - 9.21034)] = \exp(E[\ln \hat{V}_w])$.

[b] $0.564 \exp[12.62421 - 8.7996/(\ln W - 9.21034)] = 0.564 \exp(E[\ln \hat{V}_w])$.

[c] These are based upon Jensen's inequality: $a \leq E[\ln V_w] \leq \ln(E[V_w])$ with confidence of $1 - \alpha$. Thus, $e^a \leq \exp(E[\ln V_w]) \leq E[V_w]$. Here, $a = 12.62421 - 8.7996/(\ln W - 9.21034) - S_1 \cdot t_{\alpha/2}^{(n-2)}$.

where

$S_1 =$

$$S\left\{\frac{1}{n} + \frac{\left[\frac{1}{\ln(W \times 10{,}000)} - \frac{1}{\overline{\ln(W \times 10{,}000)}}\right]^2}{\sum_{i=1}^{n}\left(\frac{1}{\ln(W_i \times 10{,}000)}\right)^2 - \left(\sum_{i=1}^{n}\frac{1}{\ln(W_i \times 10{,}000)}\right)^2 \Big/ n}\right\}^{1/2} \tag{19}$$

and S is the standard error of the regression estimate $E[\ln \hat{V}]$. For example, let us compute the 90% confidence interval for the expected OE density, given a drilling density of 2.0 wells/mile^2. First, we compute $E[\ln V_w]$ and S_1:

$$E[\ln \hat{V}_{w=2}] = 12.62421 - \frac{8.7996}{\ln(2.0) + 9.21034} = 11.73567 \tag{20}$$

$$S_1 = 0.8211\left\{\frac{1}{13} + \frac{[(\ln(2.0) + 9.21034)^{-1} - 0.2279]^2}{0.14387}\right\}^{1/2}$$

$$= 0.35687 \tag{21}$$

Next, we consult a table of t values for 11 degrees of freedom and a probability level of 0.05 and find the t value of 1.796. Then, we have:

$$11.73567 - (0.35687)(1.796) \leq E[\ln V_{2.0}] \leq 11.73567 + (0.35687)(1.796) \tag{22}$$

Alternatively,

$$11.73567 - 0.6409 \leq E[\ln V_{2.0}] \leq 11.73567 + 0.6409 \tag{23}$$

$$11.09477 \leq E[\ln V_{2.0}] \leq 12.37657 \tag{24}$$

Let us now write Eq. (18) in a different form:

$$a + u \leq E[\ln V_w] + u \leq b + u, \tag{25}$$

where $a = E[\ln \hat{V}_w] - S_w \cdot t_{\alpha/2}^{(n-2)}$, and $b = E[\ln \hat{V}_w] + S_w \cdot t_{\alpha/2}^{(n-2)}$.

Then,

$$e^a e^u \leq e^{E[\ln V_w]} \cdot e^u \leq e^b e^u$$

Dividing through by e^u, we have the following:

$$e^a \leq e^{E[\ln V_w]} \leq e^b \tag{26}$$

In other words, by exponentiating the confidence limits for $E[\ln V_w]$, we have confidence limits for $\exp(E[\ln V_w])$. It must be noted that this is not a confidence interval for $\ln V_w$ or for $E[V_w]$. If we were to attempt to interpret the meaning of $\exp(E[\ln V_w]$, we might characterize it as a geometric mean value of V given W, or in the case of a log-normally distributed V, it is like a modal value of V given W.

The 90% confidence interval for $\exp(E[\ln V_w])$, given $W = 2.0$, is as follows:

$$e^{11.09477} \leq e^{E[\ln V_w]} \leq e^{12.37657} \quad (27)$$

or

$$65{,}825.99 \leq e^{E[\ln V_w]} \leq 237{,}178.90 \quad (28)$$

In a similar fashion, confidence intervals for $\exp(E[\ln V_w])$ for three additional densities were computed; these are provided in columns 4 and 5 of Table III. If we are content with the assumption that V is log-normally distributed, an interpretation of the ranges in column 4 is that we are 90% confident that the most likely value of V, given W, is within these bounds.

From Jensen's[9] inequality, we have that $E[\ln V_w] \leq \ln(E[V_w])$. Thus, $\exp a \leq \exp(E[\ln V_w]) \leq E[V_w]$. Using this inequality, we can say that $65{,}825.99 \leq E[V_{2.0}]$ with confidence of $1 - \alpha$. Table III shows this lower limit for each of the four drilling densities and for three different levels of confidence.

The statements based upon Jensen's inequality are the strongest that we can make directly about the average value of V. This regression model does not allow us to compute a confidence interval for $E[V_w]$. However, it is possible, using this model, to compute confidence intervals for V_w. While this is what we will eventually do, we do not employ this model, for reasons explained in the following section.

III. SOME PROBLEMS IN INFERENCE BY THE INITIAL MODEL

Concerning the present model, two observations are noteworthy:

- The regression model adopted cannot provide a probability statement for the expected value of V, given W.

---

[9] See Parzen (1960), p. 434.

- The expected value of V, even if it could be obtained, may not be the appropriate statement.

These observations are meaningful when one considers two things: (1) The objective of the analysis is probability statements and estimates of the recoverable OE of the entire ROW region, and (2) the predicted value of the model is an OE density.

These two facts considered jointly create a special problem not encountered routinely in the application of regression analysis. Even if our present model described an expected value of V, we would need to consider the following question: When the quantity of OE in a region is determined by multiplying a predicted OE density by the region's area, are the confidence intervals appropriate regardless of the size of the area? In other words, is the variance of the estimate independent of area size? Intuition suggests that the width of the confidence interval for the OE density should decrease as the size of the region to which it is applied increases. We can supplement this intuition by the experience of mining engineers in the estimation of average grades of mining blocks: DeWijs (1953), exploring the variance–volume relationship for average grades in the Pulacayo mine, found that, given the volume of the entire vein X, the variance of the average grades of blocks having volume x, varied according to the following law:

$$\sigma_x^2 = \alpha \ln(X/x) \tag{29}$$

where α, a constant for a particular deposit, reflects the physical circumstances of the mineralization.

Thus, DeWijs found that as the size of the mining block increased relative to the size of the population, the variance of average grade decreased. While the analogy of the estimation of average grades of mining blocks to the estimation of OE density may not be precise, there are some parallels. The basic datum of the engineer is a drill core with one or more ore intersections. From this datum, he computes an average grade for that core sample. There is an analogy to our present problem of the estimation of OE density if one considers the basic datum to be a set of wells for a region. From this datum, we compute an average OE density for the region. As the engineer would utilize one or more of his drill core grades to compute the average grade of a mining block, we would utilize the OE densities of one or more regions to estimate the density of a collection of regions. There is no point in carrying the analogy further; it may already be severely strained when one considers the particulars of the two situations. Even so, it is helpful in mentally exploring the issue of area size and variance of a density measure.

If there is a variance–area relationship, then we must consider the confidence statements arising from this regression analysis to be relevant only for the inference to OE density for the average area of the 13 regions. Thus, even if the model described the expected value of V, instead of the expected value of ln V, the confidence interval for the expected value of V based upon an unadjusted variance would not be appropriate for the entire ROW region, for its area is approximately 10 times that of the average of the areas of the 13 regions.

Finally, the expected value of V_w may not be the most relevant measure in this study. Since we have only one ROW region and it constitutes our sample, it is more relevant to seek a confidence interval for V, given W, with allowance made for the size of the ROW region, than to seek a confidence interval for $E[V_w]$. This involves both the error of the regression estimate and the distribution of V around this estimate.

Confidence intervals for V, given W, could be made using the regression model just described. But, if the variance of the expected value of V_w should vary with the size of the area, then so should the variance of V_w, for V_w (like $E[V_w]$) is a density. Consequently the regression model just described is not appropriate.

All of the foregoing issues dictate that we need a more complex regression model, one with a built-in variance–area relationship. The following section develops and estimates such a model for the ROW region.

IV. A MORE COMPLEX MODEL

A. Model Form

This model, like the one of the previous section, is a regression model that relates drilling density to expected OE density. However, it differs from the previous model in that the variance to be used in establishing a confidence interval varies inversely with the square of the logarithm of the size of the region. Specifically, we desire a regres-

$$(E[\hat{Y}_x] - Y_x)/(S\sqrt{1 + X_0'(X'X)^{-1}X_0}) = (E[\hat{Y}_x] - Y_x)/S_2 \quad (34)$$

where X is the data matrix for the independent variables. Therefore, a $1 - \alpha$ confidence interval for Y_x can be written as follows:

$$E[\hat{Y}_x] - S_2 t_{\alpha/2}^{n-2} \le Y_x \le E[\hat{Y}_x] + S_2 t_{\alpha/2}^{n-2} \quad (35)$$

Substituting $\ln A \cdot \ln V_w$ for Y_x, we have

$$E[\hat{Y}_x] - S_2 t_{\alpha/2}^{n-2} \le \ln A \cdot \ln V_w \le E[\hat{Y}_x] + S_2 t_{\alpha/2}^{n-2} \quad (36)$$

Dividing through by $\ln A$ and exponentiating we have

$$\exp\left(\frac{E[\hat{Y}_x]}{\ln A} - \frac{S_2}{\ln A} t_{\alpha/2}^{n-2}\right) \le V_w \le \exp\left(\frac{E[\hat{Y}_x]}{\ln A} + \frac{S_2}{\ln A} t_{\alpha/2}^{n-2}\right) \quad (37)$$

where $E[\hat{Y}_x]$ is the value of the regression equation at specified values of X_1 and X_2. If we substitute the equation for $E[\hat{Y}_x]$ and divide through by $\ln A$, we have the following statement of the $1 - \alpha$ confidence interval for V_w:

$$\exp\left[12.60719 - \frac{8.7709753}{\ln(W \times 10{,}000)} - \frac{S_2}{\ln A}(t_{\alpha/2}^{n-2})\right]$$

$$\le V_w \le \exp\left[12.60719 - \frac{8.7709753}{\ln(W \times 10{,}000)} + \frac{S_2}{\ln A}(t_{\alpha/2}^{n-2})\right] \quad (38)$$

Note that Eq. (38) involves $\ln A$ only as a divisor of S_2, which in effect is an adjustment of a variance for area.

C. Some Estimates

Let us use Eq. (38) to compute the 90% confidence interval for recoverable OE density, given a drilling density of 2 wells/mile^2, in the ROW region:

$$E[\hat{Y}_x]/\ln A = 12.60719 - \frac{8.7709753}{\ln(2.0 \times 10{,}000)} = 11.7215462 \quad (39)$$

$$S_2 = S(1 + X_0'(X'X)^{-1}X)^{1/2} = 13.50505 \quad (40)$$

Thus,

$$\exp\left[11.7215462 - \left(\frac{13.50505}{17.09748}\right)(1.796)\right]$$

$$\le V_{w=2.0} \le \exp\left[11.7215462 + \left(\frac{13.50505}{17.09748}\right)(1.796)\right] \quad (41)$$

This evaluates to an interval of [29819.5–508986.2]:

$$29{,}819.5 \text{ bbl/mile}^2 \leq V_{w=2.0} \leq 508{,}986.2 \text{ bbl/mile}^2 \tag{42}$$

Multiplying these densities by 26,628,200, the prospective area of the ROW region, we have the following 90% confidence interval H for recoverable OE, given a drilling density of 2 wells/mile^2:

$$940 \times 10^9 \text{ bbl} \leq H_{w=2.0} \leq 14{,}136 \times 10^9 \text{ bbl} \tag{43}$$

Taking the fraction of recoverable OE comprised by oil to be a known constant, 0.564, Eq. (43) implies an interval O for recoverable oil, given $W = 2.0$:

$$429.4 \times 10^9 \text{ bbl} \leq O_{w=2.0} \leq 7940.4 \times 10^9 \text{ bbl} \tag{44}$$

Both of these intervals are very broad, reflecting the simplicity of the model and the fact that all factors other than area and drilling density are treated as a random influence. It is noteworthy that the interval for oil is even broader than the range provided by Grossling for the entire prospective area of the world.

Point estimates for H and O from our regression model, given $W =$ 2.0, for the entire ROW region are

$$\begin{aligned} \hat{H}_{w=2.0} &= 3280.5 \times 10^9 \text{ bbl} \\ \hat{O}_{w=2.0} &= 1850 \times 10^9 \text{ bbl} \end{aligned} \tag{45}$$

These point estimates lie much closer to the lower bounds of the interval for $H_{w=2.0}$ and $O_{w=2.0}$ than to the higher bounds. This is due to the nature of the model, namely, the postulate of a multiplicative relationship. Or, looked at differently, it is due to the use of a log–log model. The intervals for ln V_w are symmetric, but they are not for V_w.

V. A CRITICAL EXAMINATION OF THE MORE COMPLEX REGRESSION MODEL

A. The Credibility Issue

At this point, we have some resource estimates. The relevant question must be: How credible are these estimates? After all, the numbers reflect directly several assumptions; one of these is the assumption that the selected form of the model is appropriate. Most certainly there are other possible forms, and a different specification would have provided different numbers. Furthermore, it is impossible to prove that a particular form is the correct one. The most that can be done to investigate the credibility of the model is to (1) examine the degree to which the other assumptions which were made to estimate the model are met; (2) examine the statistical significance of the equation and its parameters; (3) explore the effect of sample size and reaggregation; and (4) examine the accuracy and representativeness of the data. The perspective offered earlier was that the quantity of recoverable OE is a reflection of a host of important influences, e.g., geology and economics. All influences except drilling densities are represented by a random error term. Thus, even under the best of circumstances, the most that we can expect is that an estimate of OE density is unbiased and that it is a minimum variance estimate. Additionally, we can expect the probability statements to be accurate only if the deviations (residuals) of the model estimates from the sample data meet the following conditions: Their sum is zero; they exhibit no autocorrelation (trends due to influences of neglected variables or improper specification); and they are normally distributed. The deviations are provided in Appendix B.

B. The Assumptions about the Deviations

The first of these requirements of the deviations is met by virtue of the manner in which the model was fitted to the data. The presence of positive or negative serial correlation in the deviations can be tested for using the Durbin–Watson statistic, which for this model is 3.16. For each combination of sample circumstances (number of explanatory variables, sample size, and level of significance) there are two test values: a lower one D_L, and a higher one D_u. If the Durbin–Watson statistic is below D_L, we reject the null hypothesis (randomly distributed errors). If it is greater than D_u, we cannot reject the null hypothesis; the residuals are considered to be randomly distributed. Unfortunately, the Durbin–Watson table does not extend to sample sizes less than 15; consequently, we cannot rigorously test for the presence of autocorrelation. But, for a sample size of 15 and a significance level of 0.05 the upper value is 1.54. Since table values increase with the size of the sample, the Durbin–Watson statistic for this model, 3.16, strongly suggests that the residuals are not correlated. To test for negative correlations use $4.00 - 3.16 = 0.84$; here the test is inconclusive.

The remaining assumption to be examined is that the deviations are normally distributed. Because of the small sample size, $n = 13$, the chisquare test for normality of the deviations is not useful. Instead, the test for normality derived by Shapiro and Wilk (1965) is employed. A brief description of this test for normality and its application to this study are provided in Appendix C. Simply stated, this test compares the W statistic, explained in the Appendix, to test values selected from a table provided by Shapiro and Wilk (1965). For a sample of size 13, the 1 and 99 percentile test values are 0.814 and 0.986, respectively. A low value for the W statistic indicates nonnormality. Thus, if the W statistic for a sample of size 13 were 0.813, we could reject the null hypothesis of normally distributed residuals at the 1% significance level, meaning that there is a probability of 0.01 of incorrectly rejecting the hypothesis of normality. The W statistic for this study is 12.44. Even at the 99% significance level, we cannot reject the hypothesis of normality. Thus, this large value strongly indicates normally distributed deviations.

In summary, though we cannot prove that the specified form of the model is the best of possible forms, it is of such form that the assumptions upon which estimation and probability relations are based are satisfied.

C. Statistical Significance of the Model

The usual way to test how well the linear model fits is to test the null hypothesis that the coefficients of the independent variables are all zero. In order to do this, the sum of squares of residuals is partitioned to determine the extent to which the model explains the variation present in the data. There are three different sets of residuals that could be considered.

The linear model actually fitted

$$\hat{Y} = B_1X_1 + B_2X_2$$

can be tested by computing

$$R^2 = 1 - \left\{\sum_{i=1}^{13} (\hat{Y}_i - Y_i)^2 \Big/ \left[\sum_{i=1}^{13} Y_i^2 - \tfrac{1}{13}\left(\sum_{i=1}^{13} Y_i\right)^2\right]\right\} = 0.99$$

Since this model was obtained by adjusting for the nonconstant variance, this R^2 is such that $R^2(10)/(1 - R^2)2$ is F distributed. With a computed value of 495 the null hypothesis will be rejected for $\alpha = 0.001$. The linear model is then considered to explain the variance rather well.

However, the factor $\ln A$ appears in all three variables and it would seem more satisfactory to consider residuals of the form $(\hat{Y}_i - Y_i)/\ln A_i$ and compute a corresponding R_1^2 where

$$R_1^2 = 1 - \left\{ \frac{\sum_{i=1}^{13} (\hat{Y}_i/\ln A_i - Y_i/\ln A_i)^2}{\sum_{i=1}^{13} (Y_i/\ln A_i)^2 - \left[\frac{1}{13}\sum_{i=1}^{13} (Y_i/\ln A_i)\right]^2} \right\}$$

The computed value of R_1^2 is 0.54. This would correspond to the linear model

$$\ln V = B_1 + B_2[\ln(W \times 10^4)]^{-1}$$

Unfortunately, the constant variance assumption is no longer applicable and hence $R_1^2(11)/(1 - R_1^2)$ is not F distributed, thus the computed value 12.9 cannot be related to table values. R_1^2 is derived from a partitioning of the total sum of squares and is indicative of how much of the variance is explained by the model for $\ln V$ in terms of the variable $1/\ln(W \times 10^4)$. Since now only drilling density is incorporated and not area, nor any other factors affecting the production of OE, it is not surprising that only 54% of the variance is explained by the model.

In order to use $R_1^2(11)/(1 - R_1^2)$ as a test statistic it would also be necessary to show that the modified residuals still satisfy the normality and lack of autocorrelation assumptions. Because of the nonconstant variance these tests were omitted.

Finally, we could consider the original variables and thus the residuals for V. Because of the nonlinear transformation this does not seem useful since the residuals would have to be considered as multiplicative rather than additive; consequently, this sum of squares was not computed.

We note finally that the tests for normality of residuals and the lack of serial correlation were applied to the residuals for Y, since both tests require the constant variance assumption. In conclusion, we see that it is only possible to test the model actually fitted, $\hat{Y} = B_1X_1 + B_2X_2$.

D. Sample Size and Reaggregation

The fitting of a linear model with such a small sample size could result in its being very unstable, i.e., sensitive to slight changes in the values of the variables or changes obtained by combining the regions in different ways. We have already indicated that the use of drilling density and OE density is partly motivated by the objective of avoiding aggregation vagaries. To see how well the model minimizes the effects of reaggregation, we consider three cases of artificial data, as shown in Table IV (these were suggested by Z. Wurtele of Pan Heuristics).

Case II and III are obtained from I by reaggregating, in particular, reaggregating a rich region with a poor region and one for which the drilling density produces quite different data. One would expect that such an unrealistic reaggregation would seriously distort the relationship between V and W and that this effect would be reflected in the correlation coefficient, as suggested by the changed correlations of Case I, II, and III. The disparity is not as great as it might appear; the confidence intervals which are obtained are wide since the sample size is so small. This disparity might then be discounted for several reasons, but another way to judge the disparity is to compute the corresponding values of Y and X_2 and recompute the correlation coefficients using the

TABLE IV

Artificial Data—Aggregated and Reaggregated[a]

	Region	A, Area	V, drilling density	W, OE density
Case I	1	10	100	1000
	2a	50	100	3000
	2b	50	300	1000
	3a	500	200	4000
	3b	500	400	2000
	The correlation coefficient of V, W is -0.176			
Case II	1	10	1000	100
	2a + 2b	100	2000	200
	3a + 3b	1000	3000	300
	The sample correlation is $+1$			
Case III	1+2a	60	2760	100
	2b	50	1000	300
	3a	500	4000	200
	3b	500	2000	400
	The correlation coefficient is -0.514			

[a] Source: Zivia Wurtele, Pan Heuristics.

model employed in this study. In all three cases, the correlation coefficient is greater than 0.99, which indicates that the transformed values are less sensitive to the reaggregation; consequently, the small sample size is not as great a concern.

In the example considered (Cases I, II, III) no assumptions were made about whether the regions aggregated were contiguous or had common geological characteristics or other relevant properties. Obviously, any realistic reaggregation would be somewhat constrained by these factors, further softening the aggregation problem of real data.

With respect to the data (13 regions) and the model used, it is true that it does represent aggregation, but there are several reasons why such aggregating is unlikely to impair the validity of the results; they include the following:

1. The aggregation has taken place with respect to contiguous regions for which there are some similarities in overall geology, state of technology, and extent of exploration.
2. Smaller sample size will result in a wider confidence interval and hence the uncertainty exhibited by smaller sample size is reflected in the conclusions.
3. The empirical evidence is that the contrived examples do not result in distortion in the model.
4. Relatively speaking $n = 13$ is considerably larger than $n = 5$ (Case I).

If data were available for subregions it would certainly be preferable to fit the model on unaggregated data. Another way to minimize disparities in the level of exploration in the thirteen regions would be to convert drilling density into standard hole depths.

E. Accuracy and Representativeness of Data

The third issue in credibility of the estimates of the model is the accuracy of the basic data upon which the model was specified and quantified. We have little to offer by way of criticizing the accuracy of the data in an absolute sense. Clearly, we have taken as a departure point the basic data provided by Grossling (1976)—Table II. The issue of data accuracy and representativeness is too complex to consider here to the degree that it truly warrants. Such an investigation will require the reader to examine Grossling's work and the works of those whom he cites. A brief and general comment on these issues was made in Section I,E.

VI. ESTIMATES FOR THE ENTIRE WORLD

A. Information on the Middle East Region

The Middle East was excluded from the statistical analysis because of its anomalous nature. The objective of this section is to combine the results of the statistical analysis on the ROW region with independently made estimates of recoverable OE of the Middle East region.

The probabilistic description of recoverable OE of the Middle East is derived from the study of Nehring (1978) on giant oil fields and world oil resources. Nehring provides a range for the recoverable oil resources of the Middle East region: 860–1140 billion barrels. Subtracting from these limits the estimate of EVRD by Grossling (1976) for this region, we have a range for the recoverable but undiscovered oil resources of the Middle East: 212–492 billion barrels.

Grossling also provides an EVRD for the natural gas of the Middle East: 1133.2 trillion cubic feet. If we assume that the ratio of gas to oil for the recoverable undiscovered OE is the same as it is for the EVRD's, we can describe a range for recoverable undiscovered gas of $(371\text{–}861) \times 10^{12}\ \text{ft}^3$:

$$(212\text{–}492) \times 10^9\ \text{bbl} \left(\frac{1133.2 \times 10^{12}\ \text{ft}^3}{647.48 \times 10^9\ \text{bbl}}\right) \tag{46}$$

or $(371\text{–}861) \times 10^{12}\ \text{ft}^3$.

Based upon this means of inference, total recoverable gas (EVRD + undiscovered) is estimated to be within the range of $(1504.2\text{–}1994.2) \times 10^{12}\ \text{ft}^3$. Using the conversion factor of 4440 ft^3/bbl, these quantities of gas are converted to oil equivalents: $(338.7\text{–}449.1) \times 10^9$ bbl. Adding these quantities of oil equivalent to Nehring's range for recoverable oil, we have a range for recoverable OE for the Middle East region of $(1199\text{–}1589) \times 10^9$ bbl: [860 + 338.7] to [1140 + 449.1].

B. A Preliminary Estimate of World OE

In a previous section, the recoverable OE of the ROW region for a drilling density of 2.0 wells/mile^2 was estimated to be 3280.5×10^9 bbl.

Further, the number was described as a most likely estimate for that region. In the preceding section, it was suggested that the recoverable OE of the Middle East region falls within the range of $(1199–1589) \times 10^9$ bbl. The midpoint of this range is 1394×10^9. If this number is added to the 3280.5×10^9, we have as an estimate of the world's recoverable OE of 4674.5×10^9 bbl. Taking the oil component to be 56.4% of this quantity, an estimate of the recoverable oil resources of the world is 2387×10^9 bbl. Even though this number may result from "reasonable" approaches, it is a mixed number from a statistical point of view, for it is the sum of a median and a mode.

Obtaining an expected value or a modal value for the aggregate of the two regions requires more complex analysis, which is the subject of the following section.

C. Monte Carlo Analysis

The approach taken in this study to estimate expected and most likely values for the recoverable OE of the world is to consider the OE of the world H to be the sum of two random variables H_1 and H_2:

$$H = H_1 + H_2 \tag{47}$$

H_1 and H_2 are the magnitudes of the OE of the Middle East and ROW regions, respectively. Consistent with the range derived in the previous section, the probability density function for H_1 is considered to be rectangular:

$$f(h_1) = \begin{cases} 1/390 \times 10^9, & 1199 \times 10^9 \le h_1 \le 1589 \times 10^9 \\ 0, & \text{otherwise} \end{cases} \tag{48}$$

Consistent with the regression model, H_2 is considered a conditional random variable, conditional upon the drilling density postulated for the ROW region.

The procedure for obtaining the distribution for H was to sample at random from the rectangular distribution of Eq. (48) so as to obtain a value for H_1 and at the same time to draw at random a value from the t distribution to be used with the regression equation so as to generate a value for H_2, given a prespecified value for W. These quantities were added to give the size of the OE of the world.

Technically, let $F(h_1) = \int_{1199\times 10^9}^{h_1} f(h_1)\, dh_1$. A quantity of OE for the Middle East was obtained by drawing a number r_1 on the interval of [0, 1] at random and then evaluating $F^{-1}(r_1) = h_1$. Recoverable OE of

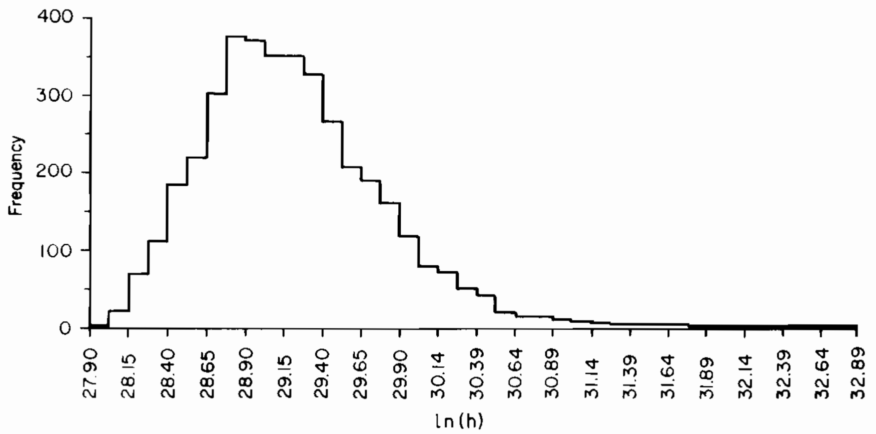

Fig. 5. Histogram of world OE resources (H), given a drilling density of 1.0 well/mile^2.

the ROW region were obtained by selecting another random number r_2 and evaluating $T^{-1}(r_2) = t_{r_2}^{n-2}$, where T is the distribution function for the t statistic. Then, for a given W, h_2 was obtained by substituting $t_{r_2}^{n-2}$ into the following equation:

$$h_2 = 26{,}628{,}200 \exp\left[12.60719 - \frac{8.7709753}{\ln(W \times 10{,}000)} + \frac{S_2}{17.09748} t_{r_2}^{(n-2)}\right]$$

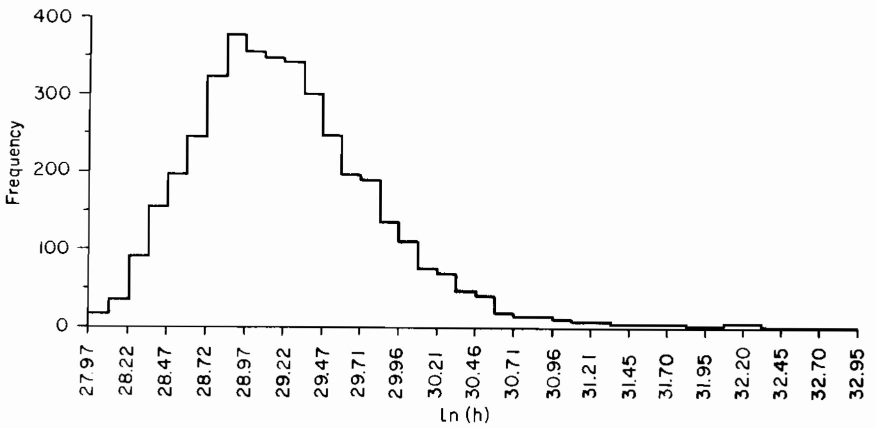

Fig. 6. Histogram of world OE resources (H), given a drilling density of 1.5 wells/mile^2.

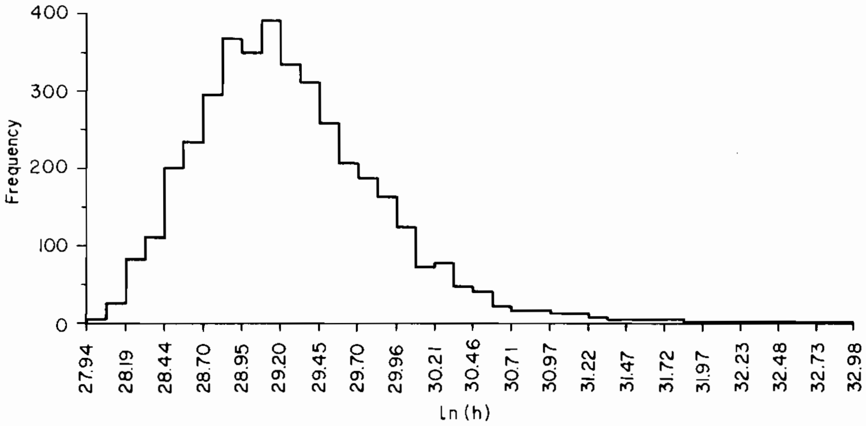

Fig. 7. Histogram of world OE resources (H), given a drilling density of 2.0 wells/mile^2.

These quantities, h_1 and h_2, were added to give h. This procedure was repeated 4000 times, giving 4000 values for h. Statistical processing of these 4000 values provided a mean value which is an estimate of the mean value of the population, and a histogram for H, which is an approximation to the probability distribution for H. From this histogram, an estimate of the most likely value of H was obtained. The histograms for five drilling densities are provided in Figs. 5–9. These

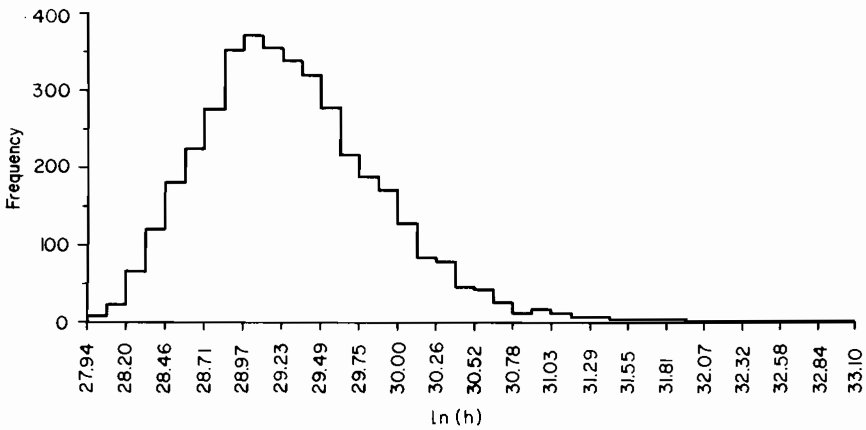

Fig. 8. Histogram of world OE resources (H), given a drilling density of 5 wells/mile^2.

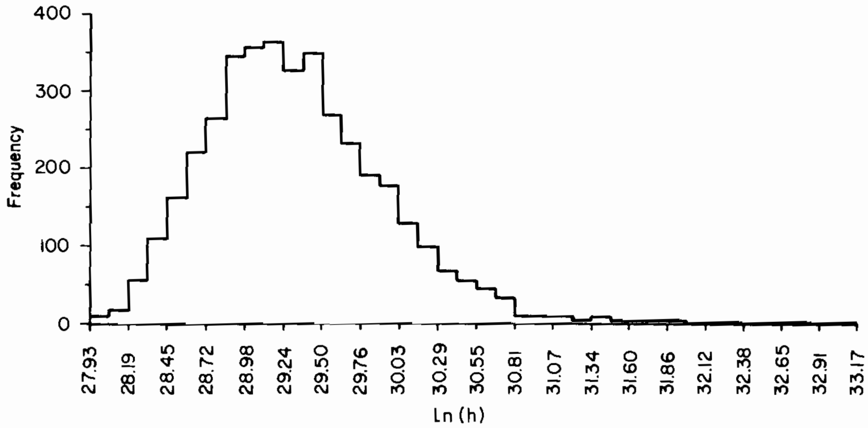

Fig. 9. Histogram of world OE resources (H), given a drilling density of 10 wells/mile^2.

histograms exhibit a skewness even when h is plotted in the logarithmic scale, indicating a high degree of skewness.

D. A Range of Values

For each of the five drilling densities, the range was approximated on the associated histogram which contained 90% of the 4000 cases. These ranges are reported in Table V. The most striking feature of a range for a given drilling density is its breadth. So that we can make comparisons, let us restrict our attention to only one of the ranges; accordingly, let us assume that the economics of exploration drilling allows a density of 2.0 wells/mile^2. This is approximately twice the current well density of the United States. Achieving such a density on the prospective area of the world would require a great amount of drilling. It is here acknowledged that the selection of the appropriate drilling density is critical in characterizing recoverable OE resources, for the cost implied by such a drilling density must be consistent with the economic reference for defining recoverable resources. Such a selection would require a marginal analysis of the cost of drilling as cumulative effort increases and the value product of drilling. While this is most important to place the result of this study in proper perspective, this issue is not taken up in this report.

It is noteworthy that the 90% range for the recoverable oil resources, given a drilling density of 2.0 wells/mile^2, is (1234–8335) $\times$ 10^9, a greater range than even Grossling's range [(1960–5600) $\times$ 10^9 bbl] which is considered by many to be extremely broad. Furthermore, the expected value of world oil resources (3558 $\times$ 10^9 bbl) for a drilling density of 2.0 wells/mile^2 is approximately 50% greater than the estimate by Hendricks (2480 $\times$ 10^9 bbl), the largest of the single point estimates (see Table I). These results seem to diverge from the statements of the experts. This divergence is eased only slightly by selecting a lower drilling density, but even at a drilling density of 1.0 (the current density of the United States) the 90% range is still very broad and the expected value of oil resources is large (3439 $\times$ 10^9 bbl) compared to existing estimates. Do these results challenge the credibility of this model and its estimates? Perhaps not, when some important facts are considered. But before considering these facts, let us examine point estimates by the model.

E. Point Estimates of Oil Resources

As previously indicated, the mean value (3641 $\times$ 10^9 bbl) from the histogram for a drilling density of 2.0 wells/mile^2 is approximately 50% greater than the largest of the point estimates provided by experts (2480 $\times$ 10^9 bbl). A judgment of the significance of this result can be

TABLE V

Ranges for World OE and Oil

Drilling density (wells/mile^2)	90% Confidence OE ($\times 10^9$ bbl)	90% Confidence Oil[a] ($\times 10^9$ bbl)	60% Confidence OE ($\times 10^9$ bbl)	60% Confidence Oil[a] ($\times 10^9$ bbl)
1.0	2,140–13,993	1,207–7,892	2,965–7,575	1,672–4,272
1.5	2,149–14,574	1,212–8,220	3,046–7,825	1,718–4,413
2.0	2,188–14,778	1,234–8,335	3,082–8,049	1,738–4,550
5.0	2,236–15,966	1,261–9,005		
10.0	2,276–17,119	1,284–9,655		

[a] Based upon the proposition that oil equals 56.4% of oil plus oil equivalent of gas.

made only after consideration of the nature of the subjective estimate provided by an expert. It is here proposed that when the thing being estimated is distributed according to a skewed distribution, like the log–normal, the geologist provides an estimate which approximates more closely a modal value than a mean value. When one considers that in order to provide a good estimate of a mean value the geologist must weight all possible outcomes by their probabilities, it becomes unlikely that the point estimate provided by him is a good estimator of the mean when the possible states of nature are distributed according to a skewed distribution. A more plausible proposition is that the geologist perceives the most frequently occurring (most likely) event as the estimate of the state of nature, thereby estimating a modal value. If this is true, then the most appropriate comparison of estimates is that of the modal value of the histograms to the subjective estimates of the experts. The modal values for the five drilling densities are provided in Table VI.

These estimates are graphed in Fig. 10. This figure shows an erratic modal value for a drilling density of 2.0 wells/mile^2. This erratic measure is believed to be a reflection of the vagaries of simulated sampling and the arbitrary selection of class intervals for the construction of a histogram. Therefore, the modal value in Table VI for a drilling density of 2.0 wells/mile^2 is an interpreted value, based upon the freehand curve of Fig. 10. It is noteworthy that an order of magnitude change in drilling density results in a change in the most likely oil resources from

TABLE VI

Estimates of Recoverable OE and Oil Resources

Drilling density (wells/mile^2)	Mean value		Modal value	
	OE ($\times 10^9$ bbl)	Oil ($\times 10^9$ bbl)	OE ($\times 10^9$ bbl)	Oil ($\times 10^9$ bbl)
1.0	6,098	3,439	3,367	1,899
1.5	6,309	3,558	3,598	2,029
2.0	6,457	3,641	3,820[a]	2,154[a]
5.0	6,897	3,890	4,350	2,453
10.0	7,208	4,065	4,674	2,636

[a] Interpolated value; see Fig. 10.

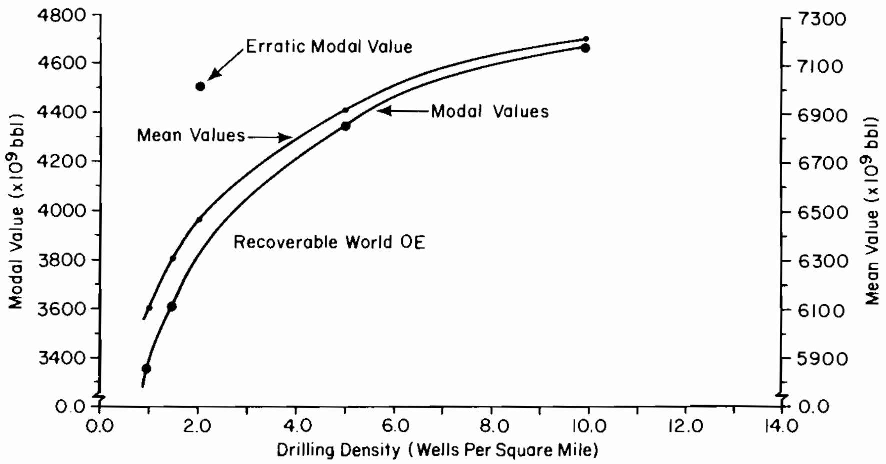

Fig. 10. Modal and mean values of world OE for various drilling densities.

approximately 1900 × 10^9 to 2600 × 10^9 bbl. Note too, that for drilling densities of 1.0 to 2.0 wells/mile^2, the modal values range only from approximately 1900 × 10^9 to 2150 × 10^9 bbl. These quantities compare well with the estimate by the experts, which seems to be converging on a number around 2000 × 10^9 bbl. Note that for drilling densities of 1.5 and 2.0 wells/mile^2 the modal values (2029 × 10^9 and 2154 × 10^9, respectively) fall within the ranges reported by Grossling [(1960–5600) × 10^9], Nehring [(1700–2300) × 10^9] and the World Energy Conference [(1920–2420) × 10^9].

VII. PROBING THE EFFECTS OF ADDITIONAL INFORMATION

A. Perspective

In this study, the effects of geologic differences among the resource subregions of the ROW region were represented by a random term (influence). Consequently, a reasonable commentary on the 90% range of values, which is very broad as compared to subjectively estimated ranges, and the mean value, which is much larger than subjectively made point estimates, is that these divergences reflect the fact that geologic information was ignored—the model employed only drilling results. This section considers such a commentary. Two different approaches are taken. The first one takes as a given the proposition that consideration of geology by the model, had such information been available, would have produced narrower ranges and smaller means for the various drilling densities. The essence of this first approach is that in spite of the acceptance of this proposition, comparisons between the estimates by the model and available geologic estimates must deal with the nature of *subjective* estimates.

The second approach seeks to estimate a log–normal distribution given a favorable evaluation of the contribution of subjective geologic estimation of nonproduced oil resources. Since the most likely value estimated by the statistical model agrees generally with geologic estimates, the geologic issue challenges primarily the estimated variance of oil resources. Consequently, this approach takes from the statistical model only the most likely estimate and makes the assumption that this value is the modal value of a log–normal distribution. The remaining (unknown) parameter of this distribution is estimated by using known resource information and subjective statements about the low range of nonproduced resources. Then this distribution is employed to compute a confidence interval and a mean value, which are compared to strictly geologic estimates and to estimates by the statistical model.

B. Subjective Geologic Resource Appraisal

Everything else being equal, if geologic data and the understanding of resource geology by the geologist contributes any true information beyond the drilling information about OE resource, the range provided by an expert geologist should be narrower than one defined by the model.

There is another factor at issue here which must be considered in comparing the results of this study with previous estimates: the nature of subjective estimates of ranges. Studies in psychometrics have shown that typically subjective estimates of the possible range of uncertain quantities understate considerably the true range. For example, Alpert and Raiffa (1969) asked 800 Harvard MBA students to provide the 0.01, 0.25, 0.50, 0.75, and 0.99 fractile estimates for each of 20 items, such as the number of automobiles imported into the United States in 1967. They found that 41% of the actual values fell outside of the 0.01–0.99 range determined from the responses of 800 students.

The ideal result would have been for 2% of the actual values to fall outside of the range. In other words, as an assessor of probabilities or uncertain quantities, man acts as though he knows more than he really does, which results in underestimating the probabilities for very large and very small values of the variable being considered.

Suppose that on average, geologic analysis provides no new information, given that we already have drilling data. While such a supposition would offend many geologists as a general proposition, the degree of offense may not be great when one considers the world as a whole and the very meager geologic data that are both available and *useful* for estimating oil resources. The contribution of geologic analysis on a world scale is severely hampered in some regions by the lack of geologic information. Suppose also that the histograms provided by the statistical analysis of this study approximate the true probability distributions. Given these suppositions, it is instructive to consider what the observation of Alpert and Raiffa would suggest as the range provided by a set of geologists. If we assume that the probabilities for the extreme values neglected by a subjective appraiser are symmetrical, i.e., 20% of the probability neglected by the experts is for extremely large values and 20% for extremely small values, we can express the observation by Alpert and Raiffa by the proposition that when asked to provide the 98% range, geologists would provide a 60% range. The 60% range for drilling densities of 1.0, 1.5, and 2.0 are provided in Table V. For a drilling density of 2.0 wells/mile^2, the 60% range $[(1738–4550) \times 10^9]$ provides a lower limit which is very close to that of Nehring (1700×10^9). The upper limit (4413×10^9) is between Nehring's and Grossling's estimates of the high bounds:

$$\underset{\text{(Nehring)}}{2300 \times 10^9} \leq 4413 \times 10^9 \leq \underset{\text{(Grossling)}}{5600 \times 10^9}$$

In other words, given what has been observed about subjective processes and the assumption that geology contributes no information about the magnitude of world resources beyond the information present in drilling results, the subjectively estimated ranges of Nehring and Grossling could be viewed as variation among experts in their subjective processes and their reaction to an uncertain world, a world which is accurately described by the model of this study.

Let us examine this from another perspective. Suppose that we treat Nehring and Grossling as a set of experts and compute a range which represents both of them by averaging their high and low bounds: (1830–

3950) $\times 10^9$. Dividing the difference of these two numbers (2120×10^9) by 0.6, we have 3533×10^9 bbl as the adjusted magnitude of the difference. Distributing this increase [$(3533–2120) \times 10^9$] equally below the low bound and above the high bound provides a composite range which has been adjusted for deficiencies of the subjective processes: (1124–4656) $\times 10^9$ bbl. The suggestion which arises from this exercise is that adjustment of subjective geologic estimates for the kinds of deficiencies which have been observed generally in subjective estimates of uncertain quantities yields a considerably broader range than those which have been reported. Adjustment of the composite range produces a lower bound that is close to the lower bound of the statistical model. However, the upper bound is much smaller than that of the statistical model.

A question that is important to consider is the following: Is it possible that if we had several hundred experts, such as Grossling and Nehring, the average of this upper bound—a bound which in this model would be a 99 percentile—would be similar to the upper bound of the 60% range of the histogram? A possible reply to this question is that such an exercise has already been achieved by the Delphi experiment conducted by the World Energy Conference (WEC), and it produced a range of only (1920–2420) $\times 10^9$ bbl. The lower bound of this range is nearly identical to Grossling's lower bound (1960×10^9), but the upper bound is much closer to Nehring's upper bound. Other than noting these results, nothing can be concluded from the result of the WEC survey, for the very purpose of a Delphi experiment is to reduce the range of estimates, not to explore the degree of variation. Unfortunately, the result of a Delphi experiment reflects more the "herding instinct" of man than persuasion based upon reasoning and the rationalization of data (see Sachman, 1974).

Here we appreciate Grossling's wonderment that there could be such accord among experts about a quantity so uncertain as the magnitude of the world's oil resources.

The foregoing discussion was motivated by the observed performance of individuals attempting to appraise uncertain events and the assumption that geology contributes no information beyond that which is present in the drilling data. Let us now relax the assumption about geologic analysis. This is equivalent to saying that given the existence of geologic data and the exercising of analysis, the true range of values for a level of confidence is narrower than that indicated by the statistical analysis. However, things are not quite this unequivocal, for the general proposition by Alpert and Raiffa may still apply. The geologist still is faced with appraising an uncertain quantity, the magnitude of the world's oil resources. While uncertainty may be reduced by the available geologic data and its processing by an expert geologist, there still remains great uncertainty. Consequently, the findings by Alpert and Raiffa and others suggest that the range reported by the expert geologist probably is narrower than it should be. In other words, since geologic analysis is to a considerable degree qualitative, the expert probably fails to consider the many combinations of events which could result in extremely high or extremely low values. Consequently the range reported by the geologist tends to be too narrow. Thus, the task of evaluating estimates is very difficult, for such evaluation must balance two opposing effects:

1. To the extent that geologic data and geologic experts provide actual resource information beyond drilling information, the ranges provided by an expert geologist should be narrower than those provided in this study.
2. Man as an assessor of probabilities or of the associated uncertain events typically underestimates considerably the possible states that could occur. Thus, the qualitative geologic analysis involved in the estimation of oil resources leads to an understatement of the possible range of values.

Unfortunately, it is difficult, if not impossible, to judge the net effect of these opposing tendencies. Furthermore, when evaluating an estimate by an expert, one must deal with the possibility that this expert does not suffer from all of the frailties of subjective processes to the same degree as does the average appraiser. In other words, even though typically subjective appraisers considerably underestimate the probabilities for extreme values, a particular expert's performance may be better than the typical performance.

C. Examination of Subjective Estimates through a Log–Normal Model

This section takes an approach quite different from that of the previous section; namely, limitations of subjective processes are ignored and the assumption is made that geologic data and subjective evaluation produce real information about the magnitude of world oil resources. In essence, this section explores what the probability distribution of world oil resources might be given these best circumstances.

The approach taken here is based upon four propositions:

- Undiscovered world oil resources are log-normally distributed.
- The effect of the additional information on the probability distribution would be a reduction in the range of possible states of nature; the most likely state would not be altered.
- The most likely state of nature implies a drilling density of 2 wells/mile^2, and this state is correctly estimated by the foregoing statistical analysis of this paper: 2154×10^9 bbl of oil.
- The most accurate subjective perceptions of oil resources are the lower bounds of possible states.

The effect of these propositions is to give us a distribution form and one point on the distribution. With these fixes, we then are motivated to examine our resource information to see if we can constrain the distribution sufficiently to estimate its parameters.

Table VII summarizes Nehring's perception of the magnitudes of various categories of world resources. Clearly, the magnitude of world recoverable resources cannot be less than cumulative production, 335×10^9 bbl. So a probability distribution for the magnitude of world oil resources which reflects our *current* knowledge must exhibit a probability of zero for quantities less than 335×10^9 bbl.

Do we have any more useful information? Again using Nehring's data (Table VII), we have the sum of cumulative production, reserves, and additional recovery. Compared to the uncertainty about the magni-

TABLE VII

Nehring's Description of World Resources[a]

Resource	Oil ($\times 10^9$ bbl)
Cumulative production	335
Proved and probable reserves	675
Additional recovery	420–730
Future discoveries	270–560
Total	1700–2300

[a] Source: Richard Nehring (1978).

tude of future discoveries, this sum is well known. But, in fact, there is considerable uncertainty about this sum. Proven reserves are required to be accurate only within ±20%. The accuracy of proven reserve estimates of some fields probably exceeds this requirement considerably. Nevertheless, it is difficult to know what accuracy to ascribe to the total category of proven plus probable reserves. Of course, estimates of additional recovery are even more tenuous. Unfortunately, while the sum looks like useful information, in order to use it to provide additional constraints on the probability distribution for recoverable oil resources, we must incorporate in our probability statement regarding the quantity a consideration of the contribution of future discoveries. In other words, we cannot state that there is a probability of 0.20 that the world oil resources are less than this quantity, because this downside risk does not consider the possible contribution of future discoveries. Therefore we cannot use this quantity directly. Only by adding estimated future discoveries do we have a consistent quantity.

Adding the lower bound of Nehring's estimate gives us the 1700×10^9 bbl which he reports as the lower bound of his range for oil resources. Suppose that we had a credible statement about the probability that the magnitude of world recoverable oil resources is less than 1700×10^9 bbl. Basically, this is the probability that the magnitude of nonproduced recoverable oil resources, O', is less than the difference of $(1700–335) \times 10^9$. We could then combine this information with the absolute barrier of 335×10^9 bbl and the most likely quantity of 2146×10^9 bbl and use them to estimate the parameters of a log–normal probability distribution. For example, let us assume that this probability is 0.05. For the purpose of estimating the implied log–normal distribution, we could summarize our information as follows:

1) $\mu = \ln(2146 \times 10^9 - 335 \times 10^9)$

2) $P[O' \leq (1700 \times 10^9 - 335 \times 10^9)] = 0.05$

The standardized normal variate is defined as follows:

$$Z = (\ln O' - \mu)/\sigma$$

where O' represents nonproduced recoverable oil resources.

Therefore σ, our unknown parameter, can be written in terms of the other quantities:

$$\sigma = (\ln O' - \mu)/Z$$

By employing the three pieces of information noted above, we can write

$$\sigma = [\ln(1700 \times 10^9 - 335 \times 10^9) - \ln(2154 \times 10^9 - 335 \times 10^9)]/(-1.645) = 0.175$$

Thus, we have defined both parameters: $\mu = 28.2293$ and $\sigma = 0.175$. Using these parameters and the quantity of cumulative production, 335×10^9 bbl, we can compute the 95% confidence interval for O:

$$1700 \times 10^9 \leq O \leq 335 \times 10^9 + \exp[28.2293 + 0.175(1.645)]$$

$$1700 \times 10^9 \leq O \leq 2761 \times 10^9$$

Additionally, we can estimate the mean value:

$$\bar{O} = 335 \times 10^9 + \exp(\mu + \sigma^2/2)$$

$$\bar{O} = 335 \times 10^9 + \exp[28.2293 + (0.175)^2/2.0)] = 2182 \times 10^9$$

Of course, a critical assumption was necessary for this exercise; namely, $P[O' \leq (1700 \times 10^9 - 335 \times 10^9)] = 0.05$. Unfortunately we do not have a fix on this number; consequently, these same computations were made for possible magnitudes of this probability; the results are provided in Table VIII.

Another assumption which was made in order to complete the exercise is that a two-parameter log–normal distribution is the appropriate one. There could indeed be other acceptable distributions, but the selection of the log–normal was not arbitrary. Many statistical studies of oil field size have found that the log–normal distribution fits the statistical data well (Barouch and Kaufman, 1977). Furthermore, the log–normal distribution has been employed as a model for the synthesis of subjective geologic statements about recoverable United States oil resources (Miller *et al.*, 1975). The use of the log–normal distribution in this exercise departs somewhat from the uses just cited. Specifically, the use here requires the assumption that nonproduced recoverable oil resources are distributed log–normally.

This exercise yields much narrower confidence bounds than those of the statistical model, given drilling densities of 1.5 to 2.0 wells/mile^2. Even so, for a conservative level for $P[O' \leq (1700 \times 10^9 - 335 \times 10^9)]$, e.g., 0.05, we have a larger quantity for an upper bound than what is implied by all subjective statements except that by Grossling:[11]

[11] All quantities are in billions of barrels.

TABLE VIII

Confidence Bounds and Expectations for Total Recoverable Oil Resources of the World O, Given: Cumulative Production, the Regression Estimate for a Drilling Density of 2.0 Wells per Square Mile (2154×10^9 bbl), and $P[O' \leq (1700 \times 10^9 – 335 \times 10^9)]$

$P[O' \leq (1700 \times 10^9 – 335 \times 10^9)]$	Confidence intervals[a]					Expected value, $\bar{O}$[a]
	98%	90%	80%	70%	40%	
0.01	1700–2761	1821–2562				2168
0.05		1700–2761				2182
0.10		1593–2964	1700–2761			2200
0.15		1488–3204		1700–2761		2225
0.30		1106–4628			1700–2761	2419

[a] All quantities are in terms of billions of barrels.

Nehring (1700–2300), Hubbert (1350–2000), Warman (1971) (1200–2000), Moody and Emmerich (1972) (1800–1900), Grossling (1960–5600). Obviously, if $P[O' \leq (1700 \times 10^9 - 335 \times 10^9)]$ is higher than 0.05 (say 0.15), the range becomes broader, $(1488–3204) \times 10^9$. A value for this probability of 0.30 yields quite a broad range: $(1106–4628) \times 10^9$ bbl. The ranges reported by Grossling, Nehring, and the World Energy Conference are preferentially examined in this report because they are the most recent, and they are all at higher levels than previously estimated ranges. Thus the earlier estimated ranges are in greater contrast with the results of this study than the more recent estimates.

Given the assumptions made in this exercise, it is not possible to approximate closely Nehring's range. A value of 0.05 for $P[O' \leq (1700 \times 10^9 - 335 \times 10^9)]$ produces the lower bound, but the upper bound of the 90% range exceeds his upper bound by approximately 500×10^9 bbl. Of course, higher values for this probability increase the difference between the 90% confidence bounds and Nehring's range. The closest conformity occurs for $P[O' \leq (1700 \times 10^9 - 335 \times 10^9)] = 0.01$.

These foregoing exercises yield very different results, for they are motivated by different assumptions about the contribution of geologic data and subjective geologic analysis to oil resource assessment. The last exercise (log–normal) and the statistical model present the extremes of probability perception, for one (the statistical model) ignores geology, and the other (the log–normal) allows for geologic and discovery information but ignores the frailties of the subjective processes. The other two exercises explore the psychometric issues and yield results between these extremes.

VIII. SUMMARY STATEMENT

The first comment in summation must be one that has been made in the report in earlier sections: This study is incomplete at present because it has not considered explicitly the economic implications of the drilling densities used for inference. Consequently, all statements must be considered conditional upon the appropriateness of the drilling density. With this caveat, a few comments are offered as a summary of the study. First, for drilling densities of 1.5 and 2.0 wells/mile^2, the modal estimates of the statistical model of this study, overall, are supportive of recent estimates. Point estimates of world oil resources for these drilling densities are 2029×10^9 and 2154×10^9 bbl, respectively. These quantities lie within the ranges reported by Nehring [(1700–2300) $\times 10^9$ bbl], Grossling [(1960–5600) $\times 10^9$ bbl] and the World Energy Conference [(1920–2420) $\times 10^9$ bbl]. It is contended in this paper that the appropriate basis for comparison of estimates is that of the most likely estimate of this model (not the mean) to subjective estimates. On this basis, there is a high degree of conformity. Since a drilling density of 1.5 wells/mile^2 is only 50% greater than the current density of the United States, it seems conceivable that such a density for the ROW region (prospective area of the world exclusive of the Middle East region) is possible and may not be incompatible with the economic reference for what has been reported by most estimators as "ultimately recoverable oil resources." Given current drilling rates, such a density for the United States is not far away. Some subregions of the ROW region may never achieve such a density, but it seems possible that others, such as the United States, may exceed it.

While point estimates by the model seem to be compatible with geological estimates of world oil resources, the 90% confidence ranges produced by the statistical analysis, given drilling densities of 1.5 and 2.0 wells/mile^2 are broader than recently estimated ranges by Grossling, Nehring, and the World Energy Conference:

Nehring	(1700–2300) $\times 10^9$ bbl
Grossling	(1960–5600) $\times 10^9$ bbl
World Energy Conference	(1920–2420) $\times 10^9$ bbl
This statistical model:	
Drilling density of 1.5	(1212–8220) $\times 10^9$ bbl
Drilling density of 2.0	(1234–8335) $\times 10^9$ bbl

Neither Nehring nor Grossling provides probability or confidence statements about his range; consequently, it is not possible to really compare unequivocally the results of the statistical model with their ranges. If these ranges are considered to bound possible states of nature at a high level of confidence, e.g., 90% to 95%, then the results of the statistical model disagree markedly with Nehring and the World Energy Conference about the breadth of the range. Disagreement is less with Grossling.

The functional form selected to relate drilling density to OE density is multiplicative; thus, it is linear only in log–log space. As a result of this specification and the assumption implicit to a linear regression model that the error term is normally distributed, the histograms generated by the Monte Carlo combining of the ROW and Middle East regions are highly skewed. A result of this skewness is that while the most likely values of world oil resources for drilling densities of 1.5 and 2.0 wells/mile^2 agree generally with geologic estimates, the mean values resulting from the statistical analysis are approximately 70% greater than most likely values: 3558×10^9 bbl and 3641×10^9 bbl for drilling densities of 1.5 and 2.0, respectively. Thus, the mean (expected) values estimated by the model exceed considerably the upper bounds of the ranges reported by Nehring and the World Energy Conference. It should be noted that the breadth of the confidence bounds and the size of the mean are in part a reflection of the model specification. For example, a linear relationship of OE density to drilling density would have produced a symmetrical distribution for recoverable oil resources of the ROW region. Consequently the histogram of the combined ROW and Middle East regions would have exhibited less skewness, and the difference between the modal and mean values of the histograms would have been considerably reduced. But a linear model does not fit the data nearly as well as the log–log model does, and it is not compatible with the evidence that quantity of oil resources is log–normally distributed. Furthermore, it does not accommodate the proposition that the relationship between OE density and drilling density is asymptotic to an upper limit as drilling density increases.

A criticism of estimates of the statistical model that merits thoughtful consideration is that the breadth of the ranges, and the magnitude of the mean values for selected drilling densities, reflect the fact that the model treats important influences, such as geology, infrastructure, and political economics as a random disturbance, and that if these factors were taken into account, the distribution of possible states would be much narrower. Everything else being equal, it must indeed be true that the variance of the estimate of a statistical model which is extended to include additional factors, where real information exists, would be less than that for a simple drilling model. While conceptually this criticism is appropriate, it loses strength when we are comparing *subjective* geological estimates to the model estimates, for everything else is not equal in this case. In other words, the fact that a range based upon subjective geologic analysis is narrower than a range estimated by the drilling model cannot be interpreted as due solely to the effect of geologic data and geologic resource understanding, because other very important elements have been introduced by the use of subjective processes: the state of the experts' geologic-resource theory, the availability of useful geologic data, and the ability of an expert resource geologist to imitate a conditional probability model for a given set of geologic circumstances. Studies in psychometrics have shown that man underestimates the spread of uncertain events, providing a 60% range when he believes he is providing a 98% range. Thus, we are presented with conflicting suggestions:

- Consideration of nondrilling data would result in a narrower distribution of possible magnitudes of recoverable world resources; therefore, the subjective geologic estimates should provide a more correct and useful perception of this distribution, given all information that is available, than a model which ignores geology and other nondrilling information.
- Subjective description of uncertain events severely understates the range of possible states; consequently, the ranges reported as a result of subjective evaluation of available data are considerably narrower than the true distribution, given the available information.

Within the text of this study, three exercises were performed to explore possible implications of the issues of geologic and other information and of the deficiencies of subjective processes. The results of these exercises and of the statistical model developed in this study are summarized in Fig. 11. This figure shows three log–normal models,

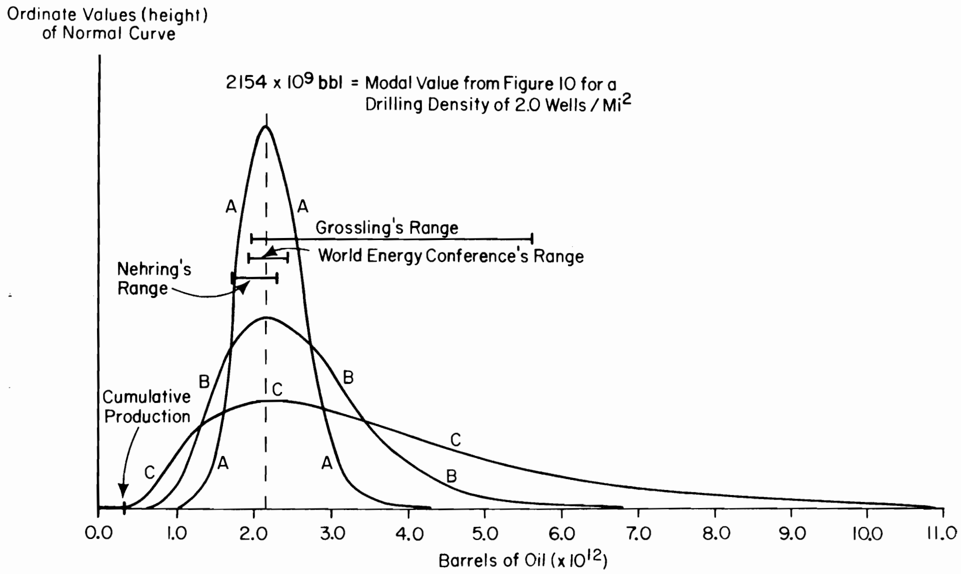

Fig. 11. Reference log–normal distributions.

each of which is forced to have as its parameter of location the logarithm of 2154×10^9, the most likely value of the histogram for a drilling density of 2.0 wells/mile^2. Distribution A[12] is the distribution arising from the third exercise of this study and here represents a very favorable perception of the contribution of geologic data and subjective geologic evaluation [$P[O' \leq (1700 \times 10^9 - 355 \times 10^9)] = 0.05$]. Distribution C[13] is a log–normal approximation to the histogram for a drilling density of 2.0 wells/mile^2. This distribution is relatively free from subjective judgments, but it ignores all geologic and other nondrilling data. Distribution B[14] is an approximation of the log–normal distribution implied by the composite ranges of Nehring and Grossling adjusted for deficiencies of the subjective processes. In essence, Fig. 11 uses the log–normal model to summarize three possible perceptions of the probability distribution of recoverable oil resources of the world. The choice of one of these three, or of any of an infinite number that lie between distributions A and C, must be predicated upon (1) an assessment of oil resource information generated by the subjective evaluation of geologic data, (2) the perceived deficiencies of subjective processes in describing uncertain events, and (3) the importance of economic effects not considered in this report.

As a geologist, the suggestion that analysis of geologic data by an expert oil geologist contributes no real information about oil resources is not acceptable as a general proposition. Nevertheless, it is a fact that drilling results on a region are dominant information when the objective of analysis is the estimation of undiscovered oil resources. One cannot ignore the possibility that it is this fact that explains the general conformity of the most likely estimates by the statistical model with

---

[12] The estimation of the parameters of this distribution was explained in the text of this report; it is based upon the assumption that the probability for nonproduced resources being less than 1365×10^9 bbl is 0.05.

[13] This approximation is based upon a simplistic fitting of the distribution: Given $\mu = \ln(2154 \times 10^9)$, σ^2 was estimated twice, once using the 5th percentile and once using the 95th percentile. The σ^2 used is an average of these two estimates.

[14] This approximation is based upon a simplistic fitting of the distribution which is identical to that used for distribution C, except that the composite range limits were assumed to be the 1st and 99th percentiles.

estimates by geologists. The identification of the appropriate distribution in Fig. 11 is not a trivial task, for doing so requires a judgment of how much useful information about oil resources is generated by the evaluation of the available geologic data and what the information gain is beyond the oil resource information already present in drilling results and OE development. Furthermore, these judgments are incomplete without consideration of the limitations and deficiencies of mental processes in the description of an event as uncertain as the magnitude of the world's oil resources. Consequently, a broad distribution, like distribution C, should not be dismissed too lightly, even though it ignores geology and other nondrilling information. At the very least, it should constitute a base case—the case of no information beyond data on drilling and OE development—to be considered by a geological expert when providing probability statements about recoverable oil resources, given the available geologic data, geologic interpretation, and his recognition of psychometric tendencies.

To the extent that a harsher assessment of the information gain from geologic data and analysis is appropriate, given data on drilling and OE development, distribution C merits greater consideration than simply being the extreme case of low information. Such an assessment may be motivated by the errors in geologic resource estimation, a notable case being the failure by geologists to perceive the large oil resources which are now being attributed to Mexico. There probably are regions of the world for which the geologic information which now is available and useful for oil resource appraisal is less than it was for Mexico prior to the discoveries of recent years. What magnitudes of errors are present in oil resource estimates for those relatively unexplored regions of the world, regions for which available useful geologic information is sparse and for which there has been little exploration drilling and OE development? It is important in considering these issues that a probability perspective be maintained. Within a probability framework, an event that is possible though not very likely cannot be ignored; it must be included in the statement of possible states and accorded an appropriate probability, for the important dimension to be captured by the probabilistic approach is the uncertainty about the actual state of nature.

A seemingly appropriate concluding remark is that although evidence suggests that we have a reasonable fix on the most likely magnitude of recoverable world oil resources, there remains considerable uncertainty about the distribution of other possible magnitudes of recoverable world oil resources. We have only begun to investigate this probability dimension. This initial investigation suggests that, given our current information, the range of possible states of nature for a high level of confidence is broader than is commonly perceived.

ACKNOWLEDGMENTS

We wish to acknowledge the assistance of Luis Chavez and Salvador Ortiz-Vertiz, who performed much of the computer work. Additionally, we are grateful to Dr. Zivia Wurtele of Pan Heuristics and Dr. Frederik P. Agterberg, Head of the Geomathematics Section of the Geological Survey of Canada, for helpful criticisms of an earlier draft of this manuscript. It is appropriate to acknowledge that many of the ideas and results here presented were initially developed by the authors in collaboration while one of us (D.P.H.) was a consultant to Pan Heuristics.

APPENDIX A

A Brief Description of the Methods Employed by Some Authors of World Oil Resource Estimates

Recent estimates of ultimate recoverable oil resources of the world are reported in Table I. Most of the data in this table were reported by Attanasi and Root (1981); we have added to their table the estimates of the World Energy Conference (1978). An in-depth examination of the methods employed by each source of an estimate is necessary to accord to each estimate a proper evaluation. But, since the primary purpose of this paper is to describe the estimates made in this study and the methodology employed to make them, only brief comment is provided regarding other estimates. In this regard, the summary description by Attanasi and Root (1981, pp. 3-10-2–3-10-3) of methodologies is useful:

> Estimates by W. P. Ryman (in Hubbert, 1969, Table 8.2, p. 194) and Shell Oil Company (cited by Warman, 1971) are

merely mentioned in their respective references without explanation of the estimation method or basic data used. The estimates presented in Warman (1971) are based upon proprietary data from British Petroleum and his estimation methods are not explained.

Estimates presented in Moody and Emmerich (1972) and Moody and Esser (1975) rely upon resource estimates for individual basins prepared by different specialists in the Mobil Corporation. According to Moody and Esser (1975), various methods were applied by Mobil researchers; the particular method used in an individual area depended on the data available. These methods included: (1) the sediment volumetric method; (2) analysis of geologic characteristics; (3) probabilistic analysis of exploration and engineering statistics; (4) analysis of production and reserve data; and (5) methods using a discovery index. Although these authors provide regional breakdowns of their estimates, they do not present basic data or identify the methods used for specific regions.

The range of estimates provided by Hubbert (1969) was based upon a scaling of Weeks's (1961) estimate and Ryman's estimate. Jodry's estimates were reported by Hubbert (1974) and appear to be based upon detailed geologic studies of the world's individual basins. The method and basic data on which Jodry based his estimates have not been published to date.

Weeks, in various publications (1961, 1958), gave estimates of 1,500 to 2,000 billion barrels for ultimate world recovery. Although the basic data have not been presented, it appears that Weeks (1958, 1965) used a volumetric approach, first determining the effective basin areas and then applying an average productivity per unit basin area to estimate expected ultimate recovery by large regions.

Hendricks (1965) followed a combination geologic analogy-volumetric approach. For the United States he estimated the quantity of oil and gas in place, on the basis of discoveries made by past exploration and estimated undiscovered petroleum. Undiscovered petroleum was estimated by using a discovery-rate technique. Explored areas of the United States were classified into four categories according to productivity.

> For each category of explored area, Hendricks derived richness factors (barrels of oil-in-place per thousand square miles of sedimentary basin). Then Hendricks divided the world's sedimentary areas into four categories corresponding to the prototype U.S. areas. To arrive at a world estimate, he multiplied measured sedimentary basin areas by the appropriate factor and summed the total. . . .
>
> The analysis by Klemme (1977) appears to rely upon the discovery history of giant fields within individual petroleum basins. He classified basins according to geologic type and reserves they contain. By unspecified methods, he estimated the amount that would be found in each basin and reported the total.

Grossling (1976) estimates ultimate recoverable oil resources of the world by first estimating the prospective area of each region (country or block of countries). Prospective area is defined as the area of region underlain by sedimentary rocks having a total thickness of at least 2000 feet. World recoverable petroleum is defined as a range: $(1960\text{–}5600) \times 10^9$ bbl. This range is determined by multiplying the total prospective area of the world by a low and a high richness factor (70,000–200,000 bbl/mile^2). Grossling estimated these richness factors by an analysis of benchmark regions (USSR, United States, Canada, Middle East). For these regions, Grossling determined cumulative production, proved reserves, and future reserve additions to known deposits. Additionally, he assembled estimates of ultimate recoverable oil for the USSR, United States, and Canada. Basically, Grossling reasoned that (1) the Middle East should be excluded when considering bounds of richness because it is so atypical; (2) the lower bound of the richness should be higher than the largest EVRD (cumulative production + reserves + future reserve additions) of the three regions; (3) the lower bound should be higher than the lower estimates of EVRD for conterminous United States; and (4) the upper bound should be lower than the higher estimate of EVRD for the USSR.

Nehring's estimate of ultimate recoverable oil resources of the world evolved from an in-depth study of the giant and super giant petroleum fields of the world. Nehring found that cumulative world oil production amounted to 335×10^9 bbl. Proved plus probable reserves were estimated to be 676×10^9 bbl. Nehring estimated that from 420×10^9 to 730×10^9 bbl of oil would be added to the reserves of known fields as these fields develop and technology improves (recoveries in the range of 45–55% and prices up to $30/bbl). Nehring estimated that exploration would add from 263×10^9 to 555×10^9 bbl. This estimate resulted from the consideration of many factors, including (1) historical discovery rates of giant oil fields in known provinces, (2) the number of untested basins, (3) the relationship of basin size to the occurrence of giant oil fields, and (4) the likelihood of the discovery of new major oil provinces.

APPENDIX B

Examination of the Fit of the Model to the Data

Definitions

Y = observation = $\ln A \cdot \ln V$, where A is the area of the region, and V is the number of barrels of hydrocarbon per square mile.

$\hat{Y}$ = prediction = $12.6071X_1 - 8.7709753X_2$, where $X_1 = \ln A$, and $X_2 = \ln A/\ln(W \times 10{,}000)$.

Residual = deviation = $Y - \hat{Y}$.

$\hat{Y}$	Y	$Y - \hat{Y}$
1.3052498E + 02	1.3182971E + 02	1.3047318E + 00
1.2939674E + 02	1.3393117E + 02	4.5344288E + 00
1.4763028E + 02	1.3798629E + 02	−9.6439906E + 00
1.4983547E + 02	1.6763894E + 02	1.7803462E + 01
1.5724449E + 02	1.3261790E + 02	−2.4626595E + 01
1.4730333E + 02	1.5213958E + 02	4.8362560E + 00
1.5201529E + 02	1.4868148E + 02	−3.3338083E + 00
1.5553076E + 02	1.4817731E + 02	−7.3534525E + 00
1.5728092E + 02	1.6604164E + 02	8.7607174E + 00
1.6318375E + 02	1.4959774E + 02	−1.3586014E + 01
1.4863741E + 02	1.5625534E + 02	7.6179355E + 00
1.7186404E + 02	1.8198674E + 02	1.0122701E + 01
1.7178770E + 02	1.7630885E + 02	4.5211434E + 00

APPENDIX C

The W Test For Normality

Method

The following description of the W test for normality is reproduced from Shapiro and Wilk (1965, pp. 602–606):

3. SUMMARY OF OPERATIONAL INFORMATION

The objective of this section is to bring together all the tables and descriptions needed to execute the W test for normality. This section may be employed independently of notational or other information from other sections.

The object of the W test is to provide an index or test statistic to evaluate the supposed normality of a complete sample. The statistic has been shown to be an effective measure of normality even for small samples ($n < 20$) against a wide spectrum of non-normal alternatives (see §5 below and Shapiro & Wilk (1964*a*)).

The W statistic is scale and origin invariant and hence supplies a test of the composite null hypothesis of normality.

To compute the value of W, given a complete random sample of size n, $x_1, x_2, \ldots, x_n$, one proceeds as follows:

(i) Order the observations to obtain an ordered sample $y_1 \leq y_2 \leq \ldots \leq y_n$.

(ii) Compute

$$S^2 = \sum_1^n (y_i - \bar{y})^2 = \sum_1^n (x_i - \bar{x})^2.$$

(iii) (*a*) If n is even, $n = 2k$, compute

$$b = \sum_{i=1}^{k} a_{n-i+1}(y_{n-i+1} - y_i),$$

where the values of a_{n-i+1} are given in Table 5.

(*b*) If n is odd, $n = 2k + 1$, the computation is just as in (iii) (*a*), since $a_{k+1} = 0$ when $n = 2k + 1$. Thus one finds

$$b = a_n(y_n - y_1) + \cdots + a_{k+2}(y_{k+2} - y_k),$$

where the value of y_{k+1}, the sample median, does not enter the computation of b.

(iv) Compute $W = b^2/S^2$.

Table 5. *Coefficients* $\{a_{n-i+1}\}$ *for the W test for normality, for* $n = 2(1)\ 50$.

i \ n	2	3	4	5	6	7	8	9	10
1	0.7071	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5888	0.5739
2	—	.0000	.1677	.2413	.2806	.3031	.3164	.3244	.3291
3	—	—	—	.0000	.0875	.1401	.1743	.1976	.2141
4	—	—	—	—	—	.0000	.0561	.0947	.1224
5	—	—	—	—	—	—	—	.0000	.0399

i \ n	11	12	13	14	15	16	17	18	19	20
1	0.5601	0.5475	0.5359	0.5251	0.5150	0.5056	0.4968	0.4886	0.4808	0.4734
2	.3315	.3325	.3325	.3318	.3306	.3290	.3273	.3253	.3232	.3211
3	.2260	.2347	.2412	.2460	.2495	.2521	.2540	.2553	.2561	.2565
4	.1429	.1586	.1707	.1802	.1878	.1939	.1988	.2027	.2059	.2085
5	.0695	.0922	.1099	.1240	.1353	.1447	.1524	.1587	.1641	.1686
6	0.0000	0.0303	0.0539	0.0727	0.0880	0.1005	0.1109	0.1197	0.1271	0.1334
7	—	—	.0000	.0240	.0433	.0593	.0725	.0837	.0932	.1013
8	—	—	—	—	.0000	.0196	.0359	.0496	.0612	.0711
9	—	—	—	—	—	—	.0000	.0163	.0303	.0422
10	—	—	—	—	—	—	—	—	.0000	.0140

i \ n	21	22	23	24	25	26	27	28	29	30
1	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407	0.4366	0.4328	0.4291	0.4254
2	.3185	.3156	.3126	.3098	.3069	.3043	.3018	.2992	.2968	.2944
3	.2578	.2571	.2563	.2554	.2543	.2533	.2522	.2510	.2499	.2487
4	.2119	.2131	.2139	.2145	.2148	.2151	.2152	.2151	.2150	.2148
5	.1736	.1764	.1787	.1807	.1822	.1836	.1848	.1857	.1864	.1870
6	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563	0.1584	0.1601	0.1616	0.1630
7	.1092	.1150	.1201	.1245	.1283	.1316	.1346	.1372	.1395	.1415
8	.0804	.0878	.0941	.0997	.1046	.1089	.1128	.1162	.1192	.1219
9	.0530	.0618	.0696	.0764	.0823	.0876	.0923	.0965	.1002	.1036
10	.0263	.0368	.0459	.0539	.0610	.0672	.0728	.0778	.0822	.0862
11	0.0000	0.0122	0.0228	0.0321	0.0403	0.0476	0.0540	0.0598	0.0650	0.0697
12	—	—	.0000	.0107	.0200	.0284	.0358	.0424	.0483	.0537
13	—	—	—	—	.0000	.0094	.0178	.0253	.0320	.0381
14	—	—	—	—	—	—	.0000	.0084	.0159	.0227
15	—	—	—	—	—	—	—	—	.0000	.0076

i \ n	31	32	33	34	35	36	37	38	39	40
1	0.4220	0.4188	0.4156	0.4127	0.4096	0.4068	0.4040	0.4015	0.3989	0.3964
2	.2921	.2898	.2876	.2854	.2834	.2813	.2794	.2774	.2755	.2737
3	.2475	.2463	.2451	.2439	.2427	.2415	.2403	.2391	.2380	.2368
4	.2145	.2141	.2137	.2132	.2127	.2121	.2116	.2110	.2104	.2098
5	.1874	.1878	.1880	.1882	.1883	.1883	.1883	.1881	.1880	.1878

(*continued*)

Table 5 (*continued*)

i \ n	31	32	33	34	35	36	37	38	39	40
6	0.1641	0.1651	0.1660	0.1667	0.1673	0.1678	0.1683	0.1686	0.1689	0.1691
7	.1433	.1449	.1463	.1475	.1487	.1496	.1505	.1513	.1520	.1526
8	.1243	.1265	.1284	.1301	.1317	.1331	.1344	.1356	.1366	.1376
9	.1066	.1093	.1118	.1140	.1160	.1179	.1196	.1211	.1225	.1237
10	.0899	.0931	.0961	.0988	.1013	.1036	.1056	.1075	.1092	.1108
11	0.0739	0.0777	0.0812	0.0844	0.0873	0.0900	0.0924	0.0947	0.0967	0.0986
12	.0585	.0629	.0669	.0706	.0739	.0770	.0798	.0824	.0848	.0870
13	.0435	.0485	.0530	.0572	.0610	.0645	.0677	.0706	.0733	.0759
14	.0289	.0344	.0395	.0441	.0484	.0523	.0559	.0592	.0622	.0651
15	.0144	.0206	.0262	.0314	.0361	.0404	.0444	.0481	.0515	.0546
16	0.0000	0.0068	0.0131	0.0187	0.0239	0.0287	0.0331	0.0372	0.0409	0.0444
17	—	—	.0000	.0062	.0119	.0172	.0220	.0264	.0305	.0343
18	—	—	—	—	.0000	.0057	.0110	.0158	.0203	.0244
19	—	—	—	—	—	—	.0000	.0053	.0101	.0146
20	—	—	—	—	—	—	—	—	.0000	.0049

i \ n	41	42	43	44	45	46	47	48	49	50
1	0.3940	0.3917	0.3894	0.3872	0.3850	0.3830	0.3808	0.3789	0.3770	0.3751
2	.2719	.2701	.2684	.2667	.2651	.2635	.2620	.2604	.2589	.2574
3	.2357	.2345	.2334	.2323	.2313	.2302	.2291	.2281	.2271	.2260
4	.2091	.2085	.2078	.2072	.2065	.2058	.2052	.2045	.2038	.2032
5	.1876	.1874	.1871	.1868	.1865	.1862	.1859	.1855	.1851	.1847
6	0.1693	0.1694	0.1695	0.1695	0.1695	0.1695	0.1695	0.1693	0.1692	0.1691
7	.1531	.1535	.1539	.1542	.1545	.1548	.1550	.1551	.1553	.1554
8	.1384	.1392	.1398	.1405	.1410	.1415	.1420	.1423	.1427	.1430
9	.1249	.1259	.1269	.1278	.1286	.1293	.1300	.1306	.1312	.1317
10	.1123	.1136	.1149	.1160	.1170	.1180	.1189	.1197	.1205	.1212
11	0.1004	0.1020	0.1035	0.1049	0.1062	0.1073	0.1085	0.1095	0.1105	0.1113
12	.0891	.0909	.0927	.0943	.0959	.0972	.0986	.0998	.1010	.1020
13	.0782	.0804	.0824	.0842	.0860	.0876	.0892	.0906	.0919	.0932
14	.0677	.0701	.0724	.0745	.0765	.0783	.0801	.0817	.0832	.0846
15	.0575	.0602	.0628	.0651	.0673	.0694	.0713	.0731	.0748	.0764
16	0.0476	0.0506	0.0534	0.0560	0.0584	0.0607	0.0628	0.0648	0.0667	0.0685
17	.0379	.0411	.0442	.0471	.0497	.0522	.0546	.0568	.0588	.0608
18	.0283	.0318	.0352	.0383	.0412	.0439	.0465	.0489	.0511	.0532
19	.0188	.0227	.0263	.0296	.0328	.0357	.0385	.0411	.0436	.0459
20	.0094	.0136	.0175	.0211	.0245	.0277	.0307	.0335	.0361	.0386
21	0.0000	0.0045	0.0087	0.0126	0.0163	0.0197	0.0229	0.0259	0.0288	0.0314
22	—	—	.0000	.0042	.0081	.0118	.0153	.0185	.0215	.0244
23	—	—	—	—	.0000	.0039	.0076	.0111	.0143	.0174
24	—	—	—	—	—	—	.0000	.0037	.0071	.0104
25	—	—	—	—	—	—	—	—	.0000	.0035

(v) 1, 2, 5, 10, 50, 90, 95, 98 and 99% points of the distribution of W are given in Table 6. *Small values of W are significant,* i.e. indicate non-normality.

(vi) A more precise significance level may be associated with an observed W value by using the approximation detailed in Shapiro & Wilk (1965*a*).

Table 6. *Percentage points of the W test* for n* = 3(1) 50

	Level								
n	0.01	0.02	0.05	0.10	0.50	0.90	0.95	0.98	0.99
3	0.753	0.756	0.767	0.789	0.959	0.998	0.999	1.000	1.000
4	.687	.707	.748	.792	.935	.987	.992	.996	.997
5	.686	.715	.762	.806	.927	.979	.986	.991	.993
6	0.713	0.743	0.788	0.826	0.927	0.974	0.981	0.986	0.989
7	.730	.760	.803	.838	.928	.972	.979	.985	.988
8	.749	.778	.818	.851	.932	.972	.978	.984	.987
9	.764	.791	.829	.859	.935	.972	.978	.984	.986
10	.781	.806	.842	.869	.938	.972	.978	.983	.986
11	0.792	0.817	0.850	0.876	0.940	0.973	0.979	0.984	0.986
12	.805	.828	.859	.883	.943	.973	.979	.984	.986
13	.814	.837	.866	.889	.945	.974	.979	.984	.986
14	.825	.846	.874	.895	.947	.975	.980	.984	.986
15	.835	.855	.881	.901	.950	.975	.980	.984	.987
16	0.844	0.863	0.887	0.906	0.952	0.976	0.981	0.985	0.987
17	.851	.869	.892	.910	.954	.977	.981	.985	.987
18	.858	.874	.897	.914	.956	.978	.982	.986	.988
19	.863	.879	.901	.917	.957	.978	.982	.986	.988
20	.868	.884	.905	.920	.959	.979	.983	.986	.988
21	0.873	0.888	0.908	0.923	0.960	0.980	0.983	0.987	0.989
22	.878	.892	.911	.926	.961	.980	.984	.987	.989
23	.881	.895	.914	.928	.962	.981	.984	.987	.989
24	.884	.898	.916	.930	.963	.981	.984	.987	.989
25	.888	.901	.918	.931	.964	.981	.985	.988	.989
26	0.891	0.904	0.920	0.933	0.965	0.982	0.985	0.988	0.989
27	.894	.906	.923	.935	.965	.982	.985	.988	.990
28	.896	.908	.924	.936	.966	.982	.985	.988	.990
29	.898	.910	.926	.937	.966	.982	.985	.988	.990
30	.900	.912	.927	.939	.967	.983	.985	.988	.990
31	0.902	0.914	0.929	0.940	0.967	0.983	0.986	0.988	0.990
32	.904	.915	.930	.941	.968	.983	.986	.988	.990
33	.906	.917	.931	.942	.968	.983	.986	.989	.990

(*continued*)

Table 6 (*continued*)

n	Level 0.01	0.02	0.05	0.10	0.50	0.90	0.95	0.98	0.99
34	.908	.919	.933	.943	.969	.983	.986	.989	.990
35	.910	.920	.934	.944	.969	.984	.986	.989	.990
36	0.912	0.922	0.935	0.945	0.970	0.984	0.986	0.989	0.990
37	.914	.924	.936	.946	.970	.984	.987	.989	.990
38	.916	.925	.938	.947	.971	.984	.987	.989	.990
39	.917	.927	.939	.948	.971	.984	.987	.989	.991
40	.919	.928	.940	.949	.972	.985	.987	.989	.991
41	0.920	0.929	0.941	0.950	0.972	0.985	0.987	0.989	0.991
42	.922	.930	.942	.951	.972	.985	.987	.989	.991
43	.923	.932	.943	.951	.973	.985	.987	.990	.991
44	.924	.933	.944	.952	.973	.985	.987	.990	.991
45	.926	.934	.945	.953	.973	.985	.988	.990	.991
46	0.927	0.935	0.945	0.953	0.974	0.985	0.988	0.990	0.991
47	.928	.936	.946	.954	.974	.985	.988	.990	.991
48	.929	.937	.947	.954	.974	.985	.988	.990	.991
49	.929	.937	.947	.955	.974	.985	.988	.990	.991
50	.930	.938	.947	.955	.974	.985	.988	.990	.991

* Based on fitted Johnson (1949) S_B approximation, see Shapiro & Wilk (1965*a*) for details.

To illustrate the process, suppose a sample of 7 observations were obtained, namely $x_1 = 6$, $x_2 = 1$, $x_3 = -4$, $x_4 = 8$, $x_5 = -2$, $x_6 = 5$, $x_7 = 0$.

(i) Ordering, one obtains

$$y_1 = -4,\ y_2 = -2,\ y_3 = 0,\ y_4 = 1,\ y_5 = 5,\ y_6 = 6,\ y_7 = 8.$$

(ii) $S^2 = \Sigma y_i^2 - \frac{1}{7}(\Sigma y_i)^2 = 146 - 28 = 118.$

(iii) From Table 5, under $n = 7$, one obtains

$$a_7 = 0.6233,\ a_6 = 0.3031,\ a_5 = 0.1401,\ a_4 = 0.0000.$$

Thus

$$b = 0.6233(8 + 4) + 0.3031(6 + 2) + 0.1401(5 - 0) = 10.6049.$$

(iv) $W = (10.6049)^2/118 = 0.9530.$

(v) Referring to Table 6, one finds the value of W to be substantially larger than the tabulated 50% point, which is 0.928. Thus there is no evidence, from the W test, of nonnormality of this sample.

4. Examples

Example 1. Snedecor (1946, p. 175), makes a test of normality for the following sample of weights in pounds of 11 men: 148, 154, 158, 160, 161, 162, 166, 170, 182, 195, 236.

The *W* statistic is found to be 0.79 which is just below the 1% point of the null distribution. This agrees with Snedecor's approximate application of the $\sqrt{b_1}$ statistic test.

Example 2. Kendall (1948, p. 194) gives an extract of 200 'random sampling numbers' from the Kendall–Babington Smith, *Tracts for Computers* No. 24. These were totalled, as number pairs, in groups of 10 to give the following sample of size 10: 303, 333, 406, 457, 461, 469, 474, 489, 515, 583.

The *W* statistic in this case has the value 0.9430, which is just above the 50% point of the null distribution.

Example 3. Davies *et al.* (1956) give an example of a 2^5 experiment on effects of five factors on yields of penicillin. The 5-factor interaction is confounded between 2 blocks. Omitting the confounded effect the *ordered* effects are

C	0.0958	ABC	0.0002
BC	.0333	CD	−0.0026
ACDE	.0293	B	−0.0036
BCE	.0246	BD	−0.0042
ACD	.0206	BCD	−0.0113
ABCE	.0194	ABE	−0.0139
DE	.0191	ABD	−0.0211
BE	.0182	AC	−0.0333
BDE	.0173	AD	−0.0341
ADE	.0132	ACE	−0.0363
BCDE	.0102	ABCD	−0.0363
ABDE	.0084	AB	−0.0402
CDE	.0077	CE	−0.0582
D	.0058	A	−0.1184
AE	.0016	E	−0.1398

In their analysis of variance, Davies *et al.* pool the 3- and 4-factor interactions for an error term. They do not find the pooled 2-factor interaction mean square to be significant. But note that CE is significant at the 5% point on a standard *F*-test. However, on the basis of a Bartlett test, they find that the significance of CE does *not* reach the 5% level.

The Data and Application in This Study

The ordered residuals (deviations, actual value–estimated value) are as follows:

−24.6266	4.5344
−13.5860	4.8363
−9.6440	7.6179
−7.3535	8.7607
−3.3338	10.0123
−1.3047	17.8035
4.5211	

The coefficients, taken from Shapiro and Wilk's Table 5, reprinted in this Appendix, are as follows:

Term	Coefficient
1	0.5350
2	0.3325
3	0.2412
4	0.1707
5	0.1099
6	0.0539

The W statistic is defined as $W = b^2/S^2$. The numerator b is computed by combining appropriately the coefficients and pairs of the order deviations:

$$
\begin{aligned}
b = {} & (0.5350)(17.8035 + 24.6266) \\
& + (0.3325)(10.0123 + 13.5860) \\
& + (0.2412)(8.7607 + 9.6440) \\
& + (0.1707)(7.6179 + 7.3535) \\
& + (0.1099)(4.8363 + 3.3338) \\
& + (0.0539)(4.5344 + 1.3047)
\end{aligned}
$$

$$b = 22.7001 + 7.8464 + 4.4392 + 2.5556 + 0.8979 + 0.3147 = 38.7539$$

Thus,

$$b^2 = (38.7539)^2 = 1501.9$$

The denominator S^2 can be computed from the standard error of the estimate SE, a statistic from the regression analysis:

$$S^2 = (\mathrm{SE})^2 \left(\frac{n-2}{n}\right) = (11.9453)^2 \left(\frac{11}{13}\right) = 142.69\left(\frac{11}{13}\right) = 120.7$$

Finally, we have a value for the W statistic of 12.44:

$$W = 1501.9/120.7 = 12.44$$

The W statistic is much larger than even the 99th percentile value, indicating strongly that the deviations are normally distributed. This statistic tests the null hypothesis that the observations are normally distributed. Small values of the statistic are indicative of nonnormality. For example, suppose that $W = 0.836$. The probability of getting a value for W this small or smaller when the observations are normally distributed is less than 0.02. Thus, given such a small value, we can expect to be incorrect in rejecting the hypothesis 2% of the time.

Let us now evaluate the W statistic of this study, 12.44. From Shapiro and Wilk's Table 6, we have for a sample of size 13, the following percentiles for W:

Probability	Percentile Value for W
0.01	0.814
0.02	0.837
0.05	0.866
0.10	0.889
0.50	0.945
0.90	0.974
0.95	0.979
0.98	0.984
0.99	0.986

The statistic is much larger than even the 99th percentile value, indicating strongly that the deviations from the regression estimates are normally distributed.

APPENDIX D

Mathematical Description of the Regression Model

Regression Analysis on Log–Log–Normal Models

The standard linear regression model is of the form

$$Y = XB + u$$

where X is a row vector, B is a column vector, and u is normally distributed with mean zero and variance σ^2. If $(Y_1, X_1), \ldots, (Y_n, X_n)$ is a random sample then the least squares and maximum likelihood estimators, for B and σ^2, are the same. In the discussion of the Initial Model, $Y = \ln V$ and $XB = \ln(f(w))$, and as noted there the assumption of constant variance for Y (e.g., u) is not a realistic one. In addition to the reasons presented there we can see that if V were to represent the total hydrocarbon in a region it could be written as a sum $V_1 + \cdots + V_A$, with each V_i being the hydrocarbons in a particular one-square-mile region and A being the area of the region. The average number of barrels per square mile is like a sample mean and the variance of the sample mean should be $\sigma^2/\sqrt{A}$, where σ^2 is the common variance of the V_i's. Unfortunately, if the V_i's were assumed log–normal then the sample mean would not be log–normal. It is clearly easier to begin with an assumption that the average number of barrels per square mile is log–normal. It is also clear that the assumption that the variance is constant with respect to the area is unrealistic. Assume then that Var $Y = d_A^2\sigma^2$, where d_A^2 is a known constant inversely related to area. In the Complex Model, $d_A^2 = 1/\ln A$.

If we let

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \qquad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix},$$

$$D = \begin{bmatrix} d_1^2 & & 0 \\ & \ddots & \\ 0 & & d_n^2 \end{bmatrix}$$

the likelihood function is

$$\begin{aligned} &L(y_1, \ldots, y_n; x_1, \ldots, x_n; B_1, B_2) \\ &\quad = \sigma^{-2n}(2\pi)^{-n/2} \exp[-(2\sigma^2)^{-1}(Y - XB)^{\mathrm{T}}D^{-1}(Y - XB)] \\ &\quad = \sigma^{-n/2}(2\pi)^{-n/2} \exp[-(2\sigma^2)^{-1}(D^{-1/2}Y - D^{-1/2}XB)^{\mathrm{T}}(D^{-1/2}Y - D^{-1/2}XB)] \end{aligned}$$

With a transformation of variables $Y \to D^{-1/2}Y$ and $X \to D^{-1/2}Y$, we have the standard case of constant variance, and the ML estimators are

$$\hat{B} = (X^T D^{-1} X)^{-1} X^T D^{-1} Y, \qquad S^2 = (n-2)^{-1}[Y^T D^{-1} Y - \hat{B}^T Y^T D^{-1} X]$$

where S^2 has already been adjusted for the bias. From known properties of $\hat{B}$, S we have that

$$\frac{([1\ x_0]\hat{B}/d_0) - d_0^{-1}E(Y)}{S\{[1\ x_0](X^T D^{-1} X)^{-1}[1\ x_0]^T\}^{1/2}} = \frac{([1\ x_0]\hat{B}/d_0) - d_0^{-1}E(Y)}{S_1}$$

is t-distributed with $df = n - 2$.

We note that for the untransformed model we could not use least squares to obtain the estimators, but in the transformed model least squares and maximum likelihood estimation coincide.

Variance of Y versus Variance of V. We began with the observation that the variance of V_w should be like the variance of a sample mean, but in the model we assumed that Var $Y = d^2\sigma^2$. We can now see that these requirements are compatible. Since Y is log–normal

$$\text{Var } Y = e^{2XB} e^{d^2\sigma^2}(e^{d^2\sigma^2} - 1)$$

If $k > 0$ and $\sigma^2 + \ln k > 0$ then letting $d^2 = 1 + \ln k/\sigma^2$ gives $e^{d^2\sigma^2} = ke^{\sigma^2}$ and Var $V_w = k[e^{2XB}e^{\sigma^2}(e^{d^2\sigma^2} - 1)]$. (1) If $e^{d^2\sigma^2} - 1$ is approximately one then $\text{Var}(V_A) \sim k\ \text{Var}(V_1)$. (2) If $e^{d^2\sigma^2}$ is large compared to one then $\text{Var}(V_A) \sim k^2 e^{\sigma^2}\ \text{Var}(V_1)$, where V_A and V_1 are the total hydrocarbons in regions of area A and 1, respectively. When $d^2 = 1/\ln A$, $k = \exp[\sigma^2(1/\ln A - 1)]$.

Confidence Intervals. It would be desirable to obtain a confidence interval for $E(V_w)$. Instead we can only obtain an interval for $E(\ln V_w)$, and furthermore $E(\ln V_w) \neq \ln E(V_w)$. However by Jensen's inequality (Parzen, 1960), $E(\ln V_w) \leq \ln E(V_w)$ and we can obtain a one-sided confidence interval, i.e., a lower bound for $E(V_w)$.

Letting

$$a = [1\ x_0]\hat{B} - t_{\alpha/2}^{n-2} S_1 d_0 = E(\ln \hat{V}_w) - t_{\alpha/2}^{n-2} S_1 d_0$$

$$b = [1\ x_0]\hat{B} + t_{\alpha/2}^{n-2} S_1 d_0 = E(\ln \hat{V}_w) + t_{\alpha/2}^{n-2} S_1 d_0$$

$$a \leq E(\ln V_w) \leq b$$

with confidence level $1 - \alpha$ gives us

$$e^a \leq e^{E(\ln V_w)} \leq e^b$$

$$e^a \leq e^{\ln E(V_w)} = E(V_w)$$

This difficulty can be partially avoided by obtaining a confidence interval for ln V_w instead of $E(\ln V_w)$; however, the variance is larger so a wider confidence interval is obtained. Instead of

$$S_1 = S\{[1\ x_0](X^TX)^{-1}[1\ x_0]^T\}^{1/2}$$

we must use

$$S_2 = S\{1 + [1\ x_0](X^TX)^{-1}[1\ x_0]^T\}^{1/2}$$

Implementing the Computer Solution. In using a standard computer library routine the transformation

$$Y \rightarrow d^{-1/2}Y, \qquad X \rightarrow D^{-1/2}X$$

is awkward, and hence the problem was reformulated as a regression problem with two independent variables,

$$X_1 = \ln A_i, \qquad X_1 = d_i^{-1/2}$$

$$X_2 = \ln A \cdot (\ln W_i \cdot 10^4)^{-1}, \qquad X_2 = d_i^{-1/2}(\ln W_i \cdot 10^4)$$

$$Y = \ln A_i \cdot \ln V_i, \qquad Y = d_i^{-1/2} \ln V_i$$

and suppressed constant term.

Unbiased Estimators. $E(\ln \hat{V}_w) = X\hat{B}$ is an unbiased estimator of $E(\ln V_w) = XB$ and S^2 is an unbiased estimator of σ^2. Hence $X\hat{B} + \frac{1}{2}S^2$ is an unbiased estimator of $XB + \frac{1}{2}\sigma^2$. However, $\exp(X\hat{B} + \frac{1}{2}S^2)$ is not an unbiased estimator of $\exp(XB + \frac{1}{2}\sigma^2) = E(V_w)$.

Since $X\hat{B}$, S^2 are independent

$$E[\exp(X\hat{B} + \tfrac{1}{2}S^2)] = E(\exp X\hat{B}) \cdot E(\exp \tfrac{1}{2}S^2)$$

$$E(e^{(1/2)S^2}) = E[(e^t)^{\sigma^2/2(n-1)}]$$

where t is t-distributed with $df = n - 1$ and

$$E(e^{X\hat{B}}) = e^{XB}E[(e^z)^\sigma]$$

where Z is standard normal.

REFERENCES

Alpert, M., and Raiffa, H. (1969). "A Progress Report on Training of Probability Assessors." Unpublished draft, Harvard Business School, Harvard University, Cambridge, Massachusetts.

Attanasi, E. D., and Root, D. H. (1981). World oil resource estimates: Differences and difficulties. *In* Workshop Proceedings: World Oil and Natural Gas Supplies, pp. 3-10-1–3-10-16. Report #WS-79-182, Electric Power Research Institute, Palo Alto, California.

Barouch, E., and Kaufman, G. M. (1977). Estimation of undiscovered oil and gas. *In* Mathematical Aspects of Production and Distribution of Energy, Proceedings of Symposia in Applied Mathematics, Vol. XXI, pp. 77–91. American Mathematical Society.

Davies, O. L., ed. (1956). "Design and Analysis of Industrial Experiments." Oliver and Boyd, London.

DeWijs, H. J. (1953). Statistics of ore distribution, Part II. *J. R. Neth. Mineral. Soc.*, January.

Grossling, B. F. (1976). "Window on Oil." Financial Times Business Information, London.

Hendricks, T. A. (1965). "Resources of Oil, Gas and Natural Gas Liquids in the United States and the World." U.S. Geological Survey Circular No. 522.

Hubbert, M. K. (1969). Energy resources. *In* "Resources and Man," Chap. 8, pp. 157–242. Freeman, San Francisco.

Hubbert, M. K. (1974). U.S. energy resources, a review as of 1972, Part 1. *In* "A National Fuels and Energy Policy Study," U.S. 93rd Congress, 2nd Session, Senate Committee on Interior and Insular Affairs, Comm. Print, Serial No. 93-40 (92-75).

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika* **36,** 149–176.

Kendall, M. G. (1948). "The Advanced Theory of Statistics." Griffin and Co., London.

Klemme, H. D. (1977). World oil and gas reserves from giant fields and petroleum basins (provinces). *In* "The Future Supply of Nature-Made Petroleum and Gas" (R. F. Meyer, ed.), pp. 217–270. Pergamon, New York.

Menard, H. W., and Sharman, G. (1975). Scientific uses of random drilling models. *Science* **190,** 337–343.

Miller, B., Thomsen, H., Dolton, G., Coury, A., Hendricks, T., Lennartz, F., Powers, R., Sable, E., and Varnes, K., (1975). "Geological Estimates of Undiscovered Recoverable Oil and Gas Resources in the United States." U.S. Geological Survey Circular No. 725.

Moody, J. D., and Emmerich, H. H. (1972). Giant oil fields of the world. *Proc. 24th Int. Geol. Congr., Montreal, Sec. 5, Mineral Fuels,* pp. 161–167.

Moody, J. D., and Esser, R. W. (1975). An estimate of the world's recoverable crude oil resources. *Proc. Ninth World Pet. Congr.* **13,** 11–20. Applied Science Publishers, London.

Nehring, R. (1978). "Giant Oil Fields and World Oil Resources." R-2284-CIA, Rand Corporation, Santa Monica, California.

Parzen, E. (1960). "Modern Probability Theory and Its Applications." Wiley, New York.

Sachman, H. (1974). "Delphi Assessment: Expert Opinion, Forecasting, and Group Process." R-1283-PR, April 1974, Rand Corporation, Santa Monica, California.

Shapiro, S. S., and Wilk, M. B. (1964a). A comparative study of various tests for normality. (Unpublished manuscript: invited paper, American Statistical Association, Annual Meeting, Chicago, Illinois, December, 1964.)

Shapiro, S. S., and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* **52,** 591–611.

Shapiro, S. S., and Wilk, M. B. (1965a). Testing the normality of several samples. (Unpublished manuscript.)

Snedecor, G. W. (1946). "Statistical Methods," 4th edition. Iowa State College Press, Ames, Iowa.

Warman, H. R. (1971). Future problems in petroleum exploration, *Pet. Rev.* **25,** No. 291, 96–101.

Weeks, L. G. (1958). Fuel reserves of the future. *Am. Assoc. Pet. Geol. Bull.* **42,** 431–438.

Weeks, L. G. (1961). The next hundred years energy demand and sources of supply. *Alberta Soc. Pet. Geol. J.* **9,** No. 5, 141–157.

Weeks, L. G. (1965). World offshore petroleum resources. *Am. Assoc. Pet. Geol. Bull.* **49,** 1680–1693.

World Energy Conference (1978). "World Energy Resources 1985–2020, Oil and Gas Resources, The Full Reports to the Conservation Commission of the World Energy Conference," Vol. 3, published for the WEC by IPC Science and Technology Press.