1660
1-2-80

DR. 460

LA-8107-MS

Informal Report

The Modeling of Off-Shore Oil Lease Data by Probit, Logit, and Tobit Models

University of California

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

LOS ALAMOS SCIENTIFIC LABORATORY
Post Office Box 1663 Los Alamos, New Mexico 87545

This report was not edited by the Technical Information staff.

Work supported by the US Geological Survey, Conservation Division.

UNITED STATES
DEPARTMENT OF ENERGY
CONTRACT W-7405-ENG. 36

LA-8107-MS
Informal Report
UC-91 92
Issued: November 1979

The Modeling of Off-Shore Oil Lease Data by Probit, Logit, and Tobit Models

George A. Milliken*
Katherine A. Ash**

*Visiting Staff Member. Department of Statistics, Kansas State University, Manhattan, KS 66502.
**Department of Statistics, Kansas State University, Manhattan, KS 66502.

THE MODELING OF OFF-SHORE OIL LEASE DATA BY PROBIT, LOGIT, AND TOBIT MODELS

by

George A. Milliken

and

Katherine A. Ash

ABSTRACT

Two quantal response models, the probit and logit models, are applied to existing bidding and production data from seven off-shore lease sales. Fifteen probit models were built using individual and combined data sets. Logit models were built for three data sets for comparison purposes. The percentage of leases correctly classified by each probit and logit model for the seven data sets was computed. In order to see if the probit and logit models order the leases from most likely to least likely to produce, the probit and logit indices were classified and the proportion of leases which were producers within the classifications was determined.

A limited dependent variable model, the tobit model, was also applied to bidding and production data and three models were built. These models were used to classify the leases as producers or nonproducers. If a lease was classified as a producer, the amount of oil production per year was estimated.

I. INTRODUCTION

The bidding practices of an oil company for off-shore oil leases in some senses reflect the degree of belief that the oil company has as to whether or not there will be production on a particular lease. Thus the bidding activity for a particular lease reflects the knowledge and degree of belief about production of all the oil companies. Therefore it should be possible to use the bidding data to order the leases of a given sale from most likely to produce to lease likely to produce.

The objective of this study is to apply quantal response models to existing bidding and production data so that the leases can be ordered and to determine if a model built from one sale's data could be used to effectively order the leases from another sale. Two quantal response models were studied, i.e., the probit model and the logit model. The probit model was used on all the provided data sets and the logit model was used on a subset of the data sets for comparison purposes.

A limited dependent variable model, the tobit model, was used to predict whether or not a lease would produce oil, and, if it would produce oil to estimate the amount of oil production per year. The application of the tobit model was merely to demonstrate its use, since it was not very successful in predicting the amount of oil production for the data sets. Since the amount bid reflected the degree of belief of all production, the failure of the tobit model resulted from considering only oil production.

This paper consists of a discussion of the data sets, the variables and transformations on the variables in order to remove some of the time dependency, the analyses by the probit, logit and tobit models and a comparison of the models for some of the data sets.

A. Bidding Activity and Production No-Production of Off-Shore Oil Leases

Before a well or series of wells is drilled on a given lease, it is not known if oil is available, but by using all available information the oil companies assess the possibility of oil and decide whether or not to bid. The amount bid surely reflects their degree of belief or degree of confidence that there is oil on a particular lease. The status oil or no-oil is not a random variable, but rather a state of nature and thus we will not assign a probability to production. Instead, the degree of confidence that oil is on a particular lease given the amount of bidding activity will be assigned. It must be kept in mind that this is an observational study and techniques which were developed for experimental studies are being used for the analysis. In that context, the experimental unit is a given lease, the stimulus is the bidding activity and the response is production or no-production. Instead of assessing probabilities of production for a lease, the degree of confidence that there is production on a given lease was assessed. The degree of confidence of production given the bidding data, DC (Production|Bidding Activity), was modeled.

As will be discussed later, the bidding data consists of several bits of information about the bids on each lease as well as generated variables. Thus the multivariable probit and logit models were used to model the degree of confidence of production as DC (Production|High Bid, number bids, number bidders, etc).

B. The Variables

Bidding data was available for leases from sales on the following dates: October 13, 1954, July 12, 1955, February 24, 1960, March 16, 1962(62A), March 13, 1962(62B), June 13, 1967 and December 15, 1970. Table 1.1 lists the variables used directly or indirectly in building the probit, logit and tobit models.

Since many of the variables were monetary and the sale dates ranged from 1954 to 1970, a technique was needed to allow comparison of these variables from year to year. Within a sale date, the maximum was found for each variable. The data for each variable within the sale were then divided by the maximum value. Hence all standardized variables ranged between 0 and 1. These standardized variables are indicated in Table 1.1 by an asterisk. The effectiveness of the standardization was evaluated visually by comparing the frequency histograms for sales 54, 60, 62A and 67 before and after standardization. It was noted that there were discrepancies from year to year in the shape of the distributions. Most of the variables listed in Table 1.1 are self-explanatory, however, a few need additional comment:

1. Geometric standard deviation = exp (standard deviation of ln bids).
2. Winner's size - 1/10 (Σ company weights), where the company weight is an ordinal number assigned to the companies as follows:

   Big 8 Company = 100
   Big 30, excluding Big 8 = 1
   Others = 0.

3. Average size of Winner = 1/10 ($\Sigma \frac{PCT}{100}$ x company weight) where PCT = percentage ownership of lease by particular company.

4. Number of Hundreds. For each lease, the weight of each company in the winning bid was available. To summarize this, the variable "number of hundreds" was created which gives the number of companies with weight 100 in the winning bid.

5. Unitized (Yes or No) Yes indicates that production from this lease was combined with production from another lease for reporting purposes.

6. Termination date - the year the lease was terminated or relinquished.

Variables 1 - 11 in Table 1.1 are the original bidding variables that were standardized before being used in a probit, logit or tobit model. Variables 12 - 13 are original variables that were not standardized. Variables 18 - 33 are generated variables.

For the probit and logit models, the Bernoulli variable production (i.e., a lease either produces or it does not) was created. If a lease had nonzero years of production or was a unitized lease it was classified as producing.

For the tobit model, sum of oil production ÷ years of production was the dependent variable. There was either zero oil production for a lease or some positive amount.

II. QUANTAL RESPONSE MODELS

A. Introduction to the Probit and Logit Models

Many experimental situations involve applying a stimulus to an experimental unit and then observing a quantal response. For example, an insect is subjected to a certain dosage of an insecticide (stimulus) and then the experimenter observes whether or not the insect dies (response). Or, an economist determines the total income (stimulus) of a family and then observes whether or not they own their home (response).

Since the response is quantal, it is assumed that the occurrence or nonoccurrence of the response depends on the intensity or level of the response. It is also assumed that for each experimental unit there is a certain level of intensity below which the response does not occur and above which the response does occur. That value of the stimulus is called the threshold or tolerance of the experimental unit. The tolerances are assumed to vary from experimental unit to experimental unit. The collection of tolerances for a population of experimental units has a distribution called the tolerance distribution.

The model for the probability that the response occurs when the value of the stimulus is x is

$$P[\text{RESPONSE} \mid X = x] = \int_{-\infty}^{x} f(x,\underline{\theta})dx \tag{1.1}$$

where $f(x,\underline{\theta})$ denotes the tolerance distribution. The integral in equation (1.1) represents the proportion of the population with a tolerance less than or equal to x.

This type of analysis depends on the assumed form of the tolerance distribution. Two different tolerance distributions are considered: the normal distribution and the logistic distribution. If the tolerance distribution is assumed to be normal, i.e.,

$$f(x,\underline{\theta}) = (2\pi\sigma^2)^{-1/2} \exp(-(x-\mu)^2/(2\sigma^2)), \tag{1.2}$$

then the analysis is called PROBIT analysis.

If the tolerance distribution is assumed to be logistic, i.e.,

$$\int_{-\infty}^{x} f(x,\underline{\theta})dx=1 \;/\; \{1 + \exp(-(\alpha + \beta x))\}, \tag{1.3}$$

then the analysis is called LOGIT analysis.

Other tolerance distributions can be used, but only the PROBIT and LOGIT models are considered in this study. The next section presents a history of the use of tolerance distributions in the analysis of quantal data.

B. History

The underlying principle of using the probit model to analyze quantal response data has been known for many years. The first use was by psychophysical investigators. Fechner (1860) discussed the frequency with which is subject correctly identified the heavier of two objects and the dependence of the correct identification on the difference between the weights of the two objects. He related the proportion of correct decisions to the corresponding difference in weight by converting the proportion to the corresponding N(0,1) deviate and then observed that there was a linear relationship between the deviate and the difference between the weights of the two objects. Müller (1879) used a type of "weighted" least squares technique to estimate both the parameters of the simple linear regression model for the standard normal deviate and the difference between the two weights.

Thomson (1919) presented an estimation technique for this problem based on maximizing the probability of the set of observations. That was essentially the method of maximum likelihood estimation, even though the maximum likelihood technique in statistics was developed much later.

Several other researchers used the probit model for psychophysical problems. In the 20's the probit model was also used for biological data in toxicity tests and insecticide studies.

Gaddum (1933) published an important report on the analysis of quantal assays in biological investigations. Bliss (1934) gave the name PROBIT to the transformation (probability unit) and in two papers (1935a, b) presented the theory of probit analysis with tables to be used for the computational problems. In the 40's several authors worked on showing that the techniques of the psychophysical sciences and the biological sciences were the same.

Finally, Finney (1947) collected this material into a book to enable researchers to carry out the computations. He then extended the methods of biological assay in his books, (1964, 1971).

C. Estimation

An experiment to which quantal analysis can be applied consists of subjecting n_i experimental units to level x_i of the stimulus and then observing the number of experimental units which respond, say r_i, $i = 1, \ldots, k$, where k is the number of different levels of x. The likelihood function is

$$L(\underline{\theta},\underline{n}) = \prod_{i=1}^{k} \left\{ \binom{n_i}{r_i} \left[\int_{-\infty}^{x_i} f(x,\underline{\theta})dx \right]^{r_i} \left[1 - \int_{-\infty}^{x_i} f(x,\underline{\theta})dx \right]^{n_i - r_i} \right\}$$

Estimates of $\underline{\theta}$, the parameters specifying the tolerance distribution are needed. For the PROBIT model the parameters are μ and σ^2 for the LOGIT† model the parameters are α and β.

(i) PROBIT model. The analysis for the probit model can be carried out by transforming to a standard normal distribution as

$$P(\text{response}|X=x) = \int_{-\infty}^{x} dN(\mu,\sigma^2) = \int_{-\infty}^{y} dN(0,1)$$

where $y = (x-\mu)/\sigma$ and $dN(\mu,\sigma^2)$ denotes the p.d.f. of a normal distribution with mean μ and variance σ^2. The variable y is called the normal equivalent deviate, Finney (1971).

Let $\hat{p}_i = r_i/n_i$ denote the proportion responding to the stimulus at level x_i and define y_i by

$$\hat{p}_i = \int_{-\infty}^{y_i} dN(0,1).$$

From the tolerance distribution, the true but unknown value of p_i is

$$p_i = \int_{-\infty}^{\frac{x_i - \mu}{\sigma}} dN(0,1).$$

Hence from $\hat{p}_i$ we get y_i which relates to x_i and the parameters of the tolerance distribution by

$$y_i = (x_i - \mu)/\sigma = \alpha + \beta x_i$$

where

$$\alpha = -\mu/\sigma \quad \text{and} \quad \beta = 1/\sigma.$$

Thus there is a linear relationship between y and x. In the days before the electronic computers, Bliss defined the probit transform as $y = 5 + \alpha + \beta x$, where 5 added so that the value of y was always positive, which was a calculator aid. The likelihood function becomes

$$L(\alpha,\beta,\underline{n}) = \prod_{i=1}^{k} \left\{ \binom{n_i}{r_i} \left[\int_{-\infty}^{\alpha+\beta x_i} dN(0,1) \right]^{r_i} \left[1- \int_{-\infty}^{\alpha+\beta x_i} dN(0,1) \right]^{n_i-r_i} \right\}.$$

The parameter estimation can be accomplished by several methods including the method of scoring and weighted nonlinear least squares.

(ii) LOGIT model. The analysis for the logit model is simpler because the integral is in closed form. Thus, the likelihood function is

$$L(\alpha,\beta,\underline{n}) = \prod_{i=1}^{k} \left\{ \binom{n_i}{r_i} \left[\frac{1}{1+e^{-\alpha-\beta x_i}} \right]^{r_i} \left[\frac{e^{-\alpha-\beta x_i}}{1+e^{-\alpha-\beta x_i}} \right]^{n_i-r_i} \right\}$$

and the maximum likelihood estimates of the parameters of the tolerance distribution may also be obtained by using one of several available numerical techniques.

D. Multiple Variable Quantal Response Models

The multiple variable quantal response models are used to model the probability of response as a function of several independent variables or several stimuli. Assume that the ith experimental unit is subjected to s stimuli in combination where the jth stimulus was at level x_{ij}. Then the probability of response is modeled by

$$P(\text{response} \mid X_1 = x_{i1}, X_2 = x_{i2}, \ldots, X_s = x_{is}) = \int_{-\infty}^{\beta_0 + \beta_1 x_{i1} + \ldots + \beta_s x_{is}} f(x)\,dx$$

$$= \int_{-\infty}^{I_i} f(x)\,dx$$

where f(x) denotes the standardized form of the tolerance distribution. The assumption is that an experimental unit will respond if the value of I_i is greater than the experimental units tolerance, i.e., the tolerance distribution is a function of the index I which has been assumed to be a linear combination of the levels of the stimuli. Nonlinear functions of the levels of the stimuli could be used to define I, but they were not considered in this study.

(i) PROBIT model. If f(x) is assumed to be normal, then

$$P(\text{response}|X_1 = x_{i1}, \ldots, X_s = x_{is}) = \int_{-\infty}^{\alpha_1 x_{i1} + \ldots + \alpha_s x_{is}} dN(\mu,\sigma^2)$$

$$= \int_{-\infty}^{\beta_0+\beta_1 x_{i1} + \ldots + \beta_s x_{is}} dN(0,1)$$

where

$$\beta_0 = \frac{-\mu}{\sigma} \text{ and } \beta_j = \frac{\alpha_j}{\sigma} \quad j = 1, \ldots, s.$$

Let Z_i denote the response of the ith experimental unit, i.e., $Z_i = 0$ if no response and $Z_i = 1$ if response. Then the likelihood function becomes

$$L(\underline{\beta}) = \prod_{i=1}^{n} \left[\int_{-\infty}^{\beta_0+\beta_1 x_{i1} + \ldots + \beta_s x_{is}} dN(0,1) \right]^{Z_i} \left[1 - \int_{-\infty}^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_s x_{is}} dN(0,1) \right]^{1-Z_i}$$

and the maximum likelihood estimates can be obtained by techniques such as the method of scoring.

(ii) LOGIT model. If f(x) is assumed to be logistic, then

$$P(\text{response}|X_1 = x_{i1}, \ldots, X_s = x_{is}) = \frac{1}{1 + e^{-\beta_0-\beta_1 x_{i1}- \ldots -\beta_s x_{is}}}$$

and the likelihood function is

$$L(\underline{\beta}) = \prod_{i=1}^{n} \left[\frac{1}{1+e^{-\beta_0-\beta_1 x_{i1}- \cdots -\beta_s x_{is}}} \right]^{Z_i} \left[\frac{e^{-\beta_0-\beta_1 x_{i1}- \cdots -\beta_s x_{is}}}{1+e^{-\beta_0-\beta_1 x_{i1}- \cdots -\beta_s x_{is}}} \right]^{1-Z_i} .$$

E. Testing the Fit of the Tolerance Distribution

After the maximum likelihood estimates of the parameters have been obtained, the model can be used to estimate the probability of response for given levels of the stimuli, $x_1, \ldots, x_s$ as

$$\hat{P}(\text{response} | x_1 \ldots x_s) = \int_{-\infty}^{\hat{\beta}_0+\hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_s x_{is}} f(\nu) d\nu$$

If n* units are subjected to level $x_1^* \ldots x_s^*$, then the expected number of responses is $E^* = n^* \hat{P}(\text{response } x_1^* \ldots x_s^*)$. The χ^2 statistic for testing the lack of fit of the tolerance distribution is $\chi_c^2 = \sum_{i=1}^{k} \frac{n_i (r_i-E_i)^2}{E_i (n_i-E_i)}$, which is distributed as a chi-square random variable with k-s degrees of freedom.

F. Building the Quantal Response Models

The multivariable logit code used to build the logit models was obtained from Nelson (1978). Some slight alterations were made in the code to suit our purposes. The multivariable probit code was written by Gladhart and Mount (1972). Both codes provided:

1) Maximum likelihood estimates of $\beta_1, \beta_2, \ldots, \beta_k$.
2) The estimated variance covariance matrix of $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$.
3) χ^2 test of $\beta_2=\beta_3= \ldots =\beta_k = 0$ [β_1 assumed to be intercept].
4) Lists $\hat{\beta}_i$/standard error $\hat{\beta}_i$, i=1 ... k.

The ratio $\hat{\beta}_i$/standard error $\hat{\beta}_i$ was used to delete variables in a stepwise fashion. Variables were deleted until a significant change in the χ^2 test was obtained. The constant was never deleted.

G. Results of the Multiple Variable Probit Model

Seven data sets were provided for which 15 different probit models were built, including the individual data sets and combinations of data sets. The list of variables and their estimated coefficients occur in Table 2.1. If the probit model is used to classify the leases as producers or nonproducers where a lease was predicted to be a producer if the index was greater than zero or if the degree of confidence was greater than .5, then the percent correct classification of the model's data set ranged from 72.3% to 79.8% with an average of 75.7% for the seven data sets. The strategy was to develop a model for a given combination of data sets and then classify the leases of those data sets used to build the model and those leases which were sold after the leases in the model's data set. The results of these classifications appear in Table 2.2. It is observed that many of the models have correctly classified more than 50% of the leases in the data sets not used to build the models.

To use the probit models more effectively, the indices were classified and then the proportion of leases which were producers within the classifications was determined. A lease was assigned a 4 if its index was between 0 and .68, a lease was assigned a 5 if its index was between .68 and 1.645, and if its index was greater than 1.645 it was assigned a 6. Table 2.3 consists of the proportion of producing leases in that sale. It is observed that for most models and sales, it would be more benificial to drill on a randomly selected lease from the 4, 5 or 6 class than to drill on a lease selected at random from the total sale.

If we restrict ourselves to leases in classes 5 and 6 then the likelihood of production greatly increases (see Table 2.4). Table 2.5 contains the proportion of producers in class 6. The results seem very good, but there are only a few leases classified as a 6.

The lack of fit of the normal tolerance distribution was examined for three of the data sets. To do this, the univariate probit code in SAS was used which provides the lack of fit statistic. For each lease, the value of the index was generated and then was classified as to their magnitude. The number falling in each class and the number of producers are in Table 2.11. The probit model was then fit to these frequencies using the mid-points of the intervals as the value of the independent variable. The last line in Table 2.11 are the significance levels of the lack of fit tests. There is an adequate fit for the 55 and 67 models but not for the 62A data. This is probabily the reason that the 62A model does not effectively order the leases for other data sets.

H. Results of the Multiple Variable Logit Model

Logit models were built for three data sets: 54, 62A and 67. Table 2.6 lists the variables and the coefficients for the final models. All three models were computed by starting with 29 variables (variables 1-13, 18-33 in Table 1.1) and eliminating variables until a significant change occurred. The constant (intercept) was never deleted from the model.

If the logit model is used to classify the leases as producers or non-producers where a lease was predicted to be a producer if the index was greater than zero or if the degree of confidence was greater than .5, then, the percentage of leases that were correctly classified when the model classified its own data ranged from 73.8% to 78.7% averaging 76%. The results of the three models classifying six data sets are in Table 2.7. In the leases not used to build the models, the models correctly classified these leases over 50% of the time in the 2/3 of the cases.

To facilitate comparison of probit and logit models the logit indices were classified in the same manner as the probit indices. A lease was assigned a 4 if its logit index was between 0 and 1.1077 (1.1077 is the 75th percentile for the logit), a 5 if its index was between 1.1077 and 2.934 (2.934 is the 95th percentile for the logit) and 6 if its index was greater than 2.934. Table 2.8 gives the proportion of producers given the index had a value of 4, 5 or 6 for each data set and model. When compared to the proportion of producers for that data (last line in Table 2.8) it is noted that for most models and sales it would be better to drill on a lease selected at random from the classes 4, 5 or 6 than on a lease selected at random from the entire sale.

Table 2.9 and 2.10 give the proportion of producers given the logit index had a value of 5 or 6 and 6, respectively. For the models classifying their own data sets, the proportion of producers is greater in the 5, 6 classes than in the 4, 5 or 6 classes. The proportion of producers is also greater in class 6 than in the 5 and 6 classes. This indicates that the logit model is effective in ordering the leases, reflecting an increasing confidence in production.

III. ANALYSIS FOR LIMITED DEPENDENT VARIABLES

A. Introduction to Tobit Analysis

Tobit analysis appears to have originated in an economic context where a variable has a limit which it assumes for a substantial proportion of the population. For the remaining portion, the variable takes on a wide range of values either above or below the limit. For example, when goods are rationed, the ration is an upper limit; some consumers take the full ration while others take less. In a survey, many households would report zero expenditures on large items such as automobiles, but among those households where such a purchase was made, considerable variability in amount of expenditure would be expected.

Tobit analysis was first described by Tobin (1958) as a hybrid of probit analysis and multiple linear regression for the analysis of economic data. Tobit analysis not only takes into account both limit and non-limit responses as does probit analysis, but also accounts for the size of non-limit responses which probit analysis does not. The tobit model is appropriate for oil-lease bidding data since for a given lease we have either zero oil production (limit response) or an observed amount of oil production (non-limit response).

B. The Limited Dependent Variable Model

Let Y_i be a limited dependent variable with a lower limit of zero (the lower limit need not be zero, but some given value, L) which is assumed for a certain proportion of the population. For the remainder of the population, the expectation of the dependent variable is assumed to be a linear combination of K independent variables $X_1, X_2, \ldots, X_k$. That is, the model can be written as

$$Y_i = \begin{cases} \underline{X}_i \underline{\beta} + e_i & \text{if RHS} > 0 \\ 0 & \text{otherwise} \end{cases}$$

where

$$\underline{X}_i = (X_{i1}, X_{i2}, \ldots, X_{ik}) \qquad i = 1, \ldots, n$$

$$\underline{\beta}' = (\beta_1, \beta_2, \ldots, \beta_k)$$

n = number of observations and

$e_i \sim NID\ (0,\sigma^2)$.

The cumulative distribution function of Y_i, given $\underline{X}_i$ is:

$$P(Y_i < y_i) = 0 \qquad \text{if } y_i < 0$$

$$P(Y_i = 0) = P(\underline{X}_i\underline{\beta} + e_i \leq 0) = P(e_i \leq -\underline{X}_i\underline{\beta}) = \Phi\left(\frac{-\underline{X}_i\underline{\beta}}{\sigma}\right) \qquad \text{if } y_i = 0$$

$$P(Y_i < y_i) = P(\underline{X}_i\underline{\beta} + e_i < y_i) = P(e_i < y_i - \underline{X}_i\underline{\beta}) = \Phi\left(\frac{y_i-\underline{X}_i\underline{\beta}}{\sigma}\right) \quad \text{if } y_i > 0$$

where

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\ e^{\frac{-t^2}{2}}\ dt.$$

C. Estimation

The sample of size n includes q observations where y is at the lower limit zero and r observations where y is above the limit. The likelihood equation is:

$$L(\underline{\beta},\sigma^2) = \prod_{i|Y_i=0} \Phi\left(\frac{-\underline{X}_i\underline{\beta}}{\sigma}\right) \prod_{i|Y_i>0} \frac{1}{\sigma}\ \phi\left(\frac{y_i-\underline{X}_i\underline{\beta}}{\sigma}\right)$$

where

$$\phi(x) = \frac{1}{\sqrt{2\Pi}} e^{\frac{-x^2}{2}}.$$

The natural logarithm of $L(\underline{\beta},\sigma^2)$ is

$$\ln L(\underline{\beta},\sigma^2) = \sum_{i|Y_i=0} \ln \Phi\left(\frac{-\underline{X}_i\underline{\beta}}{\sigma}\right) + r \ln\left(\frac{1}{\sigma}\right) - \frac{r}{2} \ln (2\Pi) - 1/2 \sum_{i|Y_i>0} \left(\frac{y_i - \underline{X}_i\underline{\beta}}{\sigma}\right)^2 .$$

The above equation can be reparameterized as

$$\ln L(\underline{\alpha},\sigma^2) = \sum_{i|Y_i=0} \ln \Phi\left(-\underline{X}_i\underline{\alpha}\right) + r \ln\left(\frac{1}{\sigma}\right) - \frac{r}{2} \ln 2\Pi - \frac{1}{2} \sum_{i|Y_i>0} \left(\frac{1}{\sigma} y_i - \underline{X}_i\underline{\alpha}\right),$$

where $\underline{\alpha} = \frac{1}{\sigma} \underline{\beta}$.

Differentiating $\ln L(\underline{\alpha},\sigma^2)$ with respect to $\alpha_1, \alpha_2, \ldots, \alpha_k$ and $\frac{1}{\sigma}$ yields k + 1 nonlinear equations. These equations were solved iteratively by Newton's method of scoring to obtain the maximum likelihood estimates of $\underline{\alpha}$ and $\frac{1}{\sigma}$ and hence of $\underline{\beta}$ and σ. The inverse of the information matrix was used to estimate the standard error of the estimates.

D. Building the Tobit Models

The code used to build the tobit models from the oil-lease data was obtained from Nelson (1976, 1978) and to suit our needs, some slight alterations were made. The code provides:

1) Maximum likelihood estimates of $\beta_1, \beta_2, \ldots, \beta_k$.
2) Standard errors for $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$.
3) The value of -2log likelihood.
4) Lists $\hat{\beta}_i$/standard error $\hat{\beta}_i$, i=1, ..., k.

The ratio $\hat{\beta}_i$/standard error $\hat{\beta}_i$ was used to delete variables in a stepwise fashion. To test the hypothesis that m variables deleted at a particular step caused a significant change in the model, the value of -2log likelihood before the variables were deleted was subtracted from the value of -2log likelihood after the variables were deleted. The resulting value has an asymptotic χ^2 distribution with m degrees of freedom. The deletion of variables continued until there was a significant reduction in χ^2 statistic.

E. Results of Tobit Models

Tobit models were built for three sale dates: 54, 62A and 67. Table 3.1 lists the maximum likelihood estimates of the coefficients of the bidding variables in the final model. The 54 and 67 tobit models were computed by starting with the final probit models for those years and eliminating variables until a significant change was obtained. The 62A tobit model was computed by starting with 29 variables (variables 1-13, 18-33 in Table 1.1 and eliminating variables until a significant change occurred. The constant (intercept) was never deleted from the model. All unitized leases were deleted from each sale since it was not known how leases were paired up.

Once the final model was determined for a particular sale date, it was desired to study how the model would classify leases from that sale date and all dates in the future. For each lease the value of $\underline{X}_i\underline{\hat{\beta}} = I_i$ was computed. If the index, I_i, was negative then it was set to zero. For each sale date and each model, 2X2 tables were tabulated: observed oil production (Yes or No) vs Predicted Oil Production (Yes or No). The proportions in Tables 3.2 and 3.3 were obtained from 2X2 tables.

The proportion of leases that were correctly classified by a data set's own model ranged from 63.3% to 77.8% averaging 71.4%. In all but one case, over 50% of the leases were correctly classified.

Table 3.3 gives the proportion of oil producers for each data set and model when the tobit had predicted oil production. In most cases it would be better to drill on a lease where the tobit model had predicted oil production than on a randomly selected lease from the entire sale.

Tables 3.4 and 3.5 give the proportion of leases where errors were made in classification for each model-data set combination. It is evident that the tobit model is much more likely to predict no oil production when there was oil production than predict oil production when there was none.

Table 3.6 are the Spearman rank correlations between predicted and observed oil production. Consideration was restricted to leases with nonzero observed oil production and nonzero predicted oil production. All the correlations except the 54 model for 54 data were essentially zero indicating that the tobit model did not perform well when predicting actual oil production.

IV. COMPARISON OF PROBIT, LOGIT AND TOBIT MODELS

The results from probit, logit and tobit analysis are compared in a number of ways. Probit and logit analyses are clearly comparable since both models are quantal response models but with different tolerance distributions. Although the tobit model is modeling essentially a quantitative response the correct classification results can be compared to the probit and logit models as well as any common variables. Logit and tobit models were built for only 3 data sets, so comparisons are limited to the common data sets.

Table 4.1 gives the common variables for the 3 models. Common variables were not indicated for 67 or 54 tobit models since these models were built from the final probit model for those data sets. Of interest is the number of variables the 62A probit and logit models share. Each model had 11 variables of which seven are the same.

The logit and probit models seem essentially equivalent in correct classification of their own data sets (Table 4.2). When classifying other data sets; in 6 of 21 cases the correct classification of the logit models exceeds the correct classifications of the probit models by over 5%. However, the probit model never exceeds the logit model by more than 5%.

The tobit model generally has a higher percentage of correct classification but the tobit model is only classifying oil production whereas the logit and probit models are classifying total production. Therefore it is difficult to compare these figures.

Table 4.3 compared $DC(\text{Production}|I = 4, 5 \text{ or } 6)$ for the probit and logit models. Here, the logit model exceeded the probit model by more than 5% in about half of the cases. The probit model exceeded the logit model by more than 5% in 3 of 21 cases. In 1/3 of the cases they are within 5% of each other.

Table 4.4 compares $DC(\text{Production}|I = 5 \text{ or } 6)$ for the probit and logit models. Here the logit model exceeded the probit model by more than 5% in over half the cases (13 of 21). The probit exceed the logit by more than 5% in 5 of 21 cases.

Table 4.5 compares $DC(\text{Production}|I = 6)$ for the probit and logit models. The logit model exceeded the probit model by more than 5% in 9 of 21 cases. The probit model exceeded the logit model by more than 5% in 4 of 21 cases.

V. SUMMARY AND FURTHER IDEAS

Seven sales of off-shore oil leases were analyzed by two quantal response models and one limited dependent variable model. Fifteen probit, three logit and three tobit models were fit to the seven data sets and combinations of the data sets. The models were used to order the leases within a sale from most likely to produce to least likely to produce.

Both the probit and logit models effectively ordered the leases, but the logit model did a more effective job of ordering the leases than the probit model based on the models built on the three common data sets.

Only oil production was used in building the tobit model and the bidding activity surely related to the degree of confidence in total production. The tobit model effectively classified the bases as producers and non-producers, but it was not able to predict oil production effectively.

It does seem promising to incorporate a total production into the tobit model by considering, say, revenue per year, which reflects the production of all products. It would also be worthwhile to change the lower limit on the limited dependent variable to a high enough level so that the very marginal producers are not considered as producers.

To improve the performance of all models other transformations could be considered, such as ranking. The ordering of the leases of future sales could possibly be ordered better if the model which has distributions of the bidding variables similar to those of the new sale was used. A procedure could be developed to use that model which matches the distributions the closest. Possibly a Kolmogorov - Smirnov type statistic could be used as a measuring stick to compare distributions of the bidding variables.

Finally, K-chotomous probit and logit models could be used so that the leases can be classified into more than two ordered classes. For example, one could use (a) no production, (b) marginal production, (c) high production, as three classes. This would be benificial since the models have the most trouble ordering the middle leases whereas they do a good job of ordering the very poor leases and the very good leases.

REFERENCES

1. G. T. Fechner, (1860). Elemente der Psychophysik. Leipzig: Breitkopf und Härtel.

2. G. E. Müller, (1879). "Veber die Maassbestimmungen des Ortisinnes der Haut mittels der Methode der richtigen und falschen Fälle." Pflug. Arch. ges. Physiol., 19, 191-235.

3. G. H. Thomson, (1919). "A direct deduction of the constant process used in the method of right and wrong cases." Physchol. Rev., 26, 454-64.

4. J. H. Gaddum, (1933). Reports on biological standards. III. Methods of biological assay depending on a quantal response. Spec. Rep. Ser. Med. Res. Coun., Lond., No. 183.

5. C. I. Bliss, (1934). "The Method of Probits." Science. 79, 38-9.

6. C. I. Bliss, (1935a). "The calculation of the dosage-mortality curve." Ann. Appl. Biol., 22, 134-67.

7. C. I. Bliss (1935b). "The comparison of dosage-mortality data." Ann. Appl. Biol., 22, 307-33.

8. D. J. Finney, (1947). Probit Analysis (1st Ed.) Cambridge: University Printing House.

9. D. J. Finney, (1971). Probit Analysis (3rd Ed.) Cambridge: University Printing House.

10. D. J. Finney, (1964). Statistical Method in Biological Assay (2nd Ed.) New York: Hafner Publishing Company.

11. F. R. Nelson, (1976). "On a General Computer Algorithm for the Analysis of Models with Limited Dependent Variables." Annals of Economic and Social Measurement, 5, 493-509.

12. F. R. Nelson, (1978). Program Documentation. Qualitative and Limited Dependent Variable Regression Models. California Institute of Technology, Pasadena, California.

13. Gladhart and Mount. Program Description for Multivariate Probit Analysis. Cornell University, Ithaca, New York.

14. J. Tobin, (1958). "Estimation of Relationships for Limited Dependent Variables." Econometrica, 26, 24-36.

Table 1.1. Variables used to build probit, logit and tobit models.

* 1. Number of bids

* 2. Number of bidders

* 3. Number of companies in winning bid (number of winners)

* 4. Median bid

* 5. High Bid

* 6. Geometric mean of bids

* 7. Geometric standard deviation of bids

* 8. Arithmetic mean of bids

* 9. Standard deviation of bids

*10. Arithmetic mean of ln(bids)

*11. Standard deviation of ln(bids)

12. Winner's size
13. Average size winner
14. Years of production
15. Sum of oil production
16. Unitized (Yes or No)
17. Termination date
18. Number of Hundreds
19. $(\text{Std. Median Bid})^2$
20. Std. Number of Winners X ln (Std. High Bid)
21. $(\text{Winner size})^2$
22. Std. Number of Winners X Winner's size
23. Average size Winner X Std. Median Bid
24. $(\text{Std. Number of Bids})^2$
25. Std. Number of Bids X Std. Number of Winners
26. Std. Number of Bids X ln(Std. High Bid)
27. Constant
28. Std. Number of Winners X Std. High Bid
29. Std. Number of Winners X Number of Hundreds
30. Std. Number of Bids X Winner's Size
31. Total Bid - High Bid
32. Number of Hundreds X ln(Std. High Bid)
33. $(\text{Std. Number of Winners})^2$

* indicates variables that were standardized.

Table 2.1 Final Probit Models for each data set and combination of data sets.

Variables	54	55	60	62A	62B	Combined 62	67
Constant	**.01412**	- 1.04585	17.9587	25.6536	-11.9019	.8601	- 1.8058
Std. No. of Bids	-26.4943	7.1901	2.5424		- 5.7571		
Std. No. of Bidders	-13.5540				5.7173	2.4393	
Std. No. of Winners		- 7.3071	18.3249	-17.6182		- 7.3349	12.6828
Std. High Bid		32.3301				- 9.0646	23.1824
Std. Geometric Mean	-24.6033	6.1497					
Std. Geometric Std. Deviation	3.5055	- 6.1723					
Std. Average Bid	45.2969		18.6039			10.7883	
Std. Std. Deviation of Bids	- 4.2134	-32.6471					-21.8666
Std. Average ln Bid			-28.3900	-21.2537	15.1419		
Std. Std. Deviation of ln Bids		6.0972	- 2.5454	- 1.7593	1.3573	.5233	1.8166
Std. Median Bid				23.3747	-10.6006		
Winner's Size	- .1014	7.7265			-40.1464		- .9863
Average Size Winner		- 2.1661	.7579		17.0434	.1352	.6344
Number of Hundreds			-15.1558		- 9.3137	- 1.4723	
Std. No. Of Winners X ln (Std. High Bid)		- 2.1784		6.5183		- 5.9446	
Number Hundreds X ln (Std. High Bid)					- .5006		
$(\text{Winner's Size})^2$			.0607	.00728			.0179
Std. No. of Winners X Winner's Size		-10.1650			30.9458	- .3288	
Ave. Size Winner X Median Bid			1.8753	- 1.8547	8.5230		
$(\text{Std. No. of Bids})^2$	29.6945			9.5754		3.1301	
Std. No. of Bids X Std. No. Winners	15.4130				- 2.4127		
Std. No. of Bids X ln (Std. High Bid)	- 6.4909	2.4079		3.4105		1.1140	.9675
$(\text{Std. No. of Winners})^2$				36.0216		6.3008	-19.3923
Std. No. of Winners X Std. High Bid		19.3879	-23.6825	-61.5377			
Std. No. of Winners X No. Hundreds					21.3990	4.4570	14.7579
Total Bid - High Bid		.00142					
Std. No. of Bids X Winners Size					5.7246		
$(\text{Std. Median Bid})^2$		35.1302	- 6.4168			- 7.3316	

70	54-60	54-62A	60-62A&B	54-60-62A	54-55-60	62A-62B-67	54-55-60-62A&B
-20.0658	- .0377	8.0235	.2538	1.0924	- .3590	- 2.6181	- 4.2231
6.8294	-15.6946	- 7.9864					
		- 3.2247		- 1.4753	- 2.6791	2.2670	
		-20.6102					
		5.3109					1.2045
	-12.0816						
3.0063					- 1.9718	- 2.9847	- 2.0308
	24.5847				3.7251		
	- 2.1611	- 2.3222		- .9446	- 1.2759	- 2.1437	- .9654
20.7361						3.3952	4.3748
					1.8607	2.6035	1.9283
- 9.7811	- 5.1103	3.7016		1.1831			
			- .4171		- .2253		- .4619
		- .30212	.3986	.2846		.0741	.3489
			- 1.0081	4.5749		- .7379	- .4405
- 1.5743		4.8494	.7787	.8770	.6608		
	.9136	- .6092	.1386				
- 1.5863	.0475		.0125	.00554	.02059		.0095
	-.6739			-12.4362		- .1837	.2596
- 4.2686	12.6772	11.2220	3.2105	3.7520	2.5954		.4459
-10.6530	7.3448		- 4.1005				
	- 3.3916			.7206		.8897	
	6.0924	32.2050	3.8627	6.3549			.5990
9.9767	-10.3434	-25.2005	- 4.2582	-6.4295	-4.6020		- 1.8943
			4.1500			2.2637	.8916
2.6360	-.3971	.2080	- .1419				
11.2683		- 3.3009					

Table 2.2 Percentage Correctly Classified Leases for Probit Models

	Data						
Model	54	55	60	62A	62B	67	70
54	77.8	30.9	51.5	42.6	36.9	31.6	40.3
55	51.1	79.8	48.5	47.2	55.8	39.2	60.5
60	54.4	52.1	75.8	50.3	51.5	66.5	55.5
62A	55.6	63.8	43.4	73.3	59.2	36.7	42.0
62B	45.6	58.5	37.4	64.1	74.8	70.9	58.0
62A & 62B	53.3	29.8	56.6	36.9	39.8	27.8	36.1
67	55.6	35.1	52.5	54.9	42.2	78.5	54.6
70	51.1	62.8	43.4	63.6	64.6	70.3	72.3
54, 60	62.2	46.8	75.8	47.7	59.2	31.6	49.6
54, 62A	70.0	59.6	44.4	70.8	43.7	35.4	53.8
60, 62A & B	57.8	58.5	65.7	69.7	67.0	63.9	58.8
54, 60, 62A	60.0	45.7	57.6	73.9	42.7	69.0	42.9
54, 55, 60	71.1	72.3	65.7	67.2	59.7	62.8	47.1
62A & B, 67	56.7	66.0	49.5	64.6	68.4	70.9	49.6
54, 55, 60, 62A & B	62.2	71.3	62.6	68.2	55.3	69.0	47.1

Table 2.3 DC (Production|I=4,5 or 6) for Probit Models

Model	Data 54	55	60	62A	62B	67	70
54	81.3	32.9	57.1	38.5	37.2	27.3	34.8
	(48)*	(85)	(77)	(179)	(188)	(139)	(89)
55	54.8	78.9	57.1	38.4	28.6	27.8	33.9
	(62)	(19)	(56)	(138)	(21)	(115)	(56)
60	60.0	29.4	77.8	31.4	39.8	42.0	34.5
	(45)	(34)	(63)	(70)	(86)	(40)	(29)
62A	58.6	50.0	55.6	61.4	41.3	25.7	36.0
	(58)	(20)	(27)	(83)	(46)	(113)	(89)
62B	66.7	37.1	100.0	100.0	71.4	00.0	33.3
	(3)	(35)	(3)	(1)	(70)	(1)	(18)
62A & B	54.7	30.4	59.1	36.4	39.8	27.7	36.4
	(86)	(92)	(88)	(195)	(206)	(155)	(118)
67	60.4	31.9	59.7	43.5	37.4	66.7	42.4
	(48)	(69)	(62)	(131)	(147)	(33)	(66)
70	87.5	43.2	55.6	50.0	66.7	41.7	70.4
	(8)	(37)	(18)	(16)	(39)	(12)	(27)
54, 60	68.2	28.3	77.8	39.7	49.1	26.3	37.5
	(44)	(46)	(63)	(151)	(116)	(133)	(64)
54, 62A	90.6	41.3	55.9	60.3	35.1	26.1	37.5
	(53)	(46)	(34)	(68)	(114)	(119)	(80)
60, 62A & B	59.7	39.0	72.7	58.1	65.2	35.7	46.4
	(62)	(41)	(55)	(74)	(46)	(42)	(69)
54, 60, 62A	63.0	32.0	62.3	66.1	39.7	44.7	38.7
	(54)	(75)	(69)	(62)	(174)	(38)	(106)
54, 55, 60	78.6	59.1	78.7	47.5	48.7	36.0	27.5
	(42)	(22)	(61)	(59)	(39)	(50)	(131)
62A & B, 67	66.7	47.4	61.0	51.8	68.1	48.3	41.8
	(45)	(38)	(41)	(56)	(47)	(29)	(98)
54, 55, 60, 62A&B	63.8	34.5	66.2	56.7	67.7	45.5	41.0
	(58)	(33)	(68)	(67)	(45)	(22)	(105)
P(Prod)	55.6	31.9	59.6	36.4	39.8	28.5	37.0
	(90)	(94)	(99)	(195)	(206)	(158)	(119)

* total number of leases with index of 4, 5 or 6.

Table 2.4 DC (Production|I = 5 or 6) for Probit Models

Model	Data 54	55	60	62A	62B	67	70
54	96.8	33.3	63.6	38.1	35.6	29.3	34.8
	(31)*	(75)	(55)	(139)	(180)	(123)	(66)
55	54.8	92.3	57.1	39.0	45.5	28.3	39.4
	(62)	(13)	(49)	(136)	(11)	(113)	(19)
60	61.1	28.6	89.8	29.8	40.0	41.2	35.7
	(36)	(28)	(49)	(47)	(60)	(34)	(28)
62A	75.0	52.9	60.0	95.5	48.4	26.7	36.8
	(20)	(17)	(20)	(22)	(31)	(105)	(87)
62B	100.0	36.0	100.0	----	82.1	----	44.4
	(2)	(25)	(2)		(28)		(9)
62A & B	52.4	30.4	57.6	36.6	39.8	27.7	36.5
	(82)	(92)	(85)	(194)	(206)	(155)	(115)
67	71.4	35.8	68.8	44.2	34.2	90.9	44.9
	(21)	53)	(48)	(104)	(117)	(11)	(49)
70	100.0	46.2	44.4	33.3	68.2	40.0	83.3
	(6)	(26)	(9)	(12)	(22)	(5)	(12)
54, 60	100.0	30.3	85.1	36.5	46.7	30.2	45.5
	(20)	(33)	(47)	(74)	(107)	(96)	(44)
54, 62A	92.0	43.8	59.1	86.4	31.2	25.9	37.3
	(25)	(32)	(22)	(22)	(77)	(112)	(75)
60, 62A & B	92.3	47.6	84.0	75.0	85.7	43.8	54.5
	(13)	(21)	(25)	(12)	(14)	(16)	(33)
54, 60, 62A	100.0	35.9	87.1	81.8	38.4	61.5	43.8
	(12)	(53)	(31)	(11)	(125)	(13)	(73)
54, 55, 60	90.9	85.7	90.3	36.4	50.0	47.8	61.9
	(11)	(7)	(31)	(22)	(4)	(23)	(21)
62A & B, 67	75.0	58.3	77.8	85.7	62.5	60.0	53.5
	(8)	(12)	(9)	(7)	(8)	(5)	(43)
54, 55, 60, 62A&B	90.9	50.0	94.4	66.7	----	100.0	55.9
	(11)	(4)	(18)	(3)		(2)	(34)

---- indicates there were no leases classified as a 5 or a 6.

* total number of leases with index of 5 or 6.

Table 2.5 DC (Production | I = 6) for Probit Models.

	Data						
Model	54	55	60	62A	62B	67	70
54	100.0	42.9	63.6	41.7	44.4	25.9	37.5
	(18)*	(42)	(30)	(89)	(99)	(85)	(40)
55	54.8	100.0	59.6	39.3	33.3	28.3	42.1
	(62)	(7)	(47)	(135)	(6)	(113)	(19)
60	64.3	31.8	91.9	27.0	44.4	39.3	35.0
	(28)	(22)	(37)	(37)	(36)	(28)	(20)
62A	84.6	45.5	64.7	100.0	60.0	28.9	38.1
	(13)	(11)	(17)	(14)	(20)	(90)	(84)
62B	100.0	37.5	100.0	----	100.0	----	75.0
	(2)	(8)	(1)		(4)		(4)
62A & B	50.6	29.2	57.7	37.0	39.9	27.7	36.4
	(77)	(89)	(78)	(192)	(203)	(155)	(107)
67	87.5	33.3	75.7	41.4	34.8	100.0	46.7
	(8)	(48)	(37)	(29)	(112)	(3)	(30)
70	100.0	62.5	80.0	50.0	66.7	00.0	100.0
	(5)	(16)	(5)	(4)	(15)	(1)	(3)
54, 60	100.0	45.0	94.1	30.2	55.6	34.7	25.0
	(20)	(20)	(34)	(43)	(45)	(49)	(16)
54, 62A	100.0	46.2	76.5	100.0	50.0	26.8	37.7
	(10)	(13)	(17)	(4)	(20)	(97)	(69)
60, 62A & B	100.0	40.0	90.9	----	100.0	100.0	63.6
	(2)	(5)	(11)		(2)	(1)	(11)
54, 50, 62A	100.0	33.3	90.0	66.7	38.7	100.0	47.3
	(12)	(48)	(10)	(3)	(119)	(2)	(55)
54, 55, 60	100.0	----	91.7	26.7	00.0	63.6	100.0
	(4)		(24)	(15)	(1)	(11)	(2)
62A & B, 67	----	00.0	77.8	----	----	----	66.7
		((2)	(9)				(9)
54, 55, 60, 62A&B	100.0	----	100.0	----	----	----	00.0
	(3)		(4)				(1)

---- indicates there were no leases classified as a 6.

* total number of leases with index of 6.

Table 2.6 Final Logit Models for Three Data Sets.

Variables	55	62A	67
Constant	-1.5887	48.788	-.4021
Std. No. of Bidders			3.8714
Std. High Bid	43.997		38.480
Std. Geometric Mean	51.738		
Std Average Bid	-96.793		
Std. Std. Deviation of Bids			-47.376
Std. Average ln Bids		-45.586	
Std. Std. Deviation ln Bids		-3.0284	4.0768
Std. Median Bid		31.131	
$(\text{Std. Median Bid})^2$	48.101		35.533
Std. No. of Winners X ln(std. high bid)		14.105	
$(\text{Winner Size})^2$		.007891	
Average Size Winner X Std. Median Bid		-2.0780	-1.4979
$(\text{Std. No. of Bids})^2$		12.057	
Std. No. of Bids Xln(Std. High Bid		5.9757	3.4013
$\text{Std. No. of Winners})^2$		44.263	
Std. No. of Winners X Std. High Bid		-145.87	
Total Bid - High Bid	-.0010829	.00054568	
Std. No. of Bid X Winner Size			.25231

Table 2.7 Percentage Leases Correctly Classified for Logit Models.

	Data						
Model	54	55	60	62A	62B	67	70
55	52.2	78.7	45.5	63.1	59.7	72.2	65.5
62A	52.2	68.1	41.4	73.8	59.7	37.3	39.5
67	53.6	71.3	47.5	65.1	57.8	76.6	65.5

Table 2.8 DC (Production | I = 4, 5, or 6) for Logit Models.

	Data						
Model	54	55	60	62A	62B	67	70
55	81.8	78.8	63.2	0	0	66.7	66.7
	(11)	(18)	(19)	(1)	(1)	(3)	(9)
62A	56.9	50.0	52.0	61.9	48.9	25.9	34.4
	(51)	(16)	(25)	(84)	(45)	(112)	(90)
67	91.7	57.1	73.3	60.0	40.0	90.0	54.1
	(12)	(21)	(15)	(15)	(25)	(10)	(37)
P(Prod)	55.6	31.9	59.6	36.4	39.8	28.5	37.0
	(90)	(94)	(99)	(195)	(206)	(158)	(119)

Table 2.9 DC (Production|I = 5 or 6) for Logit Models

Model	Data 54	55	60	62A	62B	67	70
55	81.8	81.8	72.7	0	0	66.7	57.1
	(11)	(11)	(11)	(1)	(1)	(3)	(7)
62A	73.7	44.4	69.2	88.0	57.1	27.5	36.5
	(19)	(9)	(13)	(25)	(28)	(102)	(85)
67	91.7	60.0	76.9	72.7	50.0	100.0	51.6
	(12)	(15)	(13)	(11)	(12)	(8)	(31)

Table 2.10 DC (Production|I = 6) for Logit Models

Model	Data 54	55	60	62A	62B	67	70
55	85.7	100.0	75.0	0	0	66.7	50.0
	(7)	(5)	(8)	(1)	(1)	(3)	(4)
62A	76.9	40.0	75.0	88.0	71.4	32.5	37.2
	(13)	(5)	(12)	(12)	(14)	(77)	(78)
67	88.9	70.0	77.8	71.4	33.3	100.0	57.7
	(9)	(10)	(9)	(7)	(6)	(5)	(26)

Table 2.11 Test for Lack of Fit for Selected Probit Models.

Lower limit of class	Data Set					
	55		62A		67	
	No. Prod	No. Leases	No. Prod	No. Leases	No. Prod	No. Leas
-2.140	2	19	1	21	2	26
-1.060	6	27	1	15	5	38
- .680	1	10	5	15	7	30
- .385	3	13	7	20	6	14
- .125	3	5	14	28	7	17
.125	1	4	19	33	8	12
.385	0	0	12	21	3	7
.680	3	3	4	7	4	5
1.060	4	5	7	8	5	5
2.140	3	3	14	27	3	4
Chi-square stat.	7.11		18.46		8.41	
d.f.	7		8		8	
sign. level	.4172		.0180		.3941	

Table 3.1 Final Tobit Models for Three Data Sets

Variables	54	62A	67
Std. No. of Bids	46284510.0	6401425.5	
Std. No. of Bidders	-40886239.0		
Std. No. of Winners		-14189765.0	
Std. Std. Deviation of Bids	-29151631.0	- 5140065.9	
Std. Median Bid		9191196.7	
Std. High Bid			348729.3
Average Size Winner		133693.96	
Std. Average Bid	41806740.0		
Winner Size	800040.93		-204772.09
Std. Geometric Std. Deviation	39488014.0		
Constant	-24869666.0	1561123.0	-1222458.8
Average Size Winner X Std. Median Bid		-1087037.8	
Std. No. of Bids X Std. No. Of Winners		-9305762.2	
Std. No. of Bids X ln(Std. High Bid)		2260815.7	
(Std. No. of Winners)^2		15121563.0	
(Winner Size)^2			9066.0856
Std. No. of Winners X No. of Hundreds			2579696.2

Table 3.2 Percentage of Leases Correctly Classified for Tobit Models

Model	54	55	60	62A	62B	67	70
54	63.3	67.8	56.4	57.1	71.8	44.4	76.5
62A	52.2	80.0	58.5	73.1	77.3	52.9	57.9
67	58.9	76.7	68.1	62.6	71.3	77.8	75.6

Table 3.3 Tobit Models DC (Oil Production|Predicted Oil Production)

Model	54	55	60	62A	62B	67	70
54	70.0	27.3	59.4	38.5	34.8	26.6	66.7
	(40)	(33)	(32)	(91)	(23)	(94)	(3)
62A	100.0	33.3	75.0	72.2	60.0	28.8	34.3
	(6)	(12)	(16)	(18)	(15)	(73)	(67)
67	87.5	37.9	68.9	23.8	33.3	81.8	50.0
	(16)	(29)	(45)	(21)	(24)	(11)	(28)
P(Oil Prod)	54.4	15.6	50.0	31.3	24.3	26.8	24.4
	(90)	(90)	(94)	(182)	(181)	(153)	(119)

Table 3.4 DC (Predicted Oil Production|No Oil Production)

Model	54	55	60	62A	62B	67	70
54	29.3	31.6	27.7	44.8	10.9	61.6	1.1
62A	00.0	10.5	8.5	4.0	4.4	46.4	48.9
67	4.9	23.7	29.8	12.8	11.7	1.8	15.6

Table 3.5 DC (Predicted No Oil Production|Oil Production)

Model	54	55	60	62A	62B	67	70
54	42.9	35.7	59.6	38.6	81.8	39.0	93.1
62A	87.8	71.4	74.5	77.2	79.5	48.8	20.7
67	71.4	21.4	34.0	91.2	81.8	78.0	51.7

Table 3.6 Spearman Correlation Between Observed Oil Production and Predicted Oil Production [1]

Model	Data 54	55	60	62A	62B	67	70
	.321[2]	-.267	.075	.045	.310	.001	----
54	.10[3]	.49	.76	.80	.46	.997	----
	28[4]	9	19	35	8	25	2
	49[5]	14	47	57	44	41	29
62A	-.086	0.00	.070	.149	.05	-.164	.047
	.87	1.000	.83	.63	.90	.48	.83
	5	4	12	13	9	21	23
	49	14	47	57	44	41	29
67	-.084	.073	.057	.071	-.20	.017	.059
	.78	.83	.76	.87	.75	.97	.84
	14	11	31	5	8	9	14
	49	14	47	57	44	41	29

1 Only for leases where a nonzero oil production was predicted and nonzero oil production was observed

2 Estimate of Spearman Correlation

3 Significance level

4 Number of leases which has nonzero predicted and observed production

5 Total number of oil producing leases

Table 4.1 Variables in common for probit, logit, and tobit models

55
Std. High (P, L)[1]
Std. Geometric Mean (P, L)
$(\text{Std. Median Bid})^2$ (P, L)
Total Bid - High Bid (P, L)

62A
Std. Average ln Bid (P, L)
Std. Std. Deviation ln Bid (P, L)
Std. Median Bid (P, L)
$(\text{Std. No. of Bids})^2$ (P, L)
Std. No. of Winners X ln(Std. High Bid) (P, L)
$(\text{Winner's Size})^2$ (P, L)
Std. No. of Winners X Std. High Bid (P, L)
Std. No. of Winners (T, P)
Average Size Winner X Std. Median Bid (T, P, L)
Std. No. of Bids X ln (Std. High Bid) (T, P, L)
$(\text{Std. No. of Winners})^2$ (T, P, L)

67
Std. Std. Deviation of ln Bids (P, L)
Std. No. Bids X ln(Std. High Bids) (P, L)

[1] Letters in parentheses indicates the models in which the variable occurred, P is Probit, L is Logit, and T is Tobit.

Table 4.2 Percentage of Correctly Classified Leases for Probit, Tobit, and Logit Models *

Model		54	55	60	62A	62B	67	70
	P	77.8	30.9	51.5	42.6	36.9	31.6	40.3
54	T	63.3	67.8	56.4	57.1	71.8	44.4	76.5
	P	51.1	79.8	48.5	47.2	55.8	39.2	60.5
55	L	52.2	78.7	45.5	63.1	59.7	72.2	65.5
	P	55.6	63.8	43.4	73.3	59.2	36.7	42.0
62A	L	52.2	68.1	41.4	73.8	59.7	37.3	39.5
	T	52.2	80.0	58.5	73.1	77.3	52.9	57.9
	P	55.6	35.1	52.5	54.9	42.2	78.5	54.6
67	L	55.6	71.3	47.5	65.1	57.8	76.6	65.5
	T	58.9	76.7	68.1	62.6	71.3	77.8	75.6

* Recall, the tobit model is classifying <u>oil</u> production, the probit and logit models production (overall)

Table 4.3 DC (Production | I = 4, 5, or 6) for Probit and Logit Models

		54	55	60	62A	62B	67	70
55	P	54.8	78.9	57.1	38.4	28.6	27.8	33.9
	L	81.8	77.8	63.2	0	0	66.7	66.7
62A	P	58.6	50.0	55.6	61.4	41.3	25.7	36.0
	L	56.9	50.0	52.0	61.9	48.9	25.9	34.4
67	P	60.4	31.9	59.7	43.5	37.4	66.7	42.4
	L	91.7	57.1	73.3	60.0	40.0	90.0	54.1

Table 4.4 DC (Production|I = 5 or 6) for Probit and Logit Models

		54	55	60	62A	62B	67	70
55	P	54.8	92.3	57.1	39.0	45.5	28.3	39.4
	L	81.8	81.8	72.7	0	0	66.7	57.1
62A	P	75.0	52.9	60.0	95.5	48.4	26.7	36.8
	L	73.7	44.4	69.2	88.0	57.1	27.5	36.5
67	P	71.4	35.8	68.8	44.2	34.2	90.9	44.9
	L	91.7	60.0	76.9	72.7	50.0	100.0	51.6

Table 4.5 DC (Production|I = 6) for Probit and Logit Models

		54	55	60	62A	62B	67	70
55	P	54.8	100.0	59.6	39.3	33.3	28.3	42.1
	L	85.7	100.0	75.0	0	0	66.7	50.0
62A	P	84.6	45.5	64.7	100.0	60.0	28.9	38.1
	L	76.9	40.0	75.0	88.0	71.4	32.5	37.2
67	P	87.5	33.3	75.7	41.4	34.8	100.0	46.7
	L	88.9	70.0	77.8	71.4	33.3	100.0	57.7