The Creaming Method: A Bayesian Procedure to Forecast Future Oil and Gas Discoveries in Mature Exploration Provinces
Author(s): J. Meisner and F. Demirmen
Source: *Journal of the Royal Statistical Society. Series A (General)*, Vol. 144, No. 1 (1981), pp. 1-31
Published by: Wiley for the Royal Statistical Society
Stable URL: https://www.jstor.org/stable/2982158

J. R. Statist. Soc. A (1981),
144, *Part* 1, *pp.* 1–31

The Creaming Method: A Bayesian Procedure to Forecast Future Oil and Gas Discoveries in Mature Exploration Provinces

By J. Meisner and F. Demirmen

Shell International Petroleum Co.

[Read before the Royal Statistical Society on Wednesday, July 2nd, 1980, the President Professor D. R. Cox in the Chair]

Summary

The creaming method is a statistical procedure aimed at forecasting future oil and gas discoveries in petroleum provinces, based on trends in exploration results observed in those provinces. The model recognizes the fact that, generally, the discovery rate (success rate) and the volumes (field sizes) tend to decline with advancing exploration. This is called the creaming phenomenon. Trends in success rate and field sizes are analysed separately and departures from the model are investigated.

The forecast is based on these analyses and uses as an indicator the number of exploration wells expected or planned to be drilled. The results are given in the form of predictive probability distributions. These include estimates of the number and the total volume of future discoveries.

The method is applicable to areas where discoveries are on a generally declining trend, or, for short-term estimates, on a generally constant trend.

Keywords: DISCOVERY; EXPECTATION CURVE; EXPLORATION; EXPLORATION MATURITY; EXPLORATION TREND; FORECASTING; GAS; LOGISTIC MODEL; LOGNORMAL DISTRIBUTION; OIL; PLANNING; PREDICTIVE DISTRIBUTION; PROVINCE; RESERVES; SUCCESS RATE; STATISTICAL ANALYSIS

1. Introduction

1.1. *General*

Forecasts of future discoveries of crude oil and natural gas in petroleum provinces are of great importance for planning and policy-making in the petroleum and related industry. In this paper we discuss a relatively simple statistical procedure to forecast future discoveries in exploration provinces. By "future discoveries" we mean the prospective quantity of crude oil or natural gas yet to be found by a given number of exploration wells which are planned to be drilled in a particular petroleum province. The term "petroleum province" is used in a broad sense, referring to a regional geographic area known to have commerical hydrocarbon accumulations, and need not refer to a geologic province with distinct geologic (e.g. tectonic) properties or oil-accumulation conditions. For practical purposes, national frontiers can be considered to be province boundaries.

Our method is intended basically for petroleum provinces that are in a mature exploration stage. For making short-term forecasts, however, the method is also applicable to exploration provinces that are in a transitional stage from immature to mature. A mature exploration province is one in which, after drilling a relatively large number of exploration wells, the discoveries are on a general declining trend. If the discoveries are on a general rising trend, the province is immature, and if the discoveries show a general constant trend, the area is in transition from immature to mature.

Furthermore, although our method can be used to forecast future discoveries of both crude oil and natural gas, forecasts for the two should preferably be kept separate. If it is necessary to include both oil and gas in a forecast, the volumetric units of the two should be compatible, as for example by expressing gas volumes as BTU-equivalent oil volumes.

1.2. *Past Exploration Trends*

Past exploration trends in a given petroleum province form the cornerstone of our forecasting procedure. We consider past exploration trends in terms of discoveries made as a function of the number of exploration wells drilled. Plotting the cumulative discoveries against the total number of exploration wells in a historical sequence provides a good visual description of such exploration trends (Fig. 1). Some kind of growth curve is then obtained. For a mature exploration province, the growth rate will be on a declining course and for a province in a transitional state, it will be constant. Forecasting by our method is tantamount to projection of the past exploration trend into the future over a specified number of exploration wells planned to be drilled. The number of future exploration wells serves as an indicator for the forecast.

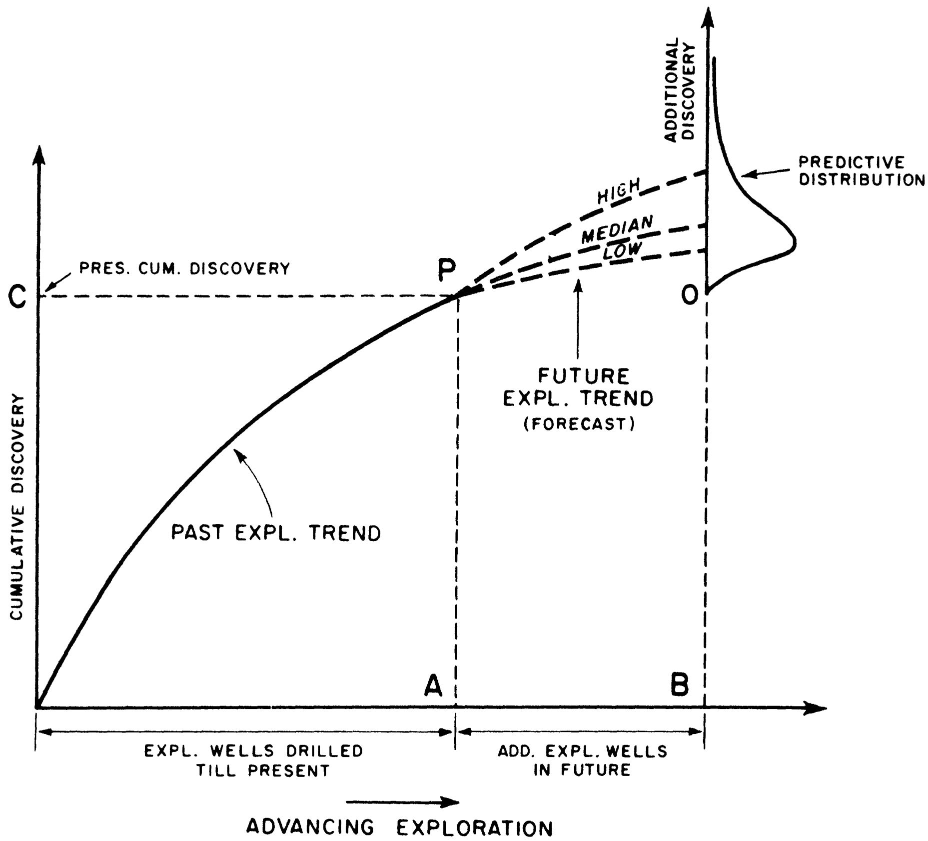

FIG. 1. Forecast of future oil discoveries.

Exploration trends of the type considered above are the combined result of two independent factors: (i) success rate, that is the proportion of exploration wells that resulted in field discoveries, and (ii) field size, that is the ultimately recoverable volume of oil or gas attributable to a field. In our method we consider both factors separately, though in the final analysis the effects of the two are combined in making predictions. Experience has shown that the field size plays a far greater role in defining the shape of the discovery trend than does the success rate (see also below).

1.3. *Creaming Phenomenon*

The statistical model we employ for prediction makes use of a generally observed phenomenon in exploration provinces. This phenomenon, which may be termed "creaming", is the diminishing effectiveness of exploration effort with advancing exploration. This phenomenon is so common that it characterizes all mature exploration provinces, and in fact, if exploration is conducted with any measure of efficiency, it is the process which all exploration provinces must eventually go through.

The creaming phenomenon is in part an expression of a decline in the success rate, i.e. decline in the probability of success. This in turn is a reflection of the fact that in a petroleum province there are a finite number of commercial fields, so that with each discovery the chance of making another discovery is reduced. This decline, however, is to some extent counteracted by the industry's increased capability to locate new fields through advances in technology. The net result is such that the probability of success, even in mature exploration provinces, declines only gently, and its effect on the overall creaming phenomenon tends to be relatively small.

A far greater role in the creaming phenomenon is that of the field size. Basically the creaming phenomenon reflects diminishing field size with advancing exploration. This attests to the fact that the industry, through its technological know-how, is generally capable of finding the larger fields early in exploration, and the smaller ones in progressively later stages. There may be exceptions to this general trend whereby a "learning period", during which the field size tends to increase, may exist in the early stages of exploration in immature provinces. But once this learning period is passed, the general declining trend is established almost invariably. Exploration trends of the type shown in Fig. 1, characterizing mature exploration provinces, are an expression of the creaming phenomenon, mostly reflecting a decline in the field size, and to a lesser extent a decline in the success rate.

The oil industry achieves the creaming phenomenon through some sort of prospect appraisal scheme whereby various exploration prospects in a province are ranked according to their prospectivity as poor, promising, etc., on the basis of available information about the particular prospects involved and a wider experience obtained from other areas. As new information (reviews, geophysical surveys, drilling results) is gathered, the evaluation of the prospects is continuously updated. All this process results in characterization of each individual prospect in terms of (i) the probability of success, and (ii) the field size, assuming discovery. Assignment of these two parameters usually involves subjective or judgemental probabilities. The estimation procedure may range from very informal, unstructured exercises to formal, well-structured decision-making processes. In a formal appraisal system, the probability of success (sometimes called the chance factor) and the field size estimates may be combined in a so-called "expectation curve" as shown in Fig. 2(a). [In fact the curve shows the so-called survival function. Thus the term "expectation curve" is misleading, as it does not refer to expectations in the statistical sense. Nonetheless, the term is widely used within the Shell Group, and we shall use it here as well.] In this diagram each potential field size is plotted in appropriate volumetric units against the probability that an actual field size equal to or larger than the potential value will materialize. Fig. 2(b) shows some examples of expectation curves. Such expectation curves can be used to rank exploration prospects and help optimise exploration drilling sequence according to the oil company's utilities. The creaming phenomenon observed in exploration provinces is essentially the result of such ranking exercises based on technical know-how.

The creaming phenomenon for a real exploration province coded XX11 is illustrated in Fig. 3. Except for a scale difference along the y-axis, the exploration trend shown was the actual observed trend for this province in the early 1970s. The general declining trend in the discoveries with advancing exploration is clearly visible.

1.4. *Gambling Aspect*

Although advances in geology and geophysics may substantially reduce the uncertainty of exploration drilling outcomes, exploration remains very much a gamble. A lottery bears much

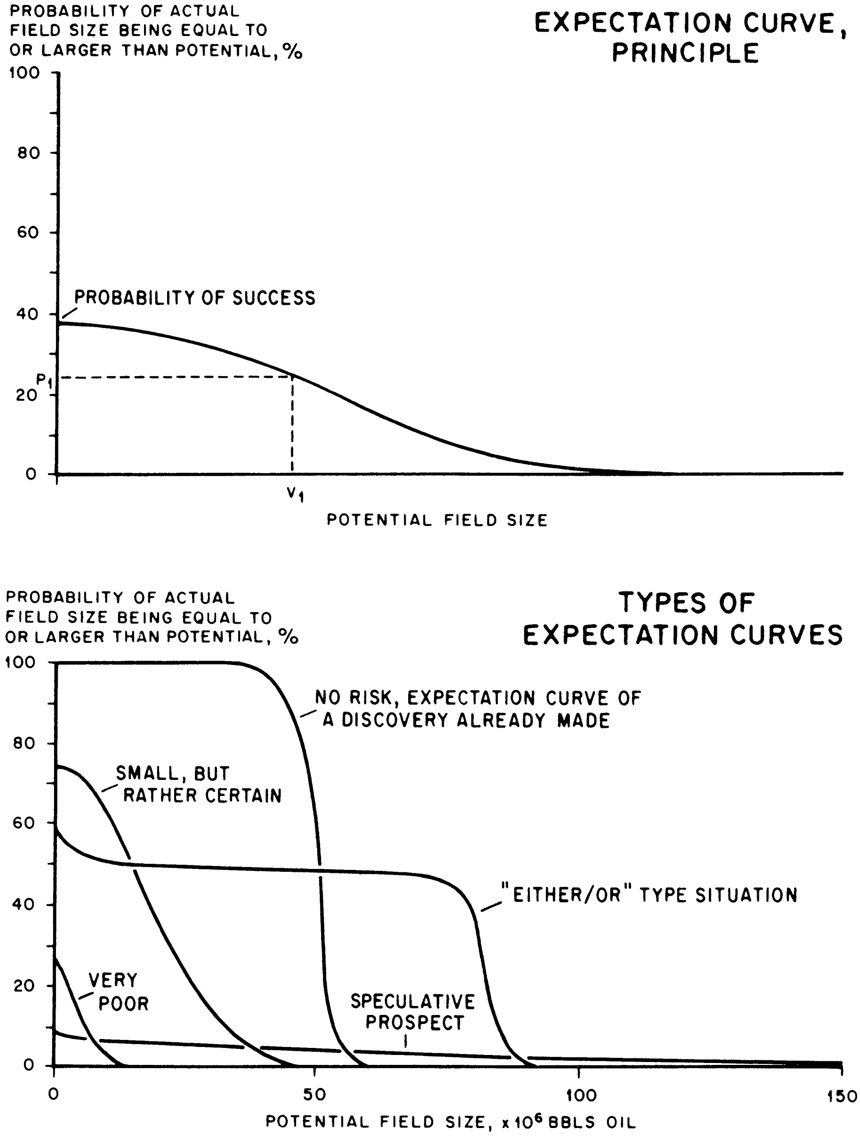

FIG. 2. Expectation curves.

resemblance to an exploration venture. The large number of structures or trapping configurations in a province which might contain oil or gas are the tickets in the lottery, and the costs of concession, surveys and exploration drilling constitute the price of a ticket. A prize is the discovery of a hydrocarbon accumulation, and the ultimately recoverable volume (the field size) is the reward, or the size of the prize. Through a prospect appraisal system, tickets believed to be most promising are bought early in exploration, and those holding lesser promise at a later stage; unpromising tickets are not bought at all. The specific order and stopping rule depend on the particular situation and the oil company's utilities. Drilling an exploration well means finding out whether or not a structure contains oil and/or gas accumulations and may consist of not only one but several drillings.

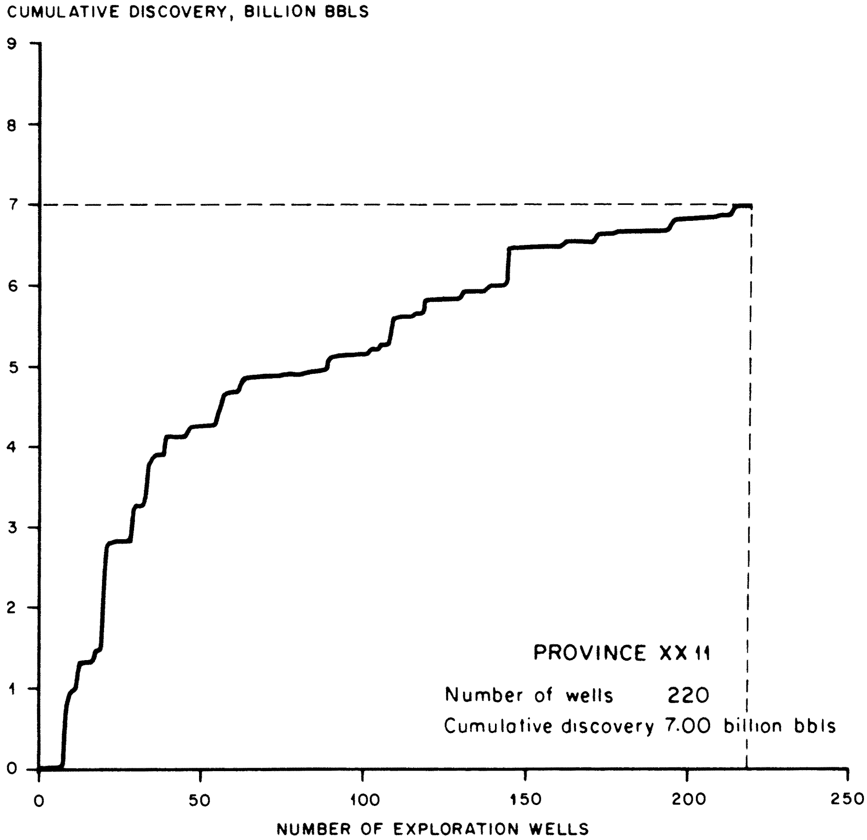

FIG. 3. Cumulative discovery against number of exploration wells to illustrate the creaming phenomenon.

Our forecasting method takes cognizance of the gambling aspect of petroleum exploration. It views the search for hydrocarbons as a stochastic process with inherent risks and rewards, and provides a forecast in terms of a predictive probability distribution that incorporates all the available information from the past exploration trends. Inferences about crude oil or natural gas discoveries from future exploration activity can then be based on the entire predictive distribution or, alternatively, on the expectation curve that corresponds to this distribution.

2. THE DATA GENERATING PROCESS

Let us denote the result of the drilling of the nth exploration well by the random binary response variable x_n where the outcome $x_n = 1$ represents a discovery or success and the outcome $x_n = 0$ a dry hole or failure. The probability of success when the nth exploration well is drilling is denoted by θ_n. Furthermore, in the case of a discovery, the field size is denoted by the random variable v_n. From experience we know that generally the frequency distribution of reported field sizes in a province can be adequately approximated by a lognormal probability distribution. Moreover, the use of the lognormal distribution can also be justified on a theoretical basis considering the formation, migration and accumulation of hydrocarbon deposits. So we shall assume that the logarithms of the field sizes (discovered and undiscovered) within a province follow a normal distribution with mean μ and variance σ^2. Throughout this paper we will label a probability density function (p.d.f. for short) by its arguments. Furthermore, if $\ln Z$ is normally distributed with mean μ and variance σ^2, its p.d.f. will be denoted by $No(\ln Z \mid \mu, \sigma^2)$ and the p.d.f. of Z itself by $LNo(Z \mid \mu, \sigma^2)$.

If exploration drilling were completely random with respect to longitude and latitude, then the probability of success would be approximately constant provided that the sum of the areal extents of the hydrocarbon accumulation within the province is very large but relatively small as compared to the total area of that province. In fact, the probability of success in random drilling is roughly equal to the ratio of the sum of the areal extents of all the hydrocarbon accumulations remaining to be discovered to the total area of the province remaining to be drilled.

The areal extent of many types of oil or gas reservoirs is highly correlated with the field size. So, even in random drilling, large oil or gas fields have a greater chance of being discovered than small ones. If a discovery is made in random drilling, the probability that the field size will lie between v and $v+dv$ will be proportional to v^{λ} where the value of exponent λ depends on the prevailing reservoir type in that area. In general the p.d.f. of the field size in the case of discovery is given by

$$p(v \mid \mu, \sigma^2, \lambda) = v^{\lambda} p(v \mid \mu, \sigma^2) \Big/ \int_0^{\infty} v^{\lambda} p(v \mid \mu, \sigma^2)\, dv, \tag{2.1}$$

where $p(v \mid \mu, \sigma^2) = LNo(v \mid \mu, \sigma^2)$. After some algebraic reduction, the right-hand side of (2.1) turns out to be equal to $LNo(v \mid \mu+\lambda\sigma^2, \sigma^2)$. Thus we may write

$$p(v \mid \mu, \sigma^2, \lambda) = p(v \mid \phi, \sigma^2) = LNo(v \mid \phi, \sigma^2), \tag{2.2}$$

where $\phi = \mu+\lambda\sigma^2$.

Let us now consider again scientific exploration drilling through the use of geology and geophysics. In binary data analysis, a convenient way to represent a decline in the probability of success with advancing exploration, so that the constraint $0<\theta_n<1$ is inevitably satisfied, is to postulate the linear logistic model

$$\theta_n = \{1+\exp(\alpha_1+\alpha_2 n)\}^{-1}, \quad \alpha_2 \geqslant 0, \tag{2.3}$$

where the cululative number of exploration wells n is used as a measure of exploratory effort. To represent a decline in the field size with advancing exploration, we postulate that the exponent λ in (2.1) is not constant but a linear decreasing function of the cumulative number of exploration wells, thus

$$\lambda_n = \gamma_1+\gamma_2 n, \quad \gamma_2 \leqslant 0. \tag{2.4}$$

As a result, the mean ϕ_n of the logarithm of the field size v_n in the case when the nth exploration well turns out to be a success can be written as

$$\phi_n = \mu+\gamma_1\sigma^2+\gamma_2\sigma^2 n = \beta_1+\beta_2 n, \quad \beta_2 \leqslant 0 \tag{2.5}$$

where $\beta_1 = \mu+\gamma_1\sigma^2$ and $\beta_2 = \gamma_2\sigma^2$. So, the p.d.f. of v_n is given by

$$p(v_n \mid \beta_1+\beta_2 n, \sigma^2) = LNo(v_n \mid \beta_1+\beta_2 n, \sigma^2). \tag{2.6}$$

Let R_n represent the total ultimately recoverable reserve from fields discovered by n exploration wells and r_n the contribution of the nth exploration well to this total reserve. Hence we have

$$r_n = x_n v_n \tag{2.7}$$

and

$$R_n = \sum_{i=1}^{n} r_i. \tag{2.8}$$

Thus it is assumed that the observed sequence of drilling results constitutes a random observation of the sequence of independent random variables $r_1, r_2, \ldots, r_n$ generated by the data generating process defined above. From the properties of the lognormal and the mixed lognormal distribution (see Aitchison and Brown, 1963) we have for the expectation and variance of R_n:

$$E(R_n) = \sum_{i=1}^{n} \theta_1 \xi_i \tag{2.9}$$

and

$$\operatorname{var}(R_n) = \sum_{i=1}^{n} \theta_i \xi_i^2(\omega - \theta_i), \tag{2.10}$$

where

$$\xi_i = \exp(\beta_1 + \beta_2 i + \tfrac{1}{2}\sigma^2) \tag{2.11}$$

and

$$\omega = \exp \sigma^2. \tag{2.12}$$

Furthermore, the cumulative distribution function (c.d.f. for short) of the additional reserve r_i from an exploration well with serial number i is completely specified by the parameter set

$$\boldsymbol{\psi}_i = \{i, \alpha_1, \alpha_2, \beta_1, \beta_2, \sigma^2\} \tag{2.13}$$

and is given by

$$P(r_i \mid \ _i) = \begin{cases} 0 & \text{for } r_1 < 0, \\ (1-\theta_i) + \theta_i \displaystyle\int_0^{r_i} LNo(r \mid \beta_1 + \beta_2 i, \sigma^2)\, dr & \text{for } r \geqslant 0. \end{cases} \tag{2.14}$$

The c.d.f. of R_n follows from the convolution of the c.d.f.s. of $r_1, r_2, \ldots, r_n$ but cannot be obtained in a mathematically tractable form.

3. DATA ANALYSIS

An analysis of data of past exploration drilling performance in a province must provide us with the building stones for prediction of future performance. The data required consist of the results of the exploration wells drilled. They should be listed in historical sequence, distinguishing discoveries (even if non-commerical) from dry holes. For the discoveries the amount of ultimate recoverable oil or gas should be given. The data thus assembled and "rolled back" to the time of discovery will identify the past exploration trends. Given the linear logistic model, the likelihood of the sequence of dry holes and discoveries is given by

$$L(\mathbf{x}_n \mid \boldsymbol{\alpha}) = \prod_{i=1}^{n} \theta_i^{x_i}(l-\theta_i)^{l-x_i} = \frac{\exp\left\{-\left(\alpha_1 \sum_{i=1}^{n} x_i + \alpha_2 \sum_{i=1}^{n} i x_i\right)\right\}}{\prod_{i=1}^{n} [1 + \exp\{-(\alpha_1 + \alpha_2 i)\}]}, \tag{3.1}$$

where $\mathbf{x}_n$ is the vector $(x_1, x_2, \ldots, x_n)^1$ and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^1$. In mature exploration areas we want to make our forecasts on the basis of the data only, so we use neutral prior distributions with constant density for α_1 and α_2. As a result the posterior distribution of $\boldsymbol{\alpha}$, apart from the normalizing constant, has the p.d.f.

$$p(\boldsymbol{\alpha} \mid \mathbf{x}_n) \propto \frac{\exp\{-(t_{11}\alpha_1 + t_{12}\alpha_2)\}}{\prod_{i=1}^{n} [1 + \exp\{-(\alpha_1 + \alpha_2 i)\}]}. \tag{3.2}$$

The quantities $t_{11} = \Sigma_{i=1}^{n} x_i$ and $t_{12} = \Sigma_{i=1}^{n} ix_i$ are minimal sufficient statistics for α_1 with α_2 fixed and α_2 with α_1 fixed, respectively. Hence they contain between them all the information inherent in the sequence of discoveries and dry holes. So, the posterior distribution is completely specified by these sufficient statistics and n, the number of exploration wells drilled:

$$p(\boldsymbol{\alpha} \mid \mathbf{x}_n) \equiv p(\boldsymbol{\alpha} \mid n, t_{11}, t_{12}). \tag{3.3}$$

Although $p(\boldsymbol{\alpha} \mid n, t_{11}, t_{12})$ can be readily evaluated numerically, it is, for large n, more convenient to approximate it by a bivariate normal distribution with means equal to the maximum likelihood estimates $\hat{\alpha}_{11}$ and $\hat{\alpha}_{22}$, and a dispersion matrix $D(\boldsymbol{\alpha})$ whose inverse has as elements minus the second partial derivatives of the logarithm of the likelihood function (3.1) with respect to the parameters α_1 and α_2 taken at its maximum. The validity of the appoximation follows from the large sample properties of the Bayesian posterior p.d.f.'s. Thus,

$$p(\boldsymbol{\alpha} \mid n, t_{11}, t_{12}) \approx No[\boldsymbol{\alpha} \mid \hat{\boldsymbol{\alpha}}, D(\boldsymbol{\alpha})], \tag{3.4}$$

where $No[\boldsymbol{\alpha} \mid \hat{\boldsymbol{\alpha}}, D(\boldsymbol{\alpha})]$ denotes the bivariate normal p.d.f. with mean vector $\hat{\boldsymbol{\alpha}}$ and dispersion matrix $D(\boldsymbol{\alpha})$.

The logarithm of the likelihood is given by

$$\ln L = -(t_{11}\alpha_1 + t_{12}\alpha_2) - \sum_{i=1}^{n} \ln[1 + \exp\{-(\alpha_1 + \alpha_2 i)\}], \tag{3.5}$$

and the maximum likelihood estimates $\hat{\alpha}_1$ and $\hat{\alpha}_2$ follow from the equations:

$$\left.\begin{aligned} \partial \ln L/\partial\alpha_1 &= -t_{11} + \sum_{i=1}^{n} \{1 + \exp(\alpha_1 + \alpha_2 i)\}^{-1} = 0 \\ \partial \ln L/\partial\alpha_2 &= -t_{12} + \sum_{i=1}^{n} \{1 + \exp(\alpha_1 + \alpha_2 i)\}^{-1} = 0 \end{aligned}\right\}. \tag{3.6}$$

The inverse of the dispersion matrix $D(\boldsymbol{\alpha})$ is given by

$$D^{-1}(\boldsymbol{\alpha}) = \begin{pmatrix} \sum_{i=1}^{n} \hat{\theta}_i(1-\hat{\theta}_i) & \sum_{i=1}^{n} i\hat{\theta}_i(1-\hat{\theta}_i) \\ \sum_{i=1}^{n} i\hat{\theta}_i(1-\hat{\theta}_i) & \sum_{i=1}^{n} i^2 \hat{\theta}_i(1-\hat{\theta}_i) \end{pmatrix}, \quad \text{where } \hat{\theta}_i = \{1 + \exp(\hat{\alpha}_1 + \hat{\alpha}_2 i)\}^{-1}. \tag{3.7}$$

The likelihood equations (3.6) are readily solved by a Newton–Raphson procedure.

The posterior distribution of the parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2)^1$ and the variance σ^2 in our lognormal linear model (2.6) follows directly from normal regression theory (see Raiffa and Schlaifer, 1961). First we summarize the data in the following set of sufficient statistics

$$T = \begin{pmatrix} t_{11} & t_{12} \\ t_{12} & t_{22} \end{pmatrix}, \tag{3.8}$$

where $t_{22} = \Sigma_{i=1}^{n} i^2 x_i$, and

$$\mathbf{g} = \left(\sum_{i=1}^{n} x_i \ln v_i, \sum_{i=1}^{n} ix_i \ln v_i \right) \quad (\text{with } x_i \ln v_i \equiv 0 \text{ for } x_i = 0). \tag{3.9}$$

From T and g we obtain

$$\hat{\boldsymbol{\beta}} = T^{-1}\mathbf{g}. \tag{3.10}$$

Furthermore,

$$v = t_{11} - 2 \quad \text{(redundant)} \tag{3.11}$$

and

$$s^2 = \left\{ \sum_{i=1}^{n} (x_i \ln v_i)^2 - \hat{\boldsymbol{\beta}}' T \hat{\boldsymbol{\beta}} \right\} \Big/ v \tag{3.12}$$

Then, writing the posterior distribution in terms of the precision constant $h = 1/\sigma^2$ as

$$p(\boldsymbol{\beta}, h \mid \hat{\boldsymbol{\beta}}, T, s^2, v) = p(\boldsymbol{\beta} \mid \hat{\boldsymbol{\beta}}, T, h)\, p(h \mid s^2, v), \tag{3.13}$$

we have, using non-informative priors for $\boldsymbol{\beta}$ and h,

$$p(\beta \mid \hat{\boldsymbol{\beta}}, T, h) = No[\boldsymbol{\beta} \mid \hat{\boldsymbol{\beta}}, (hT)^{-1}] \tag{3.14}$$

and

$$p(h \mid s^2, v) = Ga[h \mid s^2, v] \tag{3.15}$$

where $Ga[h \mid s^2, v]$ denotes the gamma p.d.f. with parameters s^2 and v. So all the information contained in the data concerning the parameters in our data generating process is completely summarized by

$$\mathbf{y} = \{n, T, \mathbf{g}, s^2, v\}.$$

We shall use the exploration drilling performance data shown in Fig. 3 from the province XXII to illustrate the data analysis. In Table 1 the field sizes of the successful exploration wells are listed together with their serial number of drilling. The data from the first 180 exploration wells will be used for our data analysis. Subsequently we will forecast the total reserve from the remaining 40 wells on the basis of the results of this analysis. The forecast can then be compared with the actual figure so as to judge their correspondence.

TABLE 1
Field sizes of 58 discoveries out of 220 exploration drillings

Exploration well No.	*Field size oil 10^6 BBLS*	*Exploration well No.*	*Field size oil 10^6 BBLS*	*Exploration well No.*	*Field size oil 10^6 BBLS*
3	28	56	177	141	8·8
7	26	57	43	144	29
8	775	58	33	145	450
9	114	62	178	152	5·9
11	31	71	15	161	8·8
12	337	76	22	162	49
17	41	78	11	171	100
18	113	81	8·1	174	10
20	1328	82	35	176	8·8
21	21	88	25	178	17
22	13	90	170	—	—
29	455	92	19	189	12
33	89	102	56	195	125
34	482	105	42	197	20
35	70	109	335	202	8·8
39	215	110	21	203	8·8
45	62	116	50	209	6·9
46	58	119	181	210	25
52	6·9	131	93	215	100
55	154	139	75	—	—

The information contained in the data from the first 180 exploration wells in Table 1 is completely summarized by

$$n = 180,$$
$$T = \begin{pmatrix} 50 & 3942 \\ 3942 & 454\,792 \end{pmatrix},$$
$$\hat{\boldsymbol{\beta}} = (4{\cdot}643,\ -8{\cdot}465 \times 10^{-3})',$$
$$s^2 = 1{\cdot}624 \quad \text{with } \nu = 48.$$

By solving the maximum likelihood equations (3.6) we obtain

$$\hat{\boldsymbol{\alpha}} = (0{\cdot}4277,\ 6{\cdot}075 \times 10^{-3})',$$
$$D(\boldsymbol{\alpha}) = \begin{pmatrix} 0.10285 & -8{\cdot}956 \times 10^{-4} \\ -8{\cdot}956 \times 10^{-4} & 1{\cdot}0752 \times 10^{-5} \end{pmatrix}.$$

Fig. 4 shows the 90 per cent Highest Density Region (HDR) for $\boldsymbol{\alpha}$. This figure shows also for the midpoints in a grid covering this 90 per cent HDR the actual posterior p.d.f. computed from (3.2)

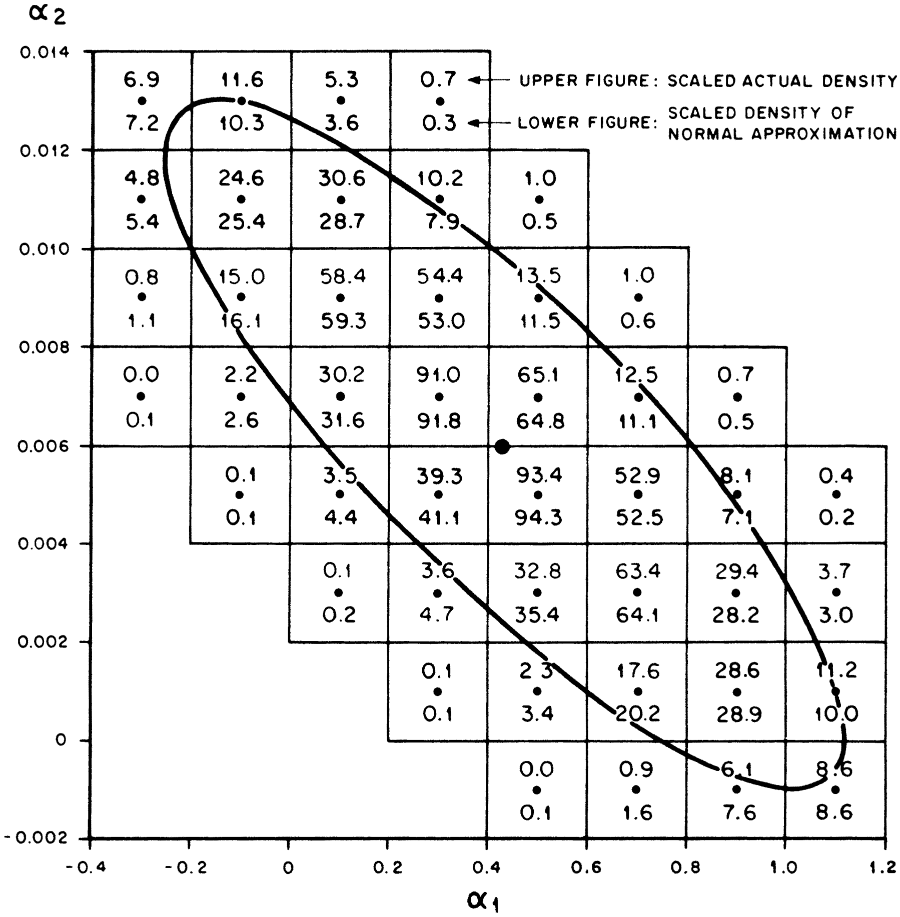

FIG. 4. Ninety per cent highest density region for α_1 and α_2 comparison of actual and approximated posterior densities of α_1 and α_2.

and the normal approximation according to (3.4). The densities shown are scaled in such a way that the density of the normal approximation at the maximum likelihood estimates equals to 100. It can be seen that the normal approximation is quite adequate. From the marginal posterior distribution of α_2 we infer that the odds in favour of a decline in the probability of success with advanced exploration are roughly 30 to 1.

In Fig. 5 we have plotted the observed probability of success in series of 10 successive exploration drillings against the mean exploration well number in each series. This figure shows also the linear logistic form fitted to the binary data so as to show the correspondence. The data

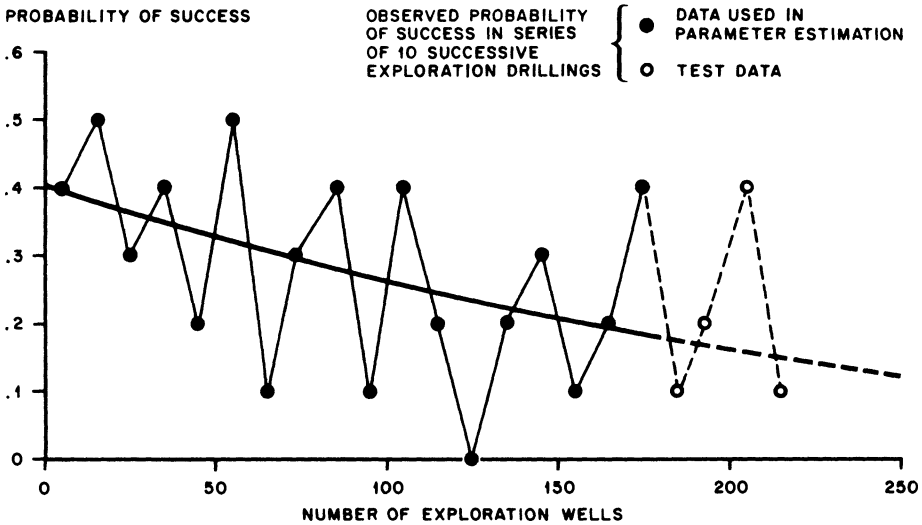

FIG. 5. Linear logistic decline of the probability of success with advancing exploration in province XX11.

analysis in our forecasting procedure includes an analysis of residuals along the lines suggested by Cox (1970). In this analysis we approximate the distribution of the number of discoveries in a series of 10 successive exploration drillings by a binomial one with a probability of success as predicted by the linear logistic form from the mean exploration well number in that series. The difference between actual and predicted number of discoveries in each series is then divided by its approximate standard error and the standardized residuals thus obtained are plotted against the mean exploration well number. This plot is used to examine the adequacy of fit by looking for systematic departures from an expected random pattern. Furthermore, we compute the ratio of the likelihoods with probability of success as observed in each series of 10 successive exploration drillings and that predicted by the linear logistic form, respectively. Two times the logarithm of this ratio has approximately a χ^2 distribution with a number of degrees of freedom equal to the number of series minus two. So, this quantity divided by its number of degrees of freedom will have an expectation equal to one, and may serve as a yardstick to assess the goodness of fit. In our examples we could not trace in the plot of standarized residuals systematic departures from a random pattern. The degree of lack of fit was equal to 1·14, which, given its distribution, is quite close to its expectation. So, the results are quite consistent with what could be expected on the basis of the assumed binary data generating process.

Fig. 6 shows the field size plotted against the exploration well number and the fitted loglinear regression line. From the marginal posterior distribution of the slope β_2 we infer that the odds in favour of a decline in the field size with advancing exploration are roughly 150 to 1. In order to check the adequacy of our loglinear normal model (2.6) the data analysis in our forecasting

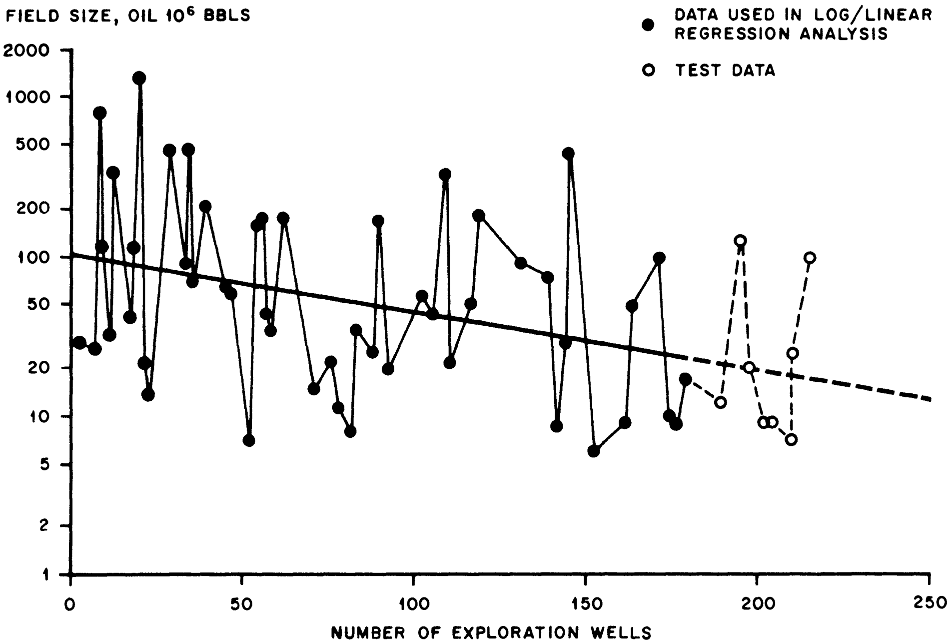

FIG. 6. Loglinear decline of the field size with advancing exploration in province XX11.

procedure includes an analysis of the so-called recursive residuals along the lines as suggested by Brown and Durbin (1968). Let $\hat{\boldsymbol{\beta}}_{k-1}$ be the coefficient vector of the fitted loglinear regression line on the basis of the first $k-1$ discoveries and let n_{k-1} refer to the serial number of the exploration drilling that yielded the $(k-1)$th discovery. According to normal regression theory the predictive distribution of the logarithm of the field size of the next discovery is normal with mean $\hat{\beta}_{1,k-1}+\hat{\beta}_{2,k-1}\,n_k$ and variance $\lambda_k\,\sigma^2$ where λ_k is given by

$$\lambda_k = 1+(k-1)^{-1}+(n_k-\bar{n}_{k-1})^2\left\{\sum_{i=1}^{k-1}(n_i-\bar{n}_{k-1})^2\right\}^{-1} \tag{3.16}$$

and $\bar{n}_{k-1} = \Sigma_{i=1}^{k-1}\,n_i/(k-1)$. The recursive residuals w_k, $k = 3,4,\dots,t_{11}$, are defined as

$$w_k = (\ln v_{n_k}-\hat{\beta}_{1,k-1}-\hat{\beta}_{2,k-1}\,n_k)/\sqrt{\lambda_k}. \tag{3.17}$$

Consequently, given our lognormal model, the w_k's are independently normally distributed with zero mean and variance σ^2. In the analysis of these recursive residuals we apply the following techniques:

(1) A normal plot of the recursive residuals together with the computation of Shapiro and Wilk's W (1965) if the number of discoveries is less than or equal to 50 or otherwise d'Agostino's D (1971), in order to assess the consistency with the assumption of lognormality.
(2) The computation of von Neumann's ratio as suggested by Phillips and Harvey (1974) in order to assess the consistency with the assumption of mutual independence.
(3) A plot of the cusum quantity of the standardized recursive residuals

$$\sum_{i=3}^{k} w_k/s, \quad k = 3,4,\dots,t_{11}$$

against k, together with appropriate control limits, in order to assess whether the loglinear relationship is constant over the exploration play.

(4) A plot of the cusum quantity of the standardized square recursive residuals

$$\sum_{i=3}^{k} w_k^2/vs^2, \quad k = 3, 4, \ldots, t_{11}$$

against k together with appropriate control limits, in order to assess whether variance σ^2 is constant over the exploration play.

Figs 7, 8 and 9 show the results of this analysis of recursive residuals for the field size data of the 50 discoveries in province XX11. The normal plot in Fig. 7 does not show serious departures from the straight line, which represents a normal distribution with zero mean and variance

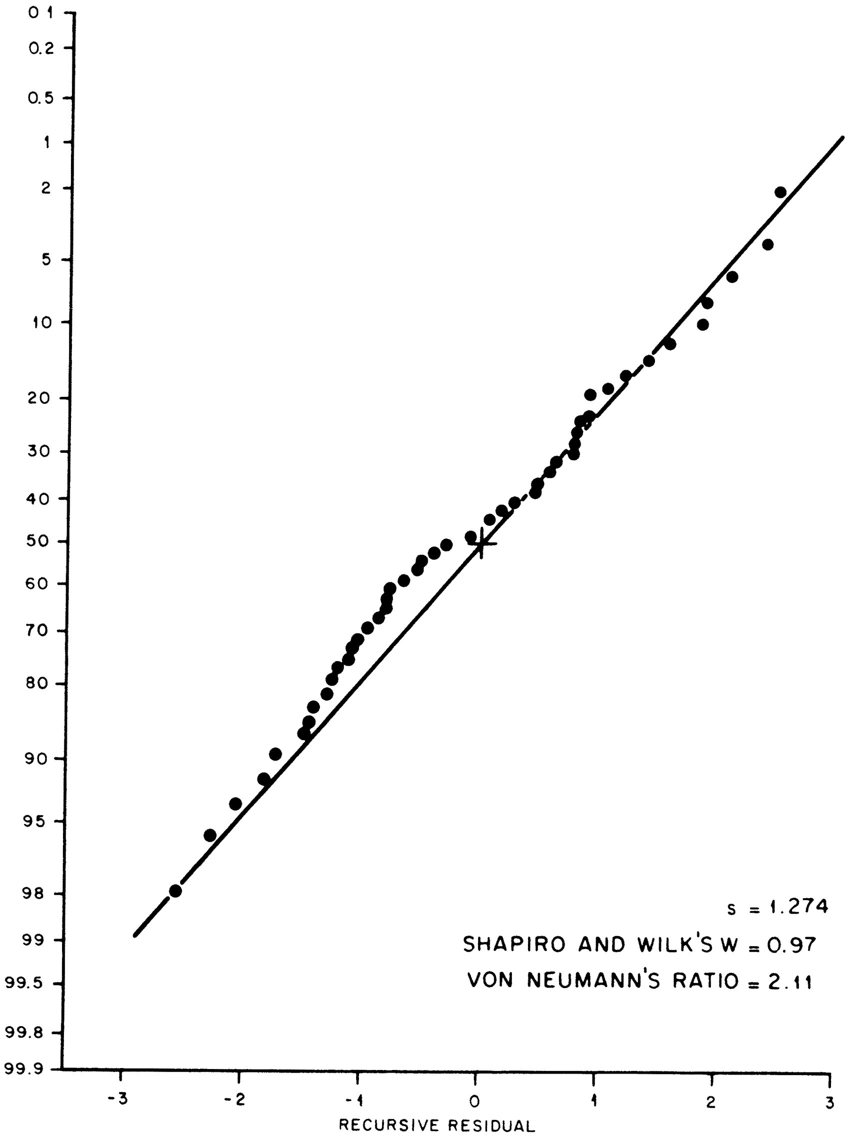

Fig. 7. Normal plot of the recursive residuals.

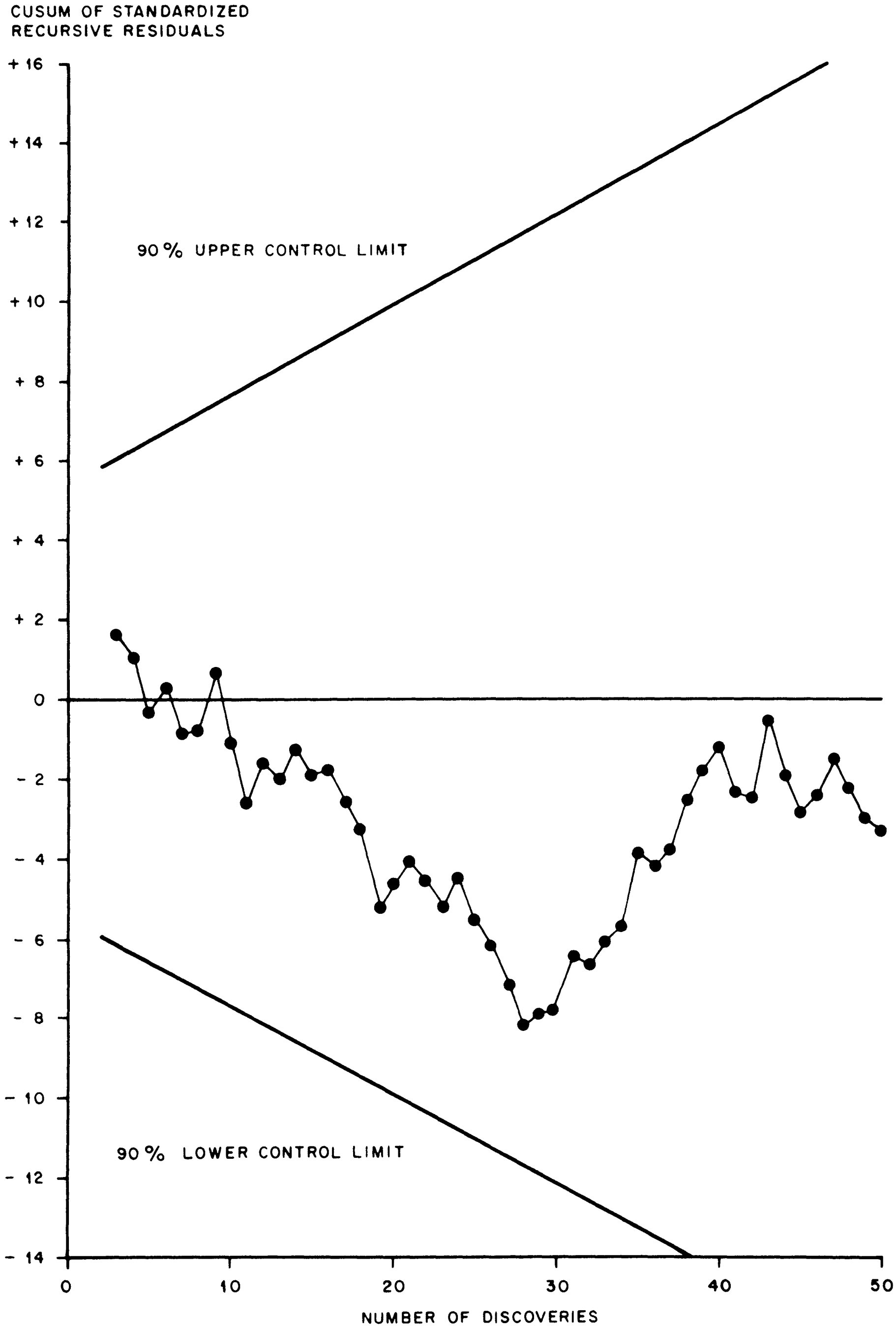

FIG. 8. Cusum plot of standardized recursive residuals.

equal to s^2. Moreover, Shapiro and Wilk's W is, given its distribution, quite close to its expectation. The same holds for the von Neumann ratio. The 90 per cent control limits in Figs 8 and 9 are drawn in such a way that the probability of the sample path crossing one or both lines is approximately 10 per cent. The slopes of the control limits for the cusum of the standardized recursive residuals in Fig. 8 are chosen such that the probability that a single point of the sample path lies outside the limits is maximum half-way between $k = 3$ and $k = t_{11}$. The proper function of the control limits is merely to provide yardsticks against which to assess the observed patterns of the sample path. It can be seen from Figs 8 and 9 that both sample paths

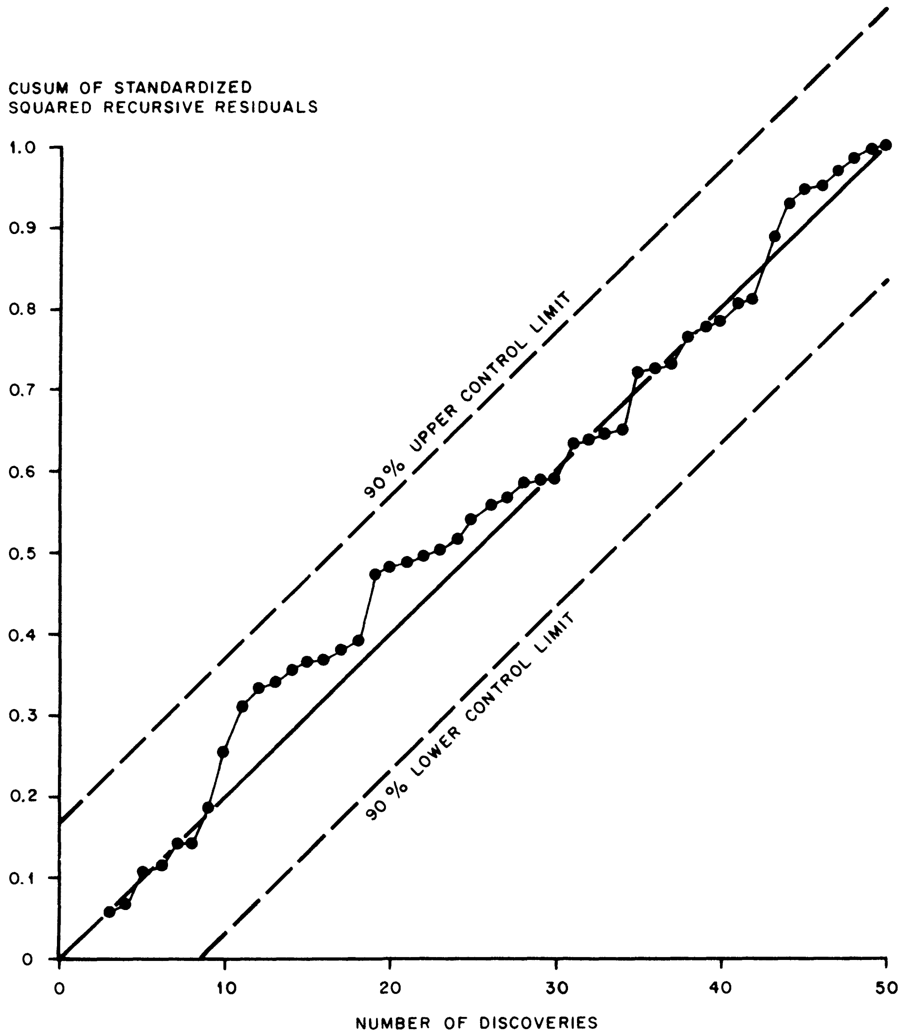

FIG. 9. Cusum plot of standardized squared recursive residuals.

look well-behaved. So, in this example the data behave quite consistently with the posed lognormal linear data generating process.

The analysis of residuals completes the data analysis. In the next section we will describe how the results of the data analysis are used to obtain the predictive distribution of the total ultimately recoverable reserve from future discoveries.

4. FORECASTING

Let m represent the number of additional exploration wells yet to be drilled and $R_{n,m}$ the total ultimately recoverable reserve from the future discoveries. Hence we have

$$R_{n,m} = \sum_{j=1}^{m} r_{n+j}. \tag{4.1}$$

According to (2.14) the c.d.f. of the additional reserve r_{n+j} from the $(n+j)$th exploration well is given by

$$P(r_{n+j}|\boldsymbol{\psi}_{n+j}) = \begin{cases} 0 & \text{for } r_{n+j}<0 \\ 1-\theta_{n+j}+\theta_{n+j}\displaystyle\int_0^{r_{n+j}} LNo[z\,|\,\beta_1+\beta_2(n+j),\sigma^2]\,dz & \text{for } r_{n+j}\geqslant 0 \end{cases} \tag{4.2}$$

where

$$\theta_{n+j} = [1+\exp\{\alpha_1+\alpha_2(n+j)\}]^{-1} \tag{4.3}$$

and

$$\boldsymbol{\psi}_{n+j} = \{n+j, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2\}. \tag{4.4}$$

Furthermore, as a result of our data analysis the respective posterior densities of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and σ^2 are given by

$$p(\boldsymbol{\alpha}\,|\,n, t_{11}, t_{12}) = No[\boldsymbol{\alpha}\,|\,\hat{\boldsymbol{\alpha}}, D(\boldsymbol{\alpha})], \tag{4.5}$$

$$p(\boldsymbol{\beta}\,|\,\hat{\boldsymbol{\beta}}, T, h) = No[\boldsymbol{\beta}\,|\,\hat{\boldsymbol{\beta}}, (hT)^{-1}] \tag{4.6}$$

and

$$p(h\,|\,s^2, v) = Ga[h\,|\,s^2, v), \tag{4.7}$$

where $h = 1/\sigma^2$.

Theoretically, the derivation of the predictive c.d.f. of $R_{n,m}$ is a straightforward process in which no particular difficulties are met. It consists of the following two steps:

(1) The c.d.f. of $R_{n,m}$ given the parameter set $\{n, m, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2\}$ follows from the convolution of the c.d.f.s of r_{n+j}, $j = 1, 2, ..., m$. The resulting c.d.f. of $R_{n,m}$ will be the mixture

$$P(R_{n,m}\,|\,n, m, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2) = \begin{cases} 0 & \text{for } R_{n,m}<0 \\ \eta_1 + \displaystyle\sum_{l=2}^{2^m} \eta_l \int_0^{R_{n,m}} p(z\,|\,l, \boldsymbol{\beta}, h^{-1})\,dz & \text{for } R_{n,m}\geqslant 0, \end{cases} \tag{4.8}$$

where η_l, $l = 1, 2, ..., 2^m$ are the successive terms in the expansion of $\Pi_{j=1}^{m}\{(1-\theta_{n+j})+\theta_{n+j}\}$ and the p.d.f. $p(z\,|\,l, \boldsymbol{\beta}, h^{-1})$ is the convolution of the p.d.f.'s $LNo[v_{n+j}\,|\,\beta_1+\beta_2(n+j), h^{-1}]$ with indices $n+j$ corresponding to the factors θ_{n+j} appearing in η_l.

(2) The η_l, $l = 1, 2, ..., 2^m$, are multiplied by the posterior p.d.f. $No[\boldsymbol{\alpha}\,|\,\hat{\boldsymbol{\alpha}}, D(\boldsymbol{\alpha})]$ and the p.d.f.s $p(z\,|\,l, \boldsymbol{\beta}, h^{-1})$, $l = 2, 3, ..., 2^m$ by the posterior p.d.f. $No[\boldsymbol{\beta}\,|\,\hat{\boldsymbol{\beta}}(hT)^{-1}]\; Ga[h\,|\,s^2, v]$. Subsequently $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and h are integrated out.

However, the resulting predictive c.d.f. $P(R_{n,m}\,|\,m, \mathbf{y})$ cannot be obtained in analytically tractable closed form. Therefore we use a Monte Carlo procedure in practical applications. An outline of this Monte Carlo procedure is given below.

Monte Carlo procedure to evaluate the predictive c.d.f. of $R_{n,m}$

Start

(1) Specify the number N of simulation runs.

(2) Specify the number n of exploration wells already drilled and the number m of exploration wells remaining to be drilled.
(3) Specify $\{\hat{\boldsymbol{\alpha}}, D(\boldsymbol{\alpha})\}$, $\{\hat{\boldsymbol{\beta}}, T\}$ and $\{s^2, v\}$.
(4) For $i = 1, 2, \ldots, N$:
  (4.1) Generate a parameter vector $\boldsymbol{\alpha}$ from the posterior p.d.f. $No[\boldsymbol{\alpha} \mid \hat{\boldsymbol{\alpha}}, D(\boldsymbol{\alpha})]$.
  (4.2) Generate a precision constant h from the posterior p.d.f. $Ga[h \mid s^2, v]$.
  (4.3) Generate a parameter vector $\boldsymbol{\beta}$ from the posterior p.d.f. $No[\boldsymbol{\beta} \mid \hat{\boldsymbol{\beta}}, (hT)^{-1}]$.
  (4.4) Set $R_{n,m}$ equal to zero.
  (4.5) For $j = 1, 2, \ldots, m$:
    (4.5.1) Compute $\theta_{n+j} = [1 + \exp\{\alpha_1 + \alpha_2(n+j)\}]^{-1}$.
    (4.5.2) Generate a value x_{n+j} of the random binary response variable with probability of success equal to θ_{n+j}.
    (4.5.3) If $x_{n+j} = 0$ then $r_{n+j} = 0$ and jump to (4.5.6), otherwise continue.
    (4.5.4) Generate a field size v_{n+j} from the p.d.f. $LNo[v_{n+j} \mid \beta_1 + \beta_2(n+j), h^{-1}]$.
    (4.5.5) $r_{n+j} = v_{n+j}$.
    (4.5.6) $R_{n,m} = R_{n,m} + r_{n+j}$.
    (4.5.7) Return to (4.5.1).
    (4.6) Save $R_{n,m}$ and the number of discoveries.
    (4.7) Return to (4.1).
(5) Group the N values of $R_{n,m}$ thus obtained and the number of discoveries into a frequency distribution.

Stop.

Figs 10, 11 and 12 show the result of the forecasting procedure for the 40 additional exploration wells in province XXII on the basis of the results of our data analysis. Fig. 10 shows the expectation curve of the total ultimate recovery, Fig. 11 the p.d.f. of the total recovery from discoveries and Fig. 12 the probability mass function of the number of discoveries. In this example 2000 simulation runs were used. The expectation curve in Fig. 10 incorporates all the available information from the past exploration trends. Relevant inferences about the prospective recovery from the 40 additional exploration wells should be based on the whole of this expectation curve.

Sometimes a simplified representation of the expectation curve may be needed in terms of a few numerical values. According to a proposed nomenclature in Shell, the expectation curve is characterized by means of the following five values:

(1) The chance factor c, which is the probability of at least one discovery, in our example $c = 0{\cdot}993$.
(2) A low value L corresponding to a probability of $\frac{1}{6}c$ that the actual ultimate recovery will be less than L, in our example $L = 75 \times 10^6$ barrels (BBLS).
(3) A middle value M corresponding to a probability of $\frac{1}{2}c$ that the actual recovery will be either more or less than M, in our example $M = 227 \times 10^6$ BBLS.
(4) A high value H corresponding to a probability of $\frac{1}{6}c$ that the actual recovery will be more than H, in our example $H = 546 \times 10^6$ BBLS.
(5) The "expectation" of the ultimate recovery defined as the chance factor multiplied by the arithemetic mean of L, M and H, in our example the expectation of the ultimate recovery $= 281 \times 10^6$ BBLS.

It must be emphasized that these five values merely serve as a simplified representation of the expectation curve. Their interpretation is directly related to the predictive distribution in the form of the expectation curve. The expectation of the ultimate recovery may be regarded as a point estimate for descriptive purposes.

To conclude this section we note that in province XXII the 40 additional exploration wells resulted in eight actual discoveries, with a total (additional) ultimate recovery of $305{\cdot}5 \times 10^6$ BBLS. These results are well in line with our forecast.

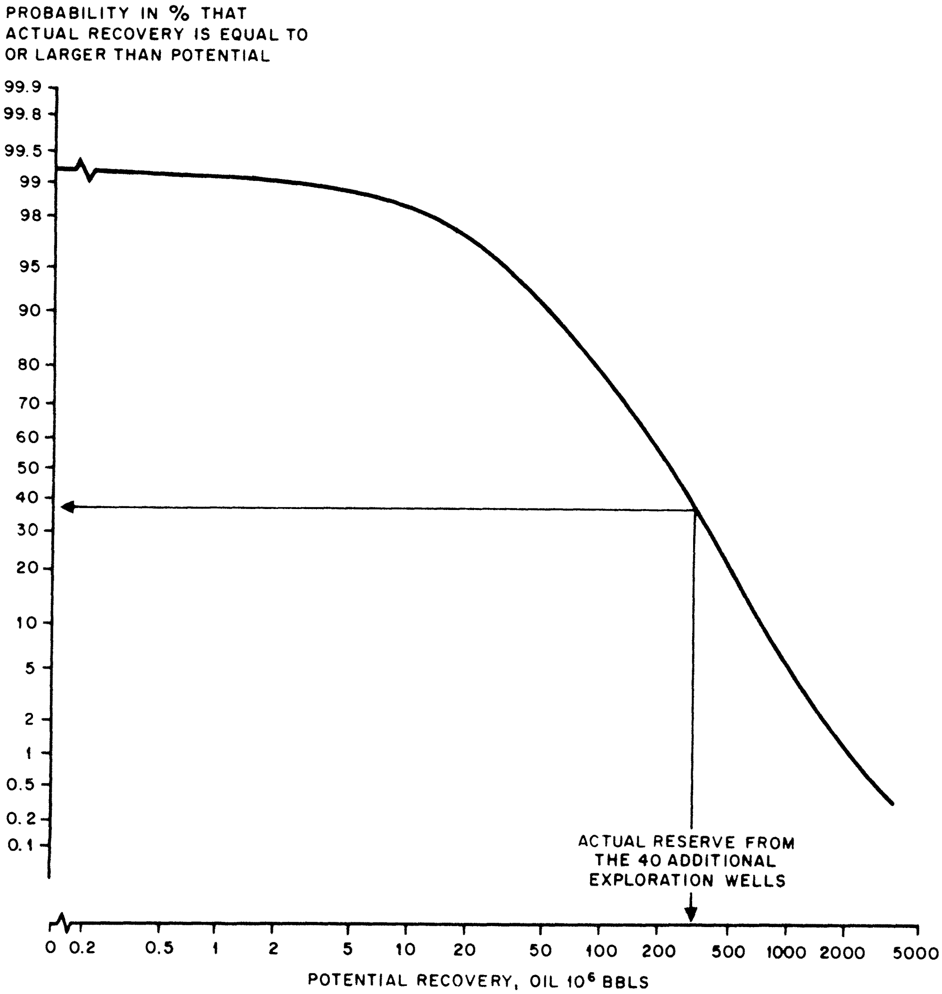

FIG. 10. Expectation curve of ultimately recoverable reserves of discoveries from 40 additonal exploration wells in province XX11.

5. DISCUSSION AND CONCLUDING REMARKS

In the preceding sections we have presented a procedure to forecast crude oil and natural gas reserves from future discoveries on the basis of the creaming phenonemon generally observed in mature petroleum provinces. Therefore we call our method the "creaming" method. Although to our knowledge the creaming method is new in its approach, forecasting procedures with the same objectives have been used in the past. A number of workers, among them Moore (1962), Zapp (1962), Hubbert (1967), Arps *et al.* (1971), Pelto (1973), Richards (1973) and Odell and Rosing (1974), have devised or utilized statistical or mathematical models to forecast hydrocarbon discoveries. These workers considered time or cumulative bore-hole length ("footage") as an indicator for forecasting whereas we employ the number of exploration wells planned to be drilled in the future as an indicator. In our view the number of exploration wells is a better indicator than time as the latter only indirectly controls future discoveries, the more direct control being the exploratory effort expended. In this respect footage as an indicator is nearly as good as the number of exploration wells but footage has the disadvantage that it does

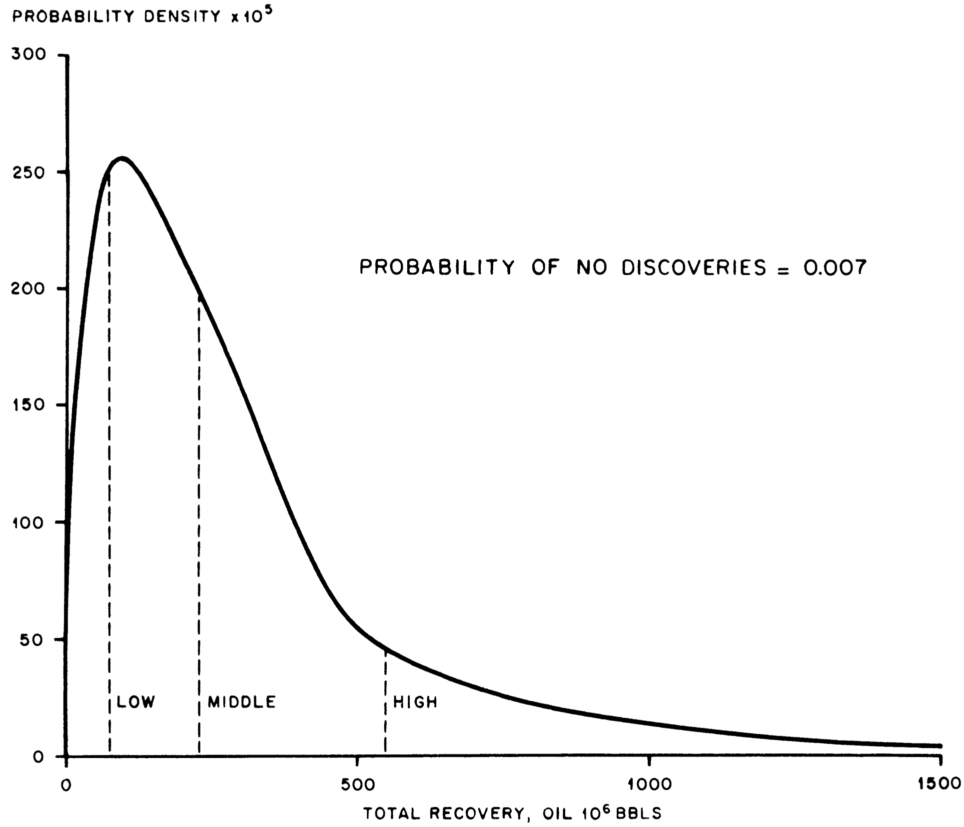

FIG. 11. Predictive probability density function of ultimately recoverable reserves of discoveries from 40 additional exploration wells in province XX11.

not allow a separate analysis of the probability of success and the field size, as in our approach. A method recently suggested by van Eek (1974) makes use of the cumulative production as an indicator. The appropriateness of this indicator is also open to question since increased production, except perhaps in the short term and in special circumstances, can hardly be considered to lead to more discoveries; in fact, provided that demand and economic incentives are there, the opposite can be said to be true. None of these methods by earlier workers yields forecast in the form of a predictive distribution and none, apart from the one proposed by Pelto (1973), gives forecasts that are interpretable in a probability context. Also noteworthy among the previous works are the empirical and theoretical studies conducted by Arps and Roberts (1958) and by Kaufman (1965) on the probability distribution of field sizes. Ryan (1973) gave a model relating discovery rate to comulative number of wells in Alberta, but his model, intended mainly for describing past discovery rates, is not readily suitable for making oil or gas discovery forecasts. Finally, we only recently became aware of a paper by Kaufman *et al.* (1975) in which a probabilistic model of oil and gas discovery is also described. They compared statistical properties of major Alberta exploration plays with properties of a Monte Carlo simulation. This model is quite different from ours and has not yet been empirically validated.

The data analysis to obtain the posterior p.d.f.'s of the parameters occurring in our model is straightforward and applies well-known techniques of maximum likelihood estimation of parameters in the linear logistic form and lognormal linear regression analysis. Since all the information from past exploration trends can be completely summarized by a set of sufficient statistics, updating the posterior p.d.f.'s of the parameters when new data became available is a

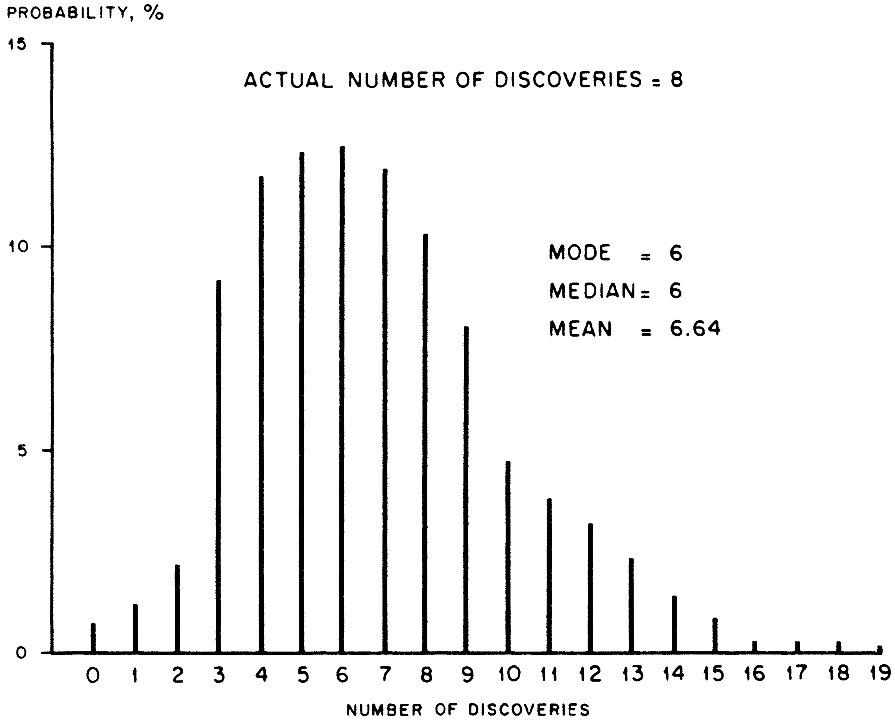

FIG. 12. Predictive probability mass function of the number of discoveries from 40 additional exploration wells in province XX11.

rather simple process (see Raiffa and Schlaifer, 1961; Walker and Duncan, 1967). In each application the adequacy of our model to describe properly the past exploration trends is assessed on the basis of an analysis of residuals. The test statistics, together with their sample distributions, are merely used as yardsticks against which the questionable discrepancies as observed in the various plots are assessed. As a result of the analysis of residuals we found that in some provinces the probability of success and/or the field size were constant or grew to a maximum before they started declining, thus indicating a "learning period". In order to cope with this kind of departure from our model we tried two different remedies: (1) to reparametrize our model to allow for the learning period and (2) to regard the learning period as irrelevant information and restrict our data analysis to the exploration trend after the learning period. The point where to start can be readily traced by running the cusum plots of standardized recursive residuals backwards through the exploration play instead of forwards. The second approach of neglecting the learning period yielded the better results. Sometimes we found that (apart from a possible incline in the early stage) past trends of both probability of success and field size are constant. Then the province is in a transition stage of exploration from immature to mature. In this case the creaming method can still be used to make a short-term forecast of the total recovery from exploration wells to be drilled in the near future, say a year.

In order to have some check on the results of the Monte Carlo procedure we compute also an estimate of the mean of the sample distribution of the total ultimate recovery $R_{n,m}$ given by

$$E(R_{n,m}) = \sum_{j=1}^{m} \frac{\exp\{\beta_1+\beta_2(n+j)+\frac{1}{2}\sigma^2\}}{1+\exp\{\alpha_1+\alpha_2(n+j)\}}. \tag{5.1}$$

To estimate the numerator $\exp\{\beta_1+\beta_2(n+j)+\frac{1}{2}\sigma^2\}$ we make use of the unbiased minimum variance Finney-type estimators for lognormal linear models derived by Bradu and Mundlak

(1970). In our applications we observed a close agreement between the point estimate of the sample distribution thus computed and the median of the predictive distribution of $R_{n,m}$. We must emphasize that this computed point estimate is used only as a check. We do not see much use in computing point estimates with specific sampling properties, such as unbiased minimum variance, minimum mean squared error or optimal relative to a squared error loss function, if they exist at all (in the convolution process of deriving the predictive distribution of $R_{n,m}$ logstudent distributions will appear which have infinite mean and variance). Inferences about the total recovery $R_{n,m}$ should be based on the entire predictive distribution which contains all the information available from the past exploration trends. In a real formal decision analysis in which future exploration effort is weighed against potential gains, the oil company's utilities are determined by many factors and cannot be approximated by relatively simple mathematical loss functions.

There is a world-wide tendency to underestimate the field sizes of recently made discoveries, which will deflate the forecasts. Therefore our procedure should make the best allowances possible for appreciation in field sizes with time, due to such causes as field extensions and improved recovery techniques, the so-called revisions. If we make the simple assumption that, on average, the revisions will enlarge the field sizes in a province by a constant factor, a solution for this problem is readily available. Systematic under- or overestimation of the field sizes in a province by a factor λ will only affect the intercept $\hat{\beta}_1$ of the loglinear regression in the following way $\hat{\beta}_1$ revised $= \hat{\beta}_1 + \ln \lambda$. So, a new expectation curve corrected for the revisions can then be readily derived.

Finally, we consider it worthwhile to add a cautionary remark regarding the application of our method. Apart from the requirement that the province for which a forecast is made should be in a mature exploration stage (or at least in a transitional stage for short-term forecasts), it is also necessary that the particular area under consideration was, at one time or another, available or accessible in its entirety for exploration. We stress the words "available" and "accessible", for there is no requirement that the area in its entirety was actually explored, only that the petroleum industry, or other agencies actively searching for oil or gas, if they chose, were in a position, technologically or in terms of permission, to do so. This is a common limitation of almost all statistical forecast procedures that try to avoid "surprises" or "freaks", but nonetheless we consider it worth a mention as it is a limitation easy to overlook and, if not heeded, may lead to misinterpretation of forecast results. In our case, the limitation follows from our use of the creaming phenomenon as an underlying concept for the forecasting procedure. Basically, the limitation means that in making forecasts we should be acutely aware of the particular geographic area or geologic domain for which we are making forecasts, by excluding from consideration those parts of the geographic area or geologic domain that were not, and could not be, explored before. Hence, for example, deep sea prospects should never be lumped with land prospects in making a forecast, and similarly prospects located in a previous concession area should be kept separate from prospects in an area that was never open to exploration. Deep prospects in an area, however, could be lumped with the shallow prospects if the industry in general had the technological capability to drill the deep prospects in the past.

We have applied the creaming method to a selected number of petroleum provinces in the world to obtain industry-wide oil discovery forecasts for the period since 1973. The potential and robustness of the method was also tested through a series of simulated forecasts using controlled situations, and practical and theoretical ramifications were investigated. The results indicate that the method should have a wide application for making oil and gas discovery forecasts for planning purposes.

ACKNOWLEDGEMENTS

We are grateful to the management of Shell Internationale Petroleum Mij. B.V. (The Hague) for permission to publish this paper. Thanks are also due to our many colleagues in the Shell organization for discussions; in particular, Dr M. H. Nederlof (SIPM) offered much useful advice and also read the manuscript.

References

AITCHISON, J. and BROWN, J. A. C. (1963). *The Lognormal Distribution*. Cambridge: University Press.

ARPS, J. J., MORTADA, M. and SMITH, A. E. (1971). Relationship between proved reserves and exploratory effort. *J. Petroleum Techn.*, 671–675.

ARPS, J. J. and ROBERTS, T. G. (1958). Economics of drilling for cretaceous oil on east flank of Denver—Julesburg basin. *Amer. Assoc. Petroleum Geologists Bull.*, **42**, 2549–2566.

BRADU, D. and MUNDLAK, Y. (1970). Estimation in lognormal linear models. *J. Amer. Statist. Ass.*, **65**, 198–211.

BROWN, R. L. and DURBIN, J. (1968). Methods for investigating whether a regression relationship is constant over time. Amsterdam: (Selected Statistical Papers from the European Meeting) Mathematical Centre Tracts No. 26.

COX, D. R. (1970). *Analysis of Binary Data*. London: Methuen.

D'AGOSTINO, R. B. (1971). An omnibus test of normality for moderate and large size samples. *Biometrika*, **58**, 341–348.

HUBBERT, M. K. (1967). Degree of advancement of petroleum exploration in the U.S. *Amer. Assoc. Petroleum Geologists Bull.*, **51**, 2207–2227.

KAUFMAN, G. M. (1965). Statistical analysis of the size distribution of oil and gas fields. SPE/AIME Symposium on Petroleum Econ. and Evaluation, Dallas Section, pp. 109–124.

KAUFMAN, G. M., BALCER, Y. and KRYT, D. (1975). A probabilistic model of oil and gas discovery. In *Methods of Estimating the Volume of Oil and Gas Resources.* (J. D. Haun, ed.). Tulsa (Oklahoma): *American Assoc. Petroleum Geologists.* (Studies in Geology No. 1.)

MOORE, C. L. (1962). *Method for Evaluating U.S. Crude Oil Resources and Projecting Domestic Crude Oil Availability.* Washington: Department Interior, Office of Oil and Gas.

ODELL, P. R. and ROSING, K. E. (1974). The North Sea oil province: a simulation model of development. *Energy Policy*, **3**, 316–329.

PELTO, C. R. (1973). Forecasting ultimate oil recovery. *Soc. Petroleum Eng. J.*, Paper No. 4261.

PHILLIPS, G. D. A. and HARVEY, A. C. (1974). A simple test for serial correlation in regression analysis. *J. Amer. Statist. Ass.*, **69**, 935–939.

RAIFFA, H and SCHAIFER, R. (1961). *Applied Statistical Decision Theory.* Boston: Harvard University.

RICHARDS, R. C. (1973). A three-dimensional model for predicting natural gas reserves and availability. *J. Canadian Petroleum Techn.*, January–March, pp. 41–46.

RYAN, J. T. (1973). An analysis of crude oil discovery rate in Alberta. *Bull. Canadian Petroleum Geology*, **21**, 219–235.

SHAPIRO, S. S. and WILK, M. B. (1965). An analysis of variance tests for normality (complete samples). *Biometrika*, **52**, 591–611.

VAN EEK, W. H. (1974). Forecasting future hydrocarbon discoveries based on a relationship between past production and reserves. *Soc. Petroleum Eng. J.*, Preprint Paper No. 4828.

WALKER, S. H. and DUNCAN, D. B. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika*, **54**, 167–179.

ZAPP, A. D. (1962). *Future Petroleum Producing Capacity of the U.S.* Washington: U.S. Geol. Survey Bull., 1142-H.

Discussion of the Paper by Mr Meisner and Mr Demirmen

Professor A. G. HAWKES (University of Swansea): In commenting on this paper I find myself operating under two handicaps which will be common to many people present tonight. First, I received the paper a few weeks ago just as examination papers were beginning to flow in, and when I had also promised to complete a long and rather complicated Ph.D. thesis examination. Secondly, I have never actually found an oil well in my life. Consequently, I shall restrict myself to making some superficial comments and suggestions, without having had the opportunity to work any of them out in detail.

The art of statistical modelling lies in obtaining a balance between a realistic representation of the physical phenomena being studied and a model which is both conceptually and computationally sufficiently simple and robust to use as a working tool. In forecasting problems especially there is a statistical creaming effect, such that it is probably necessary to work extremely hard to make any significant improvement on a good basic model. The authors of tonight's paper have presented an elegantly simple model which has been thoroughly checked in at least one province, and I assume in others which have not been reported here. More complex models probably would not do much better—and probably would do worse.

The result concerning the log normal distribution (equation (2.2)), in which the actual log normal distribution of field sizes leads to a log normal distribution of the field sizes actually found, was rather nice. However, I was rather puzzled by the assumption (2.4) concerning the λ parameter. This seems to indicate a change as exploration continues of the manner in which the probability of success depends upon field size.

It seemed to me more plausible to fix λ and to postulate that μ, the mean actual log field size, should decrease with n instead of remaining constant. In practice this makes no difference because λ could then be incorporated into the β_1 rather than into the β_2 in the model specified in equations (2.5) and (2.6). I do not think that is a practical objection, but I could not quite see the theoretical justification for it.

It appears that the mean log field size, called ϕ_n in this paper, and the log-odds ratio for the probability of success decrease in a fairly constant slope. Nevertheless, there must be a problem in deciding where to start. My inclination in modelling would be to make the model somewhat more stochastic. For example, in the model, if we have

$$\eta_n = \log\frac{(1-\theta_n)}{\theta_n},$$

where θ_n is the probability of success at the nth drilling, then the present model can be written in the form

$$\eta_n - \eta_{n-1} = \alpha_2,$$

where α_2 is a constant. I would think of adding a little extra, perhaps something like the following:

$$\eta_n - \eta_{n-1} = \alpha_2 - \alpha_3\, x_{n-1},$$

where α_3 would probably be between zero and α_2. Remember that $x_n = 1$ if the drilling is successful at the nth attempt, so that instead of having the constant slope this is slightly modified according to whether or not the drilling was successful on the previous attempt. This allows the trend to wander a little, making it slightly more sensitive to recent results, and also making the original results further back of less importance.

Similarly, for the field sizes, at the moment there is ϕ_n, which is the mean of the logarithm and thus appears in the log normal distribution.

The present model can be written essentially as follows:

$$\phi_n - \phi_{n-1} = \beta_2,$$

where β_2 is a constant. Rather vaguely, I postulate perhaps we might write + (some suitable function of recent field sizes). That would have the same kind of effect on *this* process as the extra term has on the probability of success (shown above).

In particular, one of the features of the model is the separation of looking at the probability of success at all and looking at the field size obtained if success is achieved. This helps to make the model simple, but it is one of the few assumptions that does not seem to have been checked with the data—as far as I could see, at any rate, in the paper, although it may have been done somewhere. I wonder whether this is right, and whether perhaps if there was a run of no successes that this did not provide some information about the size of the potential fields as well as whether they were zero or non-zero.

It may be that there is some kind of correlation between these two processes, the probability of success and the field size question. Perhaps a link, if there is one, might be established through this suitable function, whatever it might be, added to the expression for $\phi_n - \phi_{n-1}$.

The Monte Carlo procedure in Section 4 is a nice sensible way of handling the predictive distribution, which would be very difficult to cope with analytically. However, I wonder how large N, the number of trials, so-to-speak, has to be to provide reasonably precise information about the presumably rather important upper tail. The smooth curve in Fig. 11 must have been obtained using a very large N, or a smoothing technique such as a kernel density or histospline method. I wonder whether it would be useful to try to fit some reasonable class of distributions to describe these predictive distributions of recoverable reserves. The parameters of this class would most easily be estimated from simulated data. The idea of this is that with a fitted curve it might be possible to be rather more definite about the upper tail behaviour than is possible with direct simulation.

I was not clear about the link between the estimation procedure described in detail in the paper, based on past results, and the information available through geophysical surveys and so on about individual prospects. I assume that the former is used for strategic forward planning, with the latter used for more immediate decisions on where to drill next—although I gather from tonight's verbal presentation that perhaps it is simply used as some kind of a check on geologists' estimation. But there is no direct linking of the two things.

I cannot help feeling that the presentation of this paper is a little like watching the dance of the seven veils. The authors have taken off one veil, and that has proved quite revealing. In Section 1.3, where they talk about geophysical evidence and evaluating individual prospects and so on, I think that they have lifted the corner of a second veil and hinted at even more interesting possibilities underneath. I can see decision theorists licking their lips at the prospect. In view of the financial rewards involved, I doubt if all will be revealed just yet. Nevertheless, the authors are to be congratulated on presenting a useful and interesting paper on a matter of obvious importance. I have great pleasure in proposing the vote of thanks.

Mr A. W. Davies (British Petroleum Co. Ltd, London): Professor Hawkes has dealt with the "number crunching" aspects of the paper. I will give a few comments on the geological side. If two geologists are asked to make a report, there will be the report and two minority reports, so perhaps therefore it is fortunate that Dr Demirmen is not here tonight.

My main concern is about the definition of the "petroleum province", particularly where there is the statement (in Section 1.1) that this does not necessarily refer to a geological province. Bearing in mind unemployment in Britain at the moment, and being a geologist, I find that slightly worrying.

I accept that in the example cited the method works. I think that it works because the basin, the geological province, chosen for the example—I do not know how many other examples were tried—is a homogeneously heterogeneous basin. Where the geology will be more complex, which is the more usual, I think that the method could not work for a geological province as such, but would have to be limited to a geological horizon within a geological province. I'll try and exemplify that with respect to the U.K Continental Shelf—in the Southern North Sea in the the gas play area of the UK and of Holland virtually all the plays have the same source rock, the same reservoir rock, the same seals, the same structure, age of structural growth and time of gas migration. All these are similar, in such a situation, this system would work. Contrast with that, the Northern North Sea and the Central North Sea where it is certainly not homogeneously heterogeneous, where there are at least six sources, 15 different reservoirs and different horizons—and, if there are 15 different reservoirs there will have to be 15 seals. There are all types of structures—quite a lot more than the three examplified [by the authors]. Ages of growth, migration and re-migration: all these change the situation such that it is not possible to deal with these areas as a geological province. It would be necessary to take, say, the chalk horizons of Norway separately when this method may well work but it is not possible to throw in the chalk discoveries of Norway with the Jurassic discoveries of Norway—it is like apples and oranges.

The industry drilled quite a number of wells—33—in Norway before finding anything. There was then success and, as is the wont, when we have something we chase it, so a similar sort of play was chased. It was quite some time later that the decision was taken to drill deeper—which led to finding the new sequence. Thus, there were, in effect, two learning curves—sometimes we are rather slow in appreciating that there is more than one play.

This method may work under certain circumstances, such as in the example cited where it does appear to work. But those conditions in which it will work are thought to be rare. It is very rare now for any one company to have all the raw data available in, say, a geological province, whether it be the whole thing or just one horizon in that province. Perhaps in Nigeria Shell will have all those data, but normally oil companies are unable to get all the data in a province. Sticking by my argument that it has to be an homogeneously heterogeneous basin for the method to work, we cannot afford to have a lot of data from only one part of that basin. If all the sample or a sufficiently large aerial sample is not available, the answer may be anomalous.

Thus, the first thing is to have the data. Secondly, there has to be available the *raw* data, so that the data can be interpreted *consistently* and on a compatible basis. Equally, it is difficult to find an area where politics have not affected the oil industry, either in time or in access. Then there are the difficulties of definition. What is a discovery? Is it commerciality? If 50 million barrels [reserves] were found in say 1973 in the North Sea we would have thought that was a good find, but we would have gone on to look for something bigger. Now, however, we are looking for 50 million barrels of oil reserves. How is that well drilled in 1973 viewed? Is that the actual discovery, or is the discovery the well that has been drilled in 1980 to make sure that the 1973 discovery was right?

Turning to the size of the discovery, this indeed is difficult. It is commonly underestimated—although certainly not always, overestimation may take place for stock pushing purposes among certain other countries. Or perhaps for politics. There is a case in this regard in which a certain company was having a fight, or an acrimonious debate, with one government. It then found a large discovery in another country and, to put pressure on the first government, announced a very large discovery in that other country. It is not possible always to guarantee the figures that have been published. Of course, there are also genuine mistakes. There are cases of deliberate underestimation, not just for conservative reasons but for the following reason. In the States a bonus can be gained for a discovery on a *new* structure, so if someone can persuade the authorities that on the same main feature three separate oil or gas accumulations have been discovered he can receive three discovery bonuses.

Other factors that alter the rate of drilling include the following: a company might go into an area and not be very successful for a few years, so it marks time and hopes that someone else will be successful and once someone else is successful, the first company follows the leader. Once a company is successful, starts production and cash begins to flow, and it gets the tax dollar—or the tax pound now—it is much more prepared to drill on a geological hunch, a gut feeling or whatever if it is now spending 30 pence-£s rather than £1-£s.

In any basin, unless it is the very rare occasion when only one company is working and has the whole basin to deal with itself, I find it very difficult to see how this method can work. That is speaking as a geologist, realizing how dubious are some of the data given to our statisticians by the geologists.

One of the interesting things of this method is that the learning curve is eliminated from the statistics but when 21 years drilling in a particular country with no success is eliminated it is a very long learning curve and a great deal of data lost. For instance, Norway had 33 wells before the geologists decided that another play had better be tried, whereas on the North Slope of Alaska the third well got 10 billion barrels— and that is on the learning curve, yet it is much more mature than most basins in the world except for the Middle East and the Gulf.

My point really is that the data given to the statisticians to work on by any practical oil man are of dubious quality—sometimes highly dubious, sometimes slightly less so. It is important always to bear in mind that in going to the great and accurate extremes that one appears to with numbers the original numbers are dubious. I am sorry that I am using the word "dubious" rather often, but I think that it is the truth.

The comparison made by the authors between a lottery and the exploration business is correct. It is a useful analogy, except that in lotteries someone always wins something; in the oil business it is possible to spend $100 million and get nothing.

In summary, this mathematical approach has its value to individual companies in certain rare circumstances where they have all the data, which have all been interpreted by the companies' own people and are consistent and where they have made sure that the data deal only with a geological province which is rather simple, or with one geological horizon within one province.

Mathematically, I have no argument with the paper but I think we should be aware of the data given on which to work.

Finally, I would like to second Professor Hawkes' vote of thanks to the authors, Mr Meisner and Mr Demirmen, and particularly to Mr Meisner for presenting the paper.

The vote of thanks was passed by acclamation.

Mr G. J. A. STERN (I.C.L., London): I cannot contribute much to a theoretical evaluation of the authors' method, but I would like to be able to assess its use as an aid to management, and to do this I ask just which results from the method managers use, and how they use those results. I ask also how accurate the confidence intervals calculated have proved in practice, as I feel that what managers often want is what has been calculated here: a measure of expectation and a confidence interval. In my comments I make some assumptions as to what managers want, and suggest that cruder methods may give them what they want in a more convenient form. Firstly, the authors estimate c, the chance of making a further discovery, which for the last 40 drillings is calculated as 0·993. I suspect that the manager just wants to know whether there is a near certainty of another discovery, or a near certainty of no discovery, and all values of c between 0·05 and 0·95 are the same to him. If this is so, I would say that it seems that the graph of discoveries is tending to about 0·2 per drilling, so that the chance of one or more in 40 drillings is something of the order of $1-\exp(-8) = 0{\cdot}9997$, and that from the manager's point of view this answer does not differ from the authors' 0·993, but is calculated by a method he can believe that he understands, and can manipulate for further private calculations of his own—calculations whose nature he may not wish to discuss with a statistician. Again, a graph based on Table 1 suggests that recoveries per ten drillings are levelling out at about 75 million BBLS, or 300 for 40 drillings, close both to the forecast near the end of Section 4, and to the observed outturn.

As the authors themselves say in Section 5, and as the Seconder suggested, the "observed" recoveries are by no means accurately observed, and such a graphical method at least lets the manager make allowances for such inaccuracy, and gives him a view of the possible variance or confidence interval.

To estimate the confidence interval more closely, I regressed recoveries on a second-order polynomial in well number (grouped in tens), using for this Package X, a conversational statistical system which the Central Statistical Office were kind enough to let me use. This gave a residual standard error for groups of ten wells of 325. Approximately, this would suggest an R.S.E. of 650 for four observations, representing 40 wells, and assuming normality (which is by no means totally justified) I would take something like 300 ± 700 for the 67 per cent confidence interval. I would alter this to 0 to 1300 to allow for non-normality caused partly by the impossible nature of values below zero. I accept that something a little more scientific than this is needed, but I suggest that the final answer would still be close to this: expectation around 300, 67 per cent interval of 0 to somewhere between 1000 and 2000. I would suggest that some such method as this can be more useful in many contexts, because it is easier to see how one can make at least an approximate allowance for known inaccuracies in the data, and because it is easier to manipulate the calculations further to make other estimates, e.g. discounted values of future discoveries with confidence intervals. It may often be true too that the manager will have more confidence in such crude methods which he partly understands than he will have in methods which he cannot spare the time to get to know. I would conclude however by emphasizing that I don't know whether my picture of what managers want is valid in this case, and look to the authors for enlightenment, and also I would express the hope that their method has been or will be presented to such users in a form which makes it seem more accessible and easy to adjust. The foregoing is my own opinion and does not involve my company.

Dr F. H. HANSFORD-MILLER (Inner London Education Authority): I also would like to raise the subject of the petroleum provinces. Sub-section 1.1 of the Introduction of the paper reads as follows:

"For practical purposes, national frontiers can be considered to be province boundaries."

What, then, is the position on the North Sea continental shelf with regard to the national boundary between the United Kingdom and Norwegian oil zones? This frontier line, at present, is drawn, according to the United Kingdom–Norwegian Treaty of 1965, on the median principle. It so happens, therefore, that the line straddles the Statfjord and Frigg fields, so that in this case the national frontier cuts right across the geological provinces, about which we have heard in the discussion. The 1965 Treaty completely ignored the Norwegian Trough, of deep water up to 500 metres in places and thus not continental shelf, and I am on record (Hansford-Miller, 1980) as having pointed out that, in my view, the United Kingdom oil zone should extend eastwards to the edge of the continental shelf, i.e. to the Norwegian Trough, and if this were to happen the whole of Statfjord and Frigg would be in United Kingdom waters. How would this affect the findings of the authors? With regard to the giant Statfjord field it is estimated that 11–15 per cent is in the United Kingdom Zone, with the rest in the Norwegian Zone. If provinces in that area are taken into account surely that must affect the conclusions of the paper.

Secondly, we have recently had a decline in the estimates of future yields from the United Kingdom North Sea zone. Might this be due to the authors' forecasts, or is it possibly due to the creaming process discussed in the paper?

Mr C. J. TAYLOR (O.R. Department, British Gas Corporation): First, I would like to add my congratulations to the authors for an interesting paper. When I read the paper, what worried me most was this question of the sampling aspect of drilling. In Section 2 it states:

"If a discovery is made in random drilling, the probability that the field size will lie between v and $v+dv$ will be proportional to v^{λ}"

It was the reference to "random drilling" in the written paper that worried me because drilling is generally speaking carried out having some information from geological maps and seismic surveys. What would the effect of this be on the results? The problem has been recognized by the introduction of a λ value that changes with field size according to an empirical rule, and if this adequately covers the mathematics of the situation then I am satisfied about it.

However, secondly, there is the question that if one well is drilled and then another one is drilled close by, the second drilling might hit the same field (as the first). Under random drilling or otherwise this presumably is something that happens. I do not know what would be the effect of that on the mathematics.

Thirdly, about the authors' experience on the gas side of the subject, could Mr Meisner tell me whether he has analysed any data, in which both gas and oil are given, as measures of field size? If so, would he recommend treating the two separately, or would he try to predict some sort of bivariate distribution in forecasting?

Dr S. J. TAYLOR (Lancaster University): Tonight's proposer has suggested that the quantities $\eta_n = \log\{(1-\theta_n)/\theta_n\}$ should have more structure than the authors' model, $\eta_n - \eta_{n-1} = \alpha_2$. There is evidence in the paper favouring his suggestion. A cursory inspection of the 22 points plotted on Fig. 5 shows firstly that no consecutive pairs of the "probability of success" values are equal and secondly that the sequence of these values has negatively autocorrelated first-differences.

These two observations cast doubt on the authors' model for the η_n, although I have not yet performed significance tests. (After the meeting, I calculated the first autocorrelation coefficient to be $-0{\cdot}60$, using the 21 first-differences. This coefficient is significant at the 1 per cent level.)

Dr A. C. ATKINSON (Imperial College, London): I would like to congratulate the authors on a paper which provides an interesting example of statistical modelling that avoids over complexity but appears to produce useful answers.

I have two small comments on the data. In Fig. 5 there appears to be a see-saw effect indicative of negative correlation. On the assumption of constant θ a 2×2 table of pairs of successive results yields a χ^2 value of 0·73. There is no reason to believe that an analysis using the logistic form for decreasing probabilities would give appreciably stronger evidence of non-independence. Thus the authors' untested assumption of independence in (3.1) appears vindicated.

The first two columns of field sizes given in Table 1 appear plausible, but I am not so sure about the third column. Several of the sizes seem rounded, such as 20, 100, 125. In contrast to these there are five appearances of 8·8. Do the authors have any comments on these numbers?

Professor D. R. COX (Imperial College, London): I enjoyed this paper very much. The temptation to suggest complications to authors' models has to be resisted. Nevertheless, the exponential ultimate pessimism ($\theta_n \to 0$, $v_n \to 0$ as $n \to \infty$) of the models might for long-term forecasting be a disadvantage and extra parameters might be included. Of course, the quantities in a given "province" are finite, so that ultimately the pessimism is right, but the rate of approach assumed here is rather strong.

The analytical calculation of the predictive distribution of R_{nm}, whether by Bayesian or non-Bayesian methods, is a challenging problem. I do not doubt that the authors were correct to work numerically but I suspect that some progress could be made analytically either via asymptotic expansion of the distribution for fixed parameters, or by judicious fitting by moments, followed by perturbation to allow for errors of estimation of the parameters.

The following contributions were received in writing, after the meeting.

Mr L. A. BAXTER (Central Electricity Generating Board): The authors make use of a Monte Carlo simulation procedure to obtain values of the distribution of

$$R_{n,m} = \sum_{j=1}^{m} r_{n+j},$$

where r_{n+j} ($j = 1, 2, \ldots, m$) follows a lognormal distribution with probability θ_{n+j}, a mass of probability $1-\theta_{n+j}$ being assigned to the origin (Section 4). I would like to draw the authors' attention to a more efficient method of computing the distribution of sums of independently distributed non-negative random variables.

Cléroux and McConalogue (1976) present an algorithm for generating a cubic spline approximation to any probability distribution function $F \in C^2[0, \infty)$ (see also McConalogue, 1978). Briefly, the abscissa is divided into a series of panels of equal width, h say, and F is approximated by a cubic in each panel. This representation, together with its first and second derivatives, is continuous at each node, and hence an extremely smooth approximation is obtained. An arbitrarily close approximation can, within the limitations imposed by round-off error, be achieved by choosing h sufficiently small. The derivative of the cubic spline approximates F', and hence evaluating the Stieltjes convolution of distribution functions F and G, viz.

$$F * G(t) = \int_0^t F(t-u)\, dG(u),$$

is reduced to the straighforward task of integrating polynomials numerically. The algorithm is particularly convenient for computing recursively-defined convolutions of the form $F^{(n)}(t)$, where

$$F^{(n)}(t) = \begin{cases} F(t), & n = 1 \\ F^{(n-1)} * F(t), & n \geqslant 2. \end{cases}$$

The accuracy of the approximation does not diminish appreciably with increasing n or t.

I have used a generalization of this algorithm (McConalogue, 1980; see also Baxter, 1980) numerically to convolute a variety of probability distributions, in particular the lognormal, Weibull, gamma, inverse Gaussian and truncated normal distributions. I have found that this method is considerably more accurate than Monte Carlo simulation and is also more economical on computer time.

Mr E. A. Field (Esso Petroleum Co Ltd): The proposed procedure for forecasting oil and gas discoveries rests on assumptions made about (i) success rate and (ii) field sizes. Obviously the success rate will depend upon two factors (a) exploration efficiency and (b) area potential, and some attempt must be made to determine these. The purpose of the new procedure is to contribute towards the improvement in exploration efficiency, and this can only be determined by comparing the level of efficiency before and after the introduction of the new procedure. Cozzolino addressed this problem (Cozzolino, 1979) in his paper to the AIME. In predicting future success the dilemma is to resolve the difference between a high past exploration efficiency in a low potential search area and a low past efficiency in a high potential search area. The former implies limited future success since most fields have probably been found already, whilst the latter implies that many fields remain undiscovered.

Regarding field size distribution the choice of log normal has many precedents. The famous French study of the Algerian Sahara (Allais, 1957), even went so far as to say "As far as we can judge the log normal distribution is one of the fundamental laws of nature". For any assumption about field size distribution, even random drilling would demonstrate the phenomena called "creaming". Extensive "hindsight" analysis of exploration strategies has been reported (Haraugh *et al.*, 1977) including random drilling (Drew, 1974).

It is encouraging that work of this type is still proceeding, for in an era of energy crisis, each unit of information that can be wrung from the data carries a very high reward.

The Authors replied later, in writing, as follows.

We would first like to thank the discussants of our paper for their interest and useful comments. The vast majority of these comments are concerned with, or are related to, the structure of our model which was mainly determined by its purpose as a means of forecasting and our beliefs about the process underlying the creaming phenomenon. As outlined in the introduction, our procedure is used for strategic forward planning and aims at forecasting future discoveries from a number of exploration wells yet to be drilled and not to decide on where to drill next, as correctly understood by Professor Hawkes. We agree completely with Mr Field that the success rate depends on area potential and exploration efficiency. However, we believe also that the discovery sequence of field sizes is an important indicator of exploration efficiency, since it is in the interest of the oil industry to find the larger fields in the earlier stage of exploration. The area potential is characterized of course by the number of oil/gas fields in that area (discovered and undiscovered) and the distribution of the field sizes, which we assumed to be log normal with as parameters the mean log field size μ and the variance of the log field size σ^2. We regret that in our paper we obviously did not succeed in making clear the significance of the decline in λ as a measure of exploration efficiency since Professor Hawkes was rather puzzled by our assumption (2.4) concerning λ. Let us consider a hypothetical petroleum province with only one type of oil reservoir whose areal extent is directly proportional to its field size. In the case of random drilling the value of λ will be equal to one since, given success, the probability of obtaining a field size between v and $v+dv$ will be proportional to its areal extent which in its turn is directly proportional to v. In scientific drilling through the use of geology and geophysics, we are expected to do better. So we hope that in the early stage of exploration λ will be considerably larger than one and will decline to values below one later on. In a real petroleum province λ is unknown and varies with the type of reservoir, but some average value could be estimated by a loglinear regression of the areal extent of the field size if these data were available. An indicator of the number of oil/gas fields is the number of promising structures or trapping configurations that might contain oil or gas, of which only a fraction will turn out to be a success.

From the above it may be clear that we do not expect, as suggested by Mr Field, that our procedure will contribute towards the improvement of exploration efficiency. On the contrary, the basis of our procedure is projecting the observed past exploration efficiency into the future.

We hope also that we have now taken away the worries of Mr C. J. Taylor about random drilling. It will be clear that random drilling is only considered as a reference case against which the efficiency of scientific drilling can be judged. A successful exploration drilling is usually followed by one or more appraisal drillings to assess the field size and plan the development of the field. Hence the probability of hitting twice the same field is extremely low and would be considered as a failure in our procedure if it really did happen. In Section 1.1 the problem of forecasting oils and gas simultaneously is briefly discussed.

We greatly appreciate the comments of Professor Hawkes on the art of statistical modelling. Therefore, in our turn we are a bit puzzled by his suggestions to make the model more complicated. Writing the model for θ_n in the form he suggests yields finally:

$$\theta_n = \left\{1+\exp\left(\alpha_1+\alpha_2\, n+\alpha_3 \sum_{i=1}^{n-1} x_i\right)\right\}^{-1}$$

where he expects α_3 to be negative and in absolute value between zero and α_2. The model for θ_n in this form may be useful for one step ahead forecasts, but it is not very helpful for forecasting the sum of m step aheads when m is rather large, since except for the last term the whole sum to be forecasted is now a part of the second indicator. So, we are happy with Professor Cox's comment that the temptation to suggest complications to our model has to be resisted.

Dr S. J. Taylor seems to have a strong point in his remarks concerning the negatively auto-correlated first-differences in Fig. 5, although the comments of Mr Atkinson seem to weaken his point again. Indeed, if we assume a constant probability of success, then the data do not yield any evidence against the assumption of independence. However, we do not agree with Mr Atkinson's remark that the assumption of independence has not been tested. On p. 20 we gave a short description of our analysis of residuals and we concluded that the results of this analysis were quite consistent with what could be expected on the basis of the assumed binary data generating process. Of course, one can argue about what the most appropriate test is in this case, because we are dealing with a binary data generating process whose parameter is a non-linear function of the indicator. Let us consider the data in Fig. 5 as resulting from a binomial data generating process with sample size equal to 10 and where the probability of success follows a logistic decline with the midpoint of the number of exploration wells drilled. Let us further assume that the binomial data generating process can be adequately approximated by a normal data generating process with the same mean and variance and that an adequate approximation of the predictive distribution of a one step ahead forecast can be obtained by means of a first order linear approximation. Then we can use here too the method of the recursive residuals as described in our paper for the loglinear decline in field size, which yields a von Neumann's ratio equal to 2·65, which is not significant at the two-sided 10 per cent level. Hence we do not think that Dr S. J. Taylor's first autocorrelation coefficient constitutes the most appropriate test since no allowance is made for the heteroscedasticity of the residuals and the correlation between them introduced by the use of the common estimate $\hat{\boldsymbol{\alpha}}$.

The comments of Mr Field on area potential and exploration efficiency answer more or less the question of Professor Hawkes on the correlation between success rate and field size given success. Finally, with respect to the model structure, Professor Cox is quite right on the ultimate pessimism built into our model, but in our belief this ultimate pessimism is in agreement with the generally observed creaming phenomenon. For example, according to our model the total probability of a discovery with a field size of at least 50 10^6 BBLS when drilling the *500th* exploration well in province XX11, is roughly equal to 10^{-4}. In our belief pessimism of this order of magnitude is justified given the exploration history of this province unless additional information becomes available.

The comments of Mr Davies and Dr Hansford-Miller concern our definition of petroleum province. Mr Davies is absolutely right in pointing at the difficulty of defining a "petroleum province". The more restricted in a geological sense the more homogeneous and hence better chances that the creaming process will be observed, and can be used fruitfully for prediction. We adhere in practice as much as possible to the "play concept" where the same source rock feeds the same reservoir formation capped by the same seal. However, when this proves impracticable we may have to dilute these restrictions, usually at a price. When mixing geological apples and oranges we normally get a warning from the analysis programme that no significant creaming is observed. In such cases it becomes obviously impossible to make a forecast. This forms as if it were a built-in safety device avoiding unrealistic use of the model. In intermediate cases the degree of decline in success rate and/or field size may be so weak that the forecast also will become very vague (in terms of its predictive distribution) but may still be of some value.

Under-estimation of field sizes are dealt with outside the model. Field size "appreciation" curves can be generated in provinces and used to correct recently declared reserves in a statistical way, thus reducing the possible bias in the forecasts.

We agree with Mr Davies' conclusions that our procedure is dependent on the availability of well organized data bases and therefore not of immediate use to everybody, although the trend towards public data bases is clear. For example, the PDS system in the U.S.A. could provide much of the data required for this kind of analysis.

With respect to Dr Hansford-Miller's comments, the statement in Section 1.1 of our paper is not meant to imply that national frontiers are our preferred bounds to a petroleum province. However, when analysing the whole creaming process, the national frontiers have a significance because they may separate areas where the oil industry had great freedom to explore from others where exploration has been severely slowed down by the Government. So, although the national frontiers split a geological province rather arbitrarily, it avoids mixing apples and oranges in another sense.

Declines in estimates of future yields over the years can result from the "acounting effect", i.e. transfer of discovered oil or gas from estimated future discoveries to proven reserves. Another fairly common cause is that published estimates do not necessarily represent mean values in the sense of statistical expectations, but some other "optimistic" percentile of a predictive distribution. With increasing information these kinds of estimates become adjusted more often downward than upward.

Mr Stern asks which results from our procedure are used by management, how they use these results and whether managers have confidence in crude "on the back of an envelope" methods rather than in sophisticated procedures like ours. In Shell the usual way to present estimates of uncertain quantities such as oil and gas reserves, is by means of expectation curves as discussed in Section 1.3 of our paper. Managers can use these curves to read from them the yard-sticks which are relevant to a given decision problem. Depending on the context this may be a given percentile, the mean or the mode. In industrial practice the results of studies by geological/statistical experts (posterior distributions of parameters, expectation curves, etc.) are given to and discussed with planners (how likely is a possible scenario?) and economists (how much value can one attach to a certain area?) In this way an interface is formed between the geological/statistical experts and the decision makers.

For this kind of problem "managers" do not have the data nor the time to make even crude statistical estimates. Management employs expensive specialists to do the job of summarizing all of the available information in terms of an expectation curve showing all the uncertainties. Such an expectation curve automatically tells the manager how fuzzy the future is (see Fig. 2). In our procedure the noise inherent in the creaming process is far outweighing the inaccuracies in the data.

Finally, there were a few questions and comments on computational procedures. We are grateful to Mr Baxter for drawing our attention to a more accurate and efficient method of computing the distribution of sums of independently distributed non-negative random variables. We certainly will study the potential of this method for our problem. However, the problem is more complicated than as presented in Mr Baxter's comments. The problem is not to evaluate numerically, eq. (4.1) in our paper but the numerical evaluation of the predictive distribution of $R_{n,m}$ along the steps (1) and (2) as described in Section 4 at the bottom of p. 16. For example, one complication is that the estimates of the r_{n+j} are not mutually independent since they are all based on a common data set.

We would welcome any progress in the direction of an analytical solution of the evaluation of the predictive distribution of $R_{n,m}$. For the time being, however, we view a Monte Carlo simulation procedure followed by a simple smoothing of the resulting cumulative distribution function as an effective albeit perhaps not the most efficient solution.

References in the Discussion

Allais, M. (1957). Method of appraising economic prospects of mining exploration over large territories: Algerian Sahara case study. *Manag. Sci.*, **3**, 285–347.

Baxter, L. A. (1980). Some remarks on numerical convolution. *Commun. in Statist.* B, (to appear).

Cozzolino, J. M. (1979). Measurement and projection of exploration search efficiency. Paper delivered to the 8th Hydrocarbon Economics and Evaluation Symposium, A.I.M.E., February 1979.

Cleroux, R. and McConalogue, D. J. (1976). A numerical algorithm for recursively-defined convolution integrals involving distribution functions. *Manag. Sci.*, **22**, 1138–1146.

Drew, L. J. (1974). Estimation of petroleum exploration success and the effects of resource base exhaustion via a simulation model. US Geological Survey Bulletin No. 1328.

HANSFORD-MILLER, F. H. (1980). Territorial claims in the North Sea. *The Times*, 12 May 1980; also: North Sea boundary treaty. *The Times*, 16 June 1980.

HARBAUGH, J. W., DOVERTON, J. H. and DAVIES, J. C. (1977). *Probability Methods in Oil Exploration.* London and New York: Wiley.

McCONALOGUE, D. J. (1978). Algorithm 102. Convolution integrals involving probability distribution functions. *Computer J.*, **21**, 270–272.

——(1980). Numerical treatment of convolution integrals involving distributions with densities having singularities at the origin. *Commun. in Statist.* B, (to appear).