Prediction of Future Oil Production of an Oil Reservoir Using Time Series Analysis

by

Douglas J. Eggert

T-1907

A Thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Master of Science (Mineral Economics).

Signed: Douglas J. Eggert
Douglas J. Eggert

Approved: Hunter Lovins [signature]
Thesis Advisor

J. Paul Matthews [signature]
Head of Department

Golden, Colorado

Date: 12-9, 1976

Abstract

As this country's oil resources dwindle, the petroleum engineer finds himself repeatedly confronted with the need to predict the remaining production capacity for producing oil reservoirs. The common procedure used to make such predictions, involving the extrapolation of a least squares' line of transformed exponential or hyperbolic form, is presented. Problems and disadvantages of this procedure are pointed out. The basics and advantages of time series analysis are discussed, and the application of this method to oil production decline analysis is suggested. Production records of four actual oil reservoirs are used to test the ability of time series analysis to predict future oil production. These production predictions are presented along with comparable predictions obtained by extrapolating least squares exponential and hyperbolic decline curves. These predictions are compared to the actual realized productions during the time of the prediction. Conclusions from the comparison are presented.

Table of Contents

List of Illustrations

Acknowledgements

I would like to personally express my appreciation to my thesis committee for their suggestions concerning the format and revision of this thesis.

In addition special thanks should go to Dr. Hunter Swanson for the time he has taken to help me answer some questions concerning the development of decline curve analysis. Jerry Bergeson of Jerry Bergeson and Associates should be thanked for giving me some insight into the actual uses today of production decline curves by petroleum production engineers. I would also add a special thank you to Dr. Raymond K. Mueller who made his time series package available to me and took the time to show me how to use it.

An Introduction to Production Decline Analysis

The valuation of a producing oil property relies heavily upon the property appraiser's estimates of future production and future life for the oil reservoir. The oil property owner is dependent on his property appraisers for these estimates and requires that they be the best estimates available because they directly affect the expected income of his lands and the expected period over which this income is to be received. In addition, the elimination of the percentage depletion allowance for most oil producers has increased the need for accurate estimates of oil reserves in order that cost depletion deductions can be computed to their fullest amounts. It is important, therefore, that any oil land valuation utilize methods which give reliable estimates of these values.

The petroleum engineer has developed several methods to enable him to make estimates of oil reserves. Among these are volumetric and material balance methods which may require estimates of such reservoir parameters as reservoir area and thickness, average effective porosity of the reservoir sand, and the percent of oil present that will most likely be recovered (Hughes, 1967, p 222-227). It is necessary to resort to such techniques when determining reservoir estimates before production has begun or in the early productive life of the reservoir. But as the reservoir produces, it yields important information - a production trend - which may be used to obtain better estimates of future production and remaining reservoir life (Oduolowu, 1976, p 24).

This production trend can be shown most easily by comparing monthly, yearly, or some other periodic production statistic. In general, once a field becomes fully developed, average production for the reservoir decreases regularly with time because reservoir pressure stays approximately proportional to the amount of remaining oil. This regularity of decline requires production methods to be held constant. Constant production methods mean that outside interference of the reservoir's production such as cleaning, water-flooding, repressurizing, or government regulation of production must be kept to a minimum. Also, it must be assumed that the reservoir produced at full capacity throughout its life or at most at some constant fraction of full capacity. Study of such production decline of oil reservoirs has been seriously conducted since the 1920's (Cutler, 1924). Production decline data has been found to be most easily studied when it is plotted versus time. Curves drawn through the data points show the declining production characteristic and thus are referred to as production decline curves.

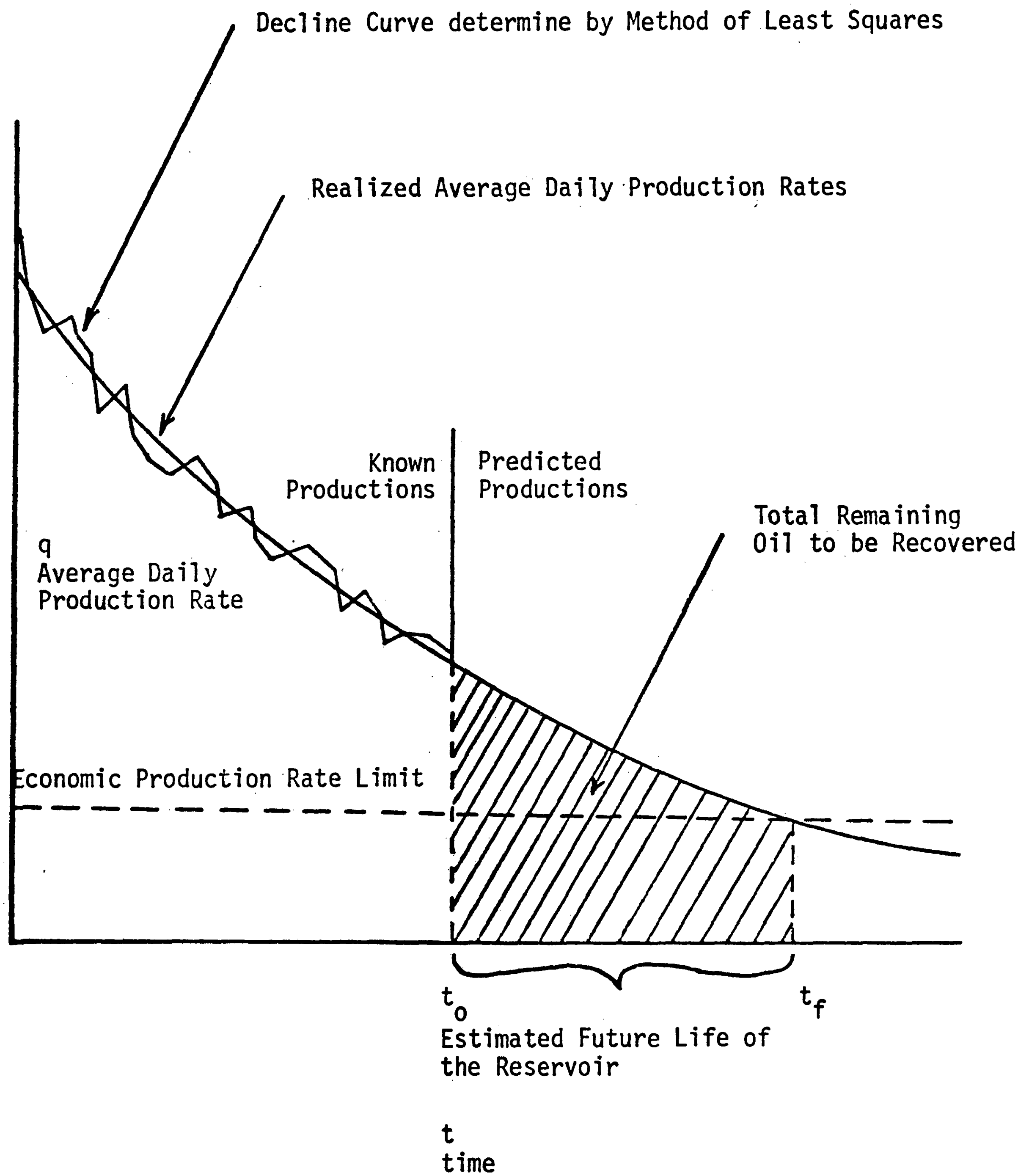
Decline curves can be drawn for individual producing wells as well as entire producing reservoirs. The production data of individual wells often show a decline with large irregularities. These irregularities can be smoothed if the average production of many wells or the production of the entire reservoir is used as the basis of the analysis. Only decline in total reservoir production was considered in this study.

Much of the study of decline curves has consisted of developing their associated mathematical relationships. Once an explicit equation for the decline curve has been determined, future productions are estimated by extrapolating the curve into future time periods. The total projected production of oil between two points in time is simply the area underneath the decline curve between the two time points as illustrated by Figure I. It is also shown in Figure I that the time at which the decline curve of future productions intersects the economic production rate limit may be used to help estimate the remaining time the lands may economically produce.

It is the goal of the author for this paper to be of use to people of a wide range of professional backgrounds. It therefore does not bog down into the pure mathematical theory behind time series analysis. Neither does it discuss the applicable details of reservoir engineering. Hopefully the discussion is sufficient to allow the reader unfamiliar with oil production decline to become acquainted with this property. Likewise the paper is structured so that the individual unfamiliar with time series should gain some knowledge of its basic concepts and mechanics.

To meet this goal, the paper is composed of three distinct parts. The first part, for those unfamiliar with oil production decline, explains production decline and shows how extrapolation of past decline has been used to predict future production. The second part generally explains what time series is and how it was applied to oil production decline. Comparison of predictions of the two methods is shown in this part, and conclusions are made. The third part consists of the appendices which follow the body of this thesis. These appendices, for the interested reader, outline the basic mathematics required to conduct a time series analysis. In addition, two sample problems are given to further explain the method. Computer output of the time series package used for this thesis is also found in the appendices.

Figure 1. Typical Production Decline Curve and Associated Data



It is hoped that readers knowledgable in other areas of science and engineering will, by seeing the way time series analysis was applied for oil production decline, see possibilities of applying these methods in their areas of knowledge. Modern time series analysis has only been completely developed in this present decade, and it appears that real physical applications of it have just begun to be realized.

Extrapolative Methods Used to Predict Future Oil Production

Early studies of production decline curves showed that such curves often become linear when the production data was displayed on a semi-log or log-log plot rather than a plot of Cartesian coordinates. Production decline curves as a result have been found to resemble either exponential (semi-log) decline relationships or hyperbolic (log-log) decline relationships. Cases have also been recorded of constant straight-line production decline (Hughes, 1967, p 209), but such occurrences are rare and were not considered in this study. The basic assumption of the extrapolative procedure is centered upon the belief that past trends should continue their influence to the same degree in the future. A reservoir whose production has been determined as declining exponentially is then assumed to continue its exponential production decline through the entire life of the reservoir. Similarly, a reservoir exhibiting hyperbolic decline is assumed to continue its hyperbolic decline for all future periods of production.

To estimate future productions of oil, a decline curve on a Cartesian plot or a transformed linear decline may be drawn and extended into the future. To save time, the petroleum engineer often will subjectively "eyeball" a curve or line through past production rates and extend this curve or line to obtain estimates of future productions. Unfortunately, many different curves and lines may be drawn through the same set of production rates, resulting in often substantially different projected productions. The placement of such curves may be affected not only by the production engineer's experience with previous production declines, but also perhaps his personal feelings toward unrelated elements at the time such a judgement is made.

The Method of Least Squares

The method of least squares can be used to provide the equation of the line which minimizes the sum of the squared deviations of the actual production rates with the associated production rates given by the determined line. The manner in which the least squares estimates of the parameters of this line are determined is probably familiar to most readers. Therefore, this procedure will only be briefly summarized. Johnston (1972, p 8-43) presents a very complete analysis of the least squares technique, and the interested reader is referred to this source.

Let X_i and Y_i denote the actual observed values to which it is desired to fit a straight line. This straight line is of the form

$$\hat{Y}_i = \hat{a} + \hat{b}X_i$$

where $\hat{a}$ is the estimate of the y-intercept and $\hat{b}$ is the estimate of the slope of the line. The values of Y_i predicted by this line are denoted by $\hat{Y}_i$. It is desired to minimize the sum of the squared differences between the realized Y values (Y_i) and the Y values predicted by the least squares line ($\hat{Y}_i$). That is minimize

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

or

$$\sum_{i=1}^{n} (Y_i - \hat{a} - \hat{b}X_i)^2$$

Rules of calculus require that the partial derivatives of this last expression with respect to $\hat{a}$ and $\hat{b}$ both be equal to zero as a necessary condition for minimization. Simplification of the two resulting equations obtained from this procedure yield what are known as the normal equations for the regression line,

$$\sum_{i=1}^{n} Y_i = n\hat{a} + \hat{b}\sum_{i=1} X_i$$

$$\sum_{i=1}^{n} X_i Y_i = \hat{a}\sum_{i=1}^{n} X_i + \hat{b}\sum_{i=1}^{n} X_i^2$$

The two unknowns in these equations, $\hat{a}$ and $\hat{b}$, are readily solved for once the necessary summations of the data points (X_i, Y_i) are made and substituted into the equations. It should be pointed out that the second derivative tests verify that the solution $(\hat{a},\hat{b})$ of the normal equations produces a minimum rather than a maximum value for the sum of squared deviations.

This method was first applied by Larkey (1925, p 1322) to help petroleum engineers make a good decision concerning a reasonable curve through exponentially declining past production rates. The hyperbolic decline curve is not as easily adaptable to the least squares technique, but with the aid of modern computers this problem can be solved efficiently today (Gopal, 1973, p 51).

The Exponential Production Decline Curve

The exponential production decline curve results when the time rate of change in the production rate is equal to a constant fraction of the production rate. The differential equation which displays this relationship is

$$\frac{dq}{dt} = -Dq$$

where

q is the production rate (production units/period of time)

t is time (period of time)

D is the constant fraction of decline per period of time (period of time)$^{-1}$

This differential equation which occurs frequently in the world of economics and natural science is easily solved. Denoting the initial production rate as q_o the solution to the differential equation is

$$q = q_o e^{-Dt} \qquad (1)$$

The cumulative production, denoted by Np, can be computed by integrating equation (1) over the desired time period, t_o to t_f

$$Np = \int_{t_o}^{t_f} q\,dt$$

$$= \int_{t_o}^{t_f} q_o e^{-Dt} dt$$

$$= -\frac{q_o}{D} \left[e^{-Dt_f} - e^{-Dt_o} \right] \qquad (2)$$

Taking natural logarithms of both sides of equation (1) results in

$$\ln q = \ln q_o - Dt$$

Thus, it can be seen that the natural log of the production rate, q, is linearly related to time, t. In applying the least squares technique then, the periods of time can be considered as independent variables and the natural logarithm of the production rates the dependent variables.

The value of the slope of the line obtained by the least squares fit is the least squares estimate of -D, and the value of the y-intercept is the least squares estimate of $\ln q_o$.

The Hyperbolic Production Decline Curve

Hyperbolic decline differs from exponential decline in that the decline of production (-D) is no longer considered constant as it was for exponential decline curves. Instead, the decline of production is assumed to be proportional to a power, n, of the production rate, q. The same differential equation solved for exponential decline is solved for the hyperbolic decline case, only with Cq^n substituted for D, where C is a constant. This gives

$$\frac{dq}{dt} = - Cq^{1+n}$$

as the hyperbolic decline differential equation. This equation is separable and easily solved yielding the result

$$q = (q_o^{-n} + nCt)^{-\frac{1}{n}} \qquad (3)$$

The initial decline in production, $-D_i$, must be equal to $-Cq_o^n$. This initial condition substituted into equation (3) gives a simplified result of

$$q = (q_o^{-n} + nD_i t q_o^{-n})^{-1/n}$$

$$= q_o(1 + nD_i t)^{-1/n} \qquad (4)$$

A more convenient form for equation (4) is obtained by defining the variables

$$h \equiv \frac{1}{nD_i}$$

and

$b \equiv \frac{1}{n}$

so that

$$q = q_o(1 + \frac{t}{h})^{-b} \tag{5}$$

The cumulative production, Np, is found by integrating equation (5) over the time of production. The result of the integration from time t_o to t_f is

$$Np = \int_{t_o}^{t_f} q_o(1 + \frac{t}{h})^{-b}dt$$

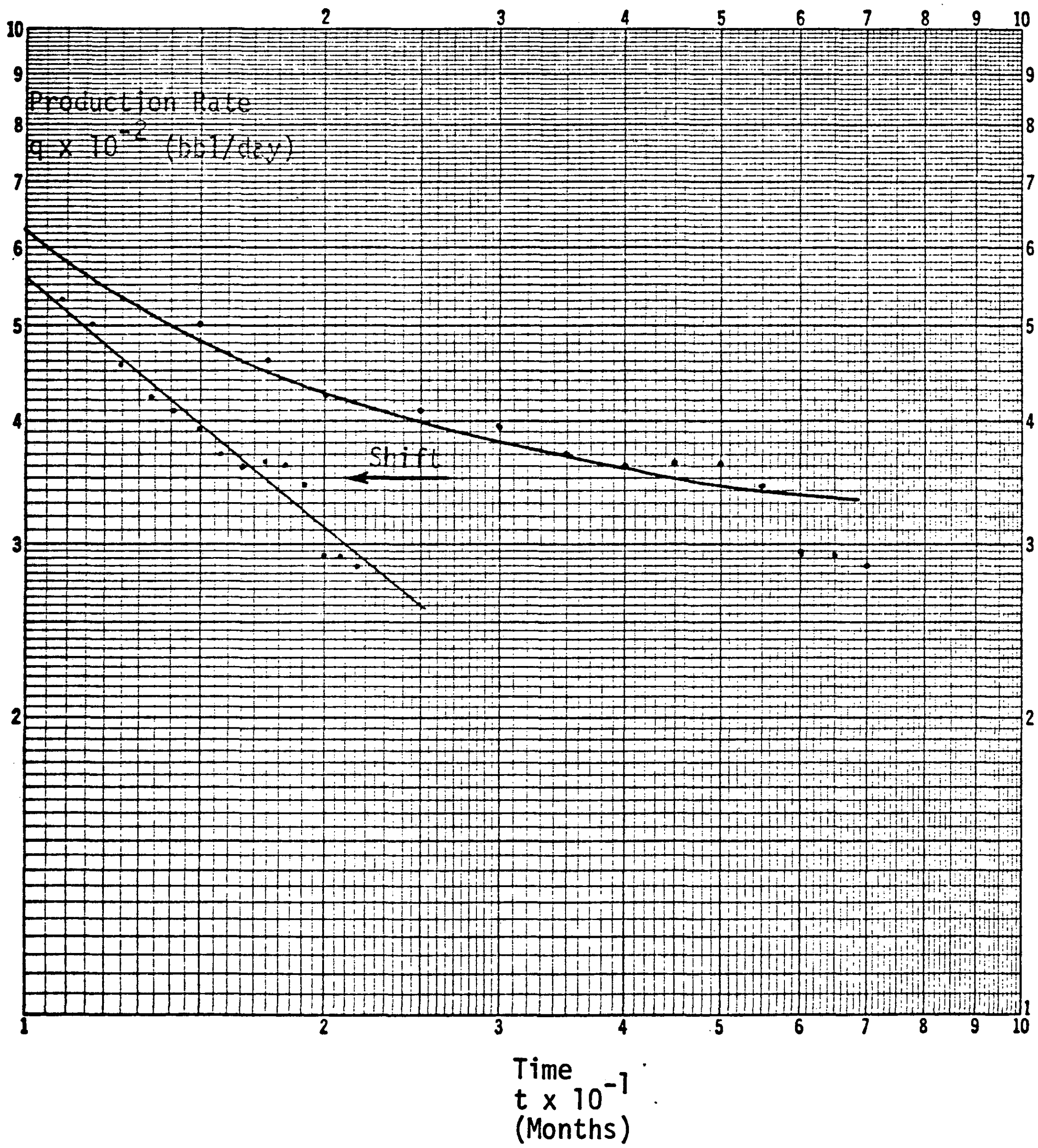
$$= \frac{q_o h}{1-b}\left[\left(\frac{t_f + h}{h}\right)^{1-b} - \left(\frac{t_o + h}{h}\right)^{1-b}\right] \tag{6}$$

Equation (5) may be linearized by taking natural logs of both sides of the equation giving

$$\ln q = \ln q_o - b \ln(1 + \frac{t}{h}) \quad .$$

This result shows that lnq is linearly related to $\ln(1 + \frac{t}{h})$. This explains why a plot of q versus t on log-log coordinates results in a straight line relation after a proper adjustment is made for the distortion of t. This adjustment is simply a horizontal shifting of the curve by some amount. Figure II shows how this shifting of the transformed production data for an actual oil reservoir, Foothills Pool, North Dakota, produces a linearized relationship. The method of least squares for the hyperbolic decline curve used in this study required a similar horizontal shift of the independent variable, $\ln(1 + \frac{t}{h})$. That is, values for h were arbitrarily chosen. The least squares problem was then solved for each value of h yielding the

Figure II. Linearization of Oil Production Decline by Horizonal Shifting Using Periodic Productions of Foothills Pool, North Dakota.



corresponding parameter values for b (the slope of the line) and ln q_0 (the y-intercept of the line). The sum of squared deviations was noted for each value of h. The value of h yielding the minimum sum of squared deviations and the corresponding parameter values for b and q_0 were used in the extrapolation equation to compute estimates of future productions.

Harmonic decline is often mentioned as a third type of oil production decline. In reality, harmonic decline is a special case of hyperbolic decline. The equation for the harmonic decline curve is simply the hyperbolic decline equation (equation (4)) with n = 1. Specifying the value of n reduces the flexibility of equation (4). Because of this, a harmonic decline curve fit of the data can never exceed the "goodness" of a hyperbolic decline curve fit, and harmonic decline was not seriously considered in this paper for that reason.

It should be pointed out that exponential decline is also a special case of hyperbolic decline. As the value of n in equation (4) approaches zero in the limit, the hyperbolic relation becomes equivalent to an exponential relationship. This means that when oil productions are actually exponentially related, the best hyperbolic fit falls exactly upon the exponential fit.

After the production data has been fitted to exponential, hyperbolic, and perhaps harmonic decline curves, it becomes necessary to make a judgement concerning the type of decline which most adequately describes the data. Stevens and Thodos (1961, p B50) suggest selecting that decline curve which exhibits the smallest value of squared deviations. Another criterion which is often used to make judgements of "goodness of fit" is the coefficient of determination (Chisholm and Whitaker, 1971, p 106). The coefficient of determination is simply the proportion of the variance of the realizations of the dependent variables about their mean which can be explained by the suggested linear relationship. Both criteria were used in this study to determine the most nearly appropriate fitted curve.

Problems with Forecasting from Production Decline Curves

1) Fitting oil production data to some preconceived curve is merely a convenience which enables the production engineer to easily extrapolate future reservoir performance. Although it is possible to correlate certain types of decline (exponential, hyperbolic, and harmonic) to a certain set of physical conditions (Brons, 1963, p 23), there is no physical basis for production decline curve analysis (Nind, 1964, p 50). It is, therefore, generally not possible to determine the type of production decline which will occur until production has actually commenced.

2) The assumption is made with production decline curve analysis that the past trend will continue unchanged in the future. This means that with no significant changes in the operation of the oil field, the parameters of the decline curve will not significantly change throughout the productive life. Also important, of course, is the need for the type of decline to not change during production. Unfortunately this frequently is not the case, and initial production decline exhibits hyperbolic characteristics while the latter portions of production show exponential tendencies (Hughes, 1967, p 210).

3) Another problem with the least squares regression method of fitting the data to a decline curve occurs with the forecasted production figures. The petroleum production engineer is interested in not only the forecasted value of production, but also in the size of the interval of most likely productions about that forecasted production. The forecasted production will rarely equal the realized production, so some idea of how much the actual production may vary about the forecasted production is useful. No presently-known analytical method of providing the smallest such probability limits for extrapolated productions is possible using decline curve analysis.

4) Probably the major weakness with the decline curve approach is that it assumes that extrapolation of the least squares line is an entirely valid procedure. This assumption is wrongly made time and time again by those seeking forecasts from their known data. The strength of regression, however, is prediction of observations requiring interpolation rather than extrapolation. Thus, the least squares line can be very helpful when predicting an observation among the known data points. There is, however, no basis whatsoever to conclude that this same degree of predictive capability holds when a forecast is made very far outside the range of sample data with no accompanying increase in information.

Philosophy of Time Series Analysis

Time series is a relatively new forecasting procedure which is gaining acceptance within the economic and scientific community. It basically treats a time dependent process as a stochastic process. This means the process is assumed to consist of an ordered collection of random variables, and the observed values are simply realizations of particular random variables.

The random variables making up the time series process need not be independent. This is important because it means time series considers the statistical dependence among observations in constructing a forecast function. This is a distinguishing feature of time series analysis as extrapolation and other forecasting techniques in making their forecasts do not assume there to be statistical dependence among the observations. Very often realizations of physical processes really seem to be randomly distributed (completely independent and identically distributed) to the layman. However, often after close examination and application of proper analysis techniques, some general repetitious pattern can be found in the observations of the process. The discovery of a pattern is very important because once the type of dependence has been determined, much more can be said about future observations as long as it is assumed these future observations have the same type and degree of dependence as their predecessors.

The Concept of Stationarity

To properly apply any time series analysis, it is necessary as a preliminary step to transform the raw observations so that all the

random variables of the process have the exact same mean and variance and form what is known as a stationary process. A formal definition for stationarity of processes can more easily be given if some notation is introduced. Let W_t be the observation at time t. A process is said to be strictly stationary when every ordered set of m + 1 observations $(W_t, W_{t+1}, \dots W_{t+m})$ possesses the same probability law as any other set of m + 1 observations of the same process shifted k units $(W_{t+k}, W_{t+1+k}, \dots, W_{t+m+k})$, for each integer $m \geq 0$.
This means that all random variables of a stationary process must have the same mean and variance if they exist. It is further implied that a stationary process must contain no longterm upward or downward trend, but instead must have an affinity for some constant mean. Most physical and economic processes are nonstationary in nature. The number of automobile registrations in the United States each year, for example, has steadily increased since WW II so one would suspect that such a process is nonstationary (Nelson, 1973, p 177). Similarly the Gross National Product of the United States has followed a growth pattern for many years and would also be considered an example of a nonstationary process (Nelson, 1973, p 64). It is not as common to find the process being studied to be stationary in its untransformed state. Nelson (1973, p 23) uses the unemployment rate of the United States as an example of a stationary process since the unemployment rate has fluctuated about a mean of 4.7% since World War II.

The process of transforming a nonstationary series to a stationary series often simply requires the removal of the upward or downward trend of the series. Most upward or downward trends can be removed by simple transformations of the raw data. Linear trends can almost always be removed by taking differences of consecutive observations. If the trend persists after this transformation, it may be necessary to further transform the data by taking differences of the previously computed differences (second differences). Processes which do not exhibit constant levels of absolute change, but rather constant levels of percentage change require a transformation involving differences of the natural logarithms of the raw data to form the stationary process. For example, Nelson (1973, p 177) has shown that it is possible to form a stationary process from the nonstationary automobile registration series by taking differences of the natural logarithms of the raw data. Similarly he has shown that the Gross National Product of the United States becomes a stationary series if first differences of the raw data are computed and used in the analysis (Nelson, 1973, p 64).

When a trend has been removed from a process, the autocorrelation function (see Appendix I) will have a value very close to zero after only a few terms. An autocorrelation function which decreases in value very slowly indicates the trend has not been entirely removed and further transformation is required to obtain a truly stationary process.

Advantages of Time Series Analysis

One major advantage of using time series analysis as opposed to other forecasting methods is that time series takes the interdependence of observations into account. An observation may show dependence on one or more observations of the past. (Recall such dependence is ignored and assumed away by regression-type forecasting methods). Furthermore, the correlations between observations may be strong or mild, positive or negative. Recognizing such dependence, having this type of information available and quantifiable, and using it to forecast is obviously preferable to incorrectly assuming the observed values establish a smooth trend which can be extrapolated in order to forecast future values of the process.

A second advantage of the use of time series arises when it is compared to the regression method used to derive the decline curve equations presented earlier. In developing the decline curve for a particular reservoir, the data is forced to fit a model of either exponential or hyperbolic form. Recall that least squares regression determines the necessary parameters to "best fit" a given model. Brons (1963, p 22) points out that such forecasting logic is inverted and that what is needed instead is a method in which the formula or curve is instead determined by the data. The forecasting function arrived at through time series analysis accomplishes this exactly since it is uniquely determined by the correlations among the data.

A final advantage of time series analysis results since the observations are treated as realizations of a stationary process. It is possible to calculate the variances of the distribution of the random variables, and then to calculate probability limits for individual forecasts. These probability limits give the forecaster an idea of how much the future realized observation may vary from his forecast of that observation. No technique is available to determine probability intervals for forecasts using extrapolation, although so called confidence intervals are often incorrectly interpreted as such.

The Application of Time Series Analysis to Oil Production Decline

Past production rates for four actual oil fields were used as case examples to test the ability of time series analysis to predict future production rates for those fields. The results were compared with those obtained by fitting the same production data to regression decline curves. The production statistics used in the analysis are those compiled by the North Dakota Geological Survey in their bi-yearly publication, Production Statistics and Engineering Data - Oil in North Dakota.

The four North Dakotan oil reservoirs used in the study were Foothills Pool, Beaver Lodge Madison Pool, North Black Slough Pool, and Haas Madison Pool. The author had no personal knowledge of the operations of these reservoirs and was therefore not able to include any substantial subjective judgement in the analysis. These reservoirs were selected because of 1) their histories of reasonably regular decline; and 2) the number of producing wells in each field during the decline was fairly constant, making these fields nearly perfect examples of production decline. It is necessary that the number of producing wells be relatively constant so that any observed decline in production is attributed directly to actual production decline rather than to the fact that fewer wells are producing.

It often happens in practice, however, that the number of producing wells in a given field varies greatly throughout the oil field life. It is still possible to perform a production decline analysis for such a field, if the data is first put on a constant basis. This is done by dividing each monthly, semi-annual, or annual production figure by the average number of wells producing during that time. Because the number of producing wells per period was nearly constant for each of the four sample cases, such transformations were not called for and were not carried out.

One refinement of the raw production data from the four North Dakotan oil fields was required. Monthly production statistics were recorded by the North Dakota Geological Survey, but because the lengths of months vary from 28 to 31 days, it was necessary to divide each monthly production figure for all four wells by the number of days in their respective months. The result is a common basis for the data of average barrels per day over each month. Such a transformation is necessary because time series analysis will detect in the untransformed data an undesirable yearly trend due to unequal periods of production. The resulting data, the average daily production for each of the four fields by month, are listed in Table I.

Average production rates which obviously were not in line with the general decline trend were not considered in either the regression or time series analyses. Such irregularities result when a significant number of the wells of a field are closed down for routine maintenance, repair, or modification, and thus produce no oil for a short period. Such an interruption in the well's production is referred to by the petroleum engineer as a shut-in. If a significant number of the wells of a field are shut in, then the total production figures for the field during the shut-ins will be much lower than if the shut-ins had not

Table I. Average Daily Production of Oil by Months for North Dakotan Oil Fields (bbls/day).

Time Period (Month)	Foothills Pool	Beaver Lodge Madison Pool	North Black Slough Pool	Haas Madison Pool
1	566.9	5206.	605.3	627.3
2	535.2	5092.	604.3	580.4
3	552.6	4834.	561.7	584.7
4	541.1	4567.	558.9	598.4
5	534.5	4709.	568.2	582.6
6	521.1	4579.	613.6	557.5
7	516.9	4547.	549.0	568.4
8	516.9	4520.	537.1	563.7
9	515.1	4421.	582.1	560.0
10	502.7	4435.	490.6	556.3
11	473.3	4201.	542.7	503.6
12	467.9	4142.	531.6	537.3
13	458.0	4254.	530.7	536.3
14	479.2	4297.	427.0	533.3
15	462.0	4295.	439.1	523.5
16	460.8	2888.*	445.5	518.5
17	467.9	4388.	384.0	496.0
18	443.3	4189.	431.3	503.9
19	447.7	4218.	384.0	504.6
20	426.6	4278.	361.0	494.3
21	416.8	4102.	371.2	488.4
22	427.3	4010.	313.9	479.5

Time Period (Month)	Foothills Pool	Beaver Lodge Madison Pool	North Black Slough Pool	Haas Madison Pool
23	421.9	3869.	379.4	480.9
24	403.1	3792.	366.3	496.3
25	412.6	3651.	351.3	521.4
26	408.8	3592.	338.3	508.3
27	393.8	3613.	357.3	491.5
28	390.9	3494.	309.3	495.8
29	321.4*	3337.	299.4	488.5
30	426.8	3356.	281.9	478.3
31	395.9	3174.	304.5	482.7
32	385.0	3197.	282.4	464.6
33	407.1	3220.	247.5	465.9
34	333.9*	3033.	252.3	460.1
35	387.3	2979.	253.5	481.0
36	375.8	3030.	247.6	449.3
37	371.2	2975.	257.8	464.4
38	364.0	2954.	273.5	461.6
39	366.9	2839.	269.9	460.7
40	347.1	2967.	261.2	480.3
41	367.5	2760.	245.6	472.6
42	360.7	2718.	251.5	449.6
43	366.3	2580.	244.2	456.6
44	350.8	2639.	236.3	443.0
45	366.6	2706.	223.7	444.4
46	362.2	2598.	234.6	431.0

Time Period (Month)	Foothills Pool	Beaver Lodge Madison Pool	North Black Slough Pool	Haas Madison Pool
47	367.9	2625.	209.4	433.9
48	364.6	2619.	208.2	424.2
49	354.4	2519.	191.6	434.0
50	342.6	2379.	204.3	429.2
51	367.3	2370.	207.8	433.3
52	363.2	2275.		428.0
53	340.2	2328.		429.2
54	330.1	2189.		418.0
55	339.8	2272.		413.5
56	337.9	2271.		
57	346.3	2140.		
58	298.5	2169.		
59	319.0	2128.		
60	299.6	2022.		
61	294.8	2058.		
62	294.4	2043.		
63	284.5	2137.		
64	283.0	2041.		
65	310.0	1943.		
66	295.0			
67	293.1			
68	287.2			
69	289.3			
70	290.6			

Time Period (Month)	Foothills Pool	Beaver Lodge Madison Pool	North Black Slough Pool	Haas Madison Pool
71	289.0			
72	285.1			
73	288.5			
74	288.4			
75	291.0			

For Foothills Pool Month 1 = January 1967

For Beaver Lodge Madison Pool Month 1 = February 1968

For North Black Slough Pool Month 1 = September 1969

For Haas Madison Pool Month 1 = June 1969

*Production at these periods was apparently curtailed by the field operators. These production rates were thus eliminated as data and not considered in the forecasting analyses.

occurred. The field productions during significant shut-ins were thus dropped from the analyses. The productions used in both analyses therefore contained the production rates up to the shut-in and then continue with the production rates after the shut-in. The shut-in is thus treated as if it had never occurred. The assumption is made here that once the shut-in wells begin to produce again, the repairs or modifications made during the shut-in will not significantly change the structure of the production record. Especially important is the assumption that the shut-in will not cause the dependence among the observed production rates to significantly change. The assumption that the productions during the period of shut-in may be excluded from the analyses and that the shut-in changes and modifications will not change the production structure is not entirely valid; however, Brons (1963, p 23) feels the effect of a shut-in on future production is insignificant enough to exclude it from his analysis. For the four case examples sudden reductions in production, probably due to shut-ins of a significant number of wells, occurred at three points in the productions listed in Table I. Two of these shut-ins occurred in the Foothills Pool production record and the third in the productions of Beaver Lodge Madison Pool. The eliminated productions are indicated in Table I.

Results Using Extrapolative Forecasting Methods

The production data of Table I was fitted to exponential (equation (1)) and hyperbolic (equation (5)) decline curves so that the sum of squared deviations was at a minimum for each of these curves. The statistical computer package, Stat Pack, developed by Western Michigan University was used to accomplish this task. The resulting regression equations are listed in Table II.

The statistician in least squares analysis is most interested in the "goodness" of the regression fit and whether there is any evidence to suggest that an improper model has initially been specified. To conduct such an analysis the statistician will look at several quantities. The coefficient of determination, R^2, is the proportion of the variance of the production rates about the mean of the production rates explained by the linear relationship of the production rate, q, on time, t. It may be interpreted as a measure of how well the regression line explains the relationship between the production rate and time. Values for the coefficient of determination may range from 1.0 to 0.0 inclusive. A value close to 1.0 indicates a very good fit while a value in the neighborhood of 0.0 indicates a poor fit. Estimates for the standard error (standard deviation of the sampling distribution) of the regression coefficient, denoted by $SE(\hat{b})$, also give a measure of the strength of the regression fit. A small value for $SE(\hat{b})$ in relation to the value of $\hat{b}$ indicates a very strong relationship has been found to exist between the production rate and time. It is also possible to test the goodness of the fit by means of an F-test. This is done by selecting some significance level, α, which is essentially the proportion of time the tester is willing to concede that he will incorrectly reject the null hypothesis. The significance level used for this analysis was 5%. Then the null hypothesis, $H_0:\hat{b} = 0$, is tested against the alternate

Table II. Exponential and Hyperbolic Decline Curve Regression Equations for Foothills Pool, Beaver Lodge Madison Pool, North Black Slough Pool, and Haas Madison Pool. (See further details in discussion on page 42)

Pool	Exponential Equation	Hyperbolic Equation
Foothills	$q = 532.6e^{-0.0093t}$	$q = 561.1 \left(\frac{t}{60} + 1\right)^{-0.8657}$
Beaver Lodge Madison	$q = 5187.12e^{-0.0155t}$	$q = 5187.12 \left(\frac{t}{10{,}000} + 1\right)^{-155.05}$
North Black Slough	$q = 624.1e^{-0.0235t}$	$q = 659.0 \left(\frac{t}{80} + 1\right)^{-2.4486}$
Haas Madison	$q = 580.3e^{-0.0062t}$	$q = 608.6 \left(\frac{t}{20} + 1\right)^{-0.2711}$

hypothesis, $H_1:\hat{b} \neq 0$. The null hypothesis is rejected if the F value calculated from the regression exceeds the critical value of F found from the F statistical tables having (1, n-2) degrees of freedom and significance level α. These statistics are listed in Table III and all seem to indicate very good fits for the regression lines.

It next becomes necessary to determine for each oil reservoir which regression curve best describes the production decline for that reservoir. The sum of the squared deviations (related to the standard error of the estimate in regression) and the coefficient of determination may each be used to help make such a judgement. These statistics are tabulated in Table IV. A low value for the sum of squared deviations or a value close to 1.0 for the coefficient of determination indicate a "good" fit for the data. The application of these criteria to the regression curves for these data seem to indicate that the declines of Foothills Pool, North Black Slough Pool, and Haas Madison Pool are hyperbolic in nature, while Beaver Lodge Madison Pool more nearly exhibits exponentially declining production. Such behavior is not surprising since studies have shown that most decline curves are of the hyperbolic type (Cutler, 1924, p 22).

Results Using Time Series Forecasting Methods

A time series analysis was also conducted for the data of Table I. It was initially determined that a transformation involving differences of natural logarithms of the raw data was necessary to form the stationary series. This is to be expected since the absolute changes in production rates tend to decrease as the actual production rate

Table III. Statistics for Determination of "Goodness" of Regression Fit for North Dakotan Oil Fields (assume $\alpha = 5\%$)

Pool	Exponential or Hyperbolic Fit	Coefficient of Determination	SE($\hat{b}$)	F_{cal}	F_{crit}	Is $H_0: \hat{b} = 0$ rejected
Foothills	Exponential	0.9598	0.00022	1694	3.98	Yes
Foothills	Hyperbolic	0.9683	0.01860	2166	3.98	Yes
Beaver Lodge Madison	Exponential	0.9856	0.00024	4257.	3.99	Yes
Beaver Lodge Madison	Hyperbolic	0.9856	2.37900	4248.	3.99	Yes
North Black Slough	Exponential	0.9588	0.00070	1139.	4.04	Yes
North Black Slough	Hyperbolic	0.9633	0.06825	1287.	4.04	Yes
Haas Madison	Exponential	0.9179	0.00025	592.	4.03	Yes
Haas Madison	Hyperbolic	0.9374	0.00962	794.	4.03	Yes

Table IV. Sum of Squared Deviations and Coefficient of Determination for Curve Fits of Table II.

Pool	Exponential		Hyperbolic	
	Sum of Sq. Dev.	Coef. of Deter.	Sum of Sq. Dev.	Coef. of Deter.
Foothills	0.1163	0.96000	0.0917	0.96800
Beaver Lodge Madison	0.0760	0.98564	0.0762	0.98561
North Black Slough	0.2620	0.95900	0.2330	0.96300
Haas Madison	0.0473	0.91788	0.0361	0.93742

decreases, indicating that it is perhaps percent changes in the data which are stationary. For the reader unfamiliar with modern time series analysis, Appendix I outlines the preliminary steps necessary to build a time series model. Appendix II uses the procedure of Appendix I in a simple sample problem. Appendix III shows how time series was applied to an actual case of oil production decline. The summarized results of the time series analysis for the four cases are given in Table V. The corresponding computer output is found in Appendix IV.

The time series models listed in Table V were analyzed extensively to check that they accurately modeled the known production rates. The first test in this analysis is known as overfitting. This test requires the modeling of the production rates with a more elaborate time series model than the original one. In this case all models were of the form MA (1). A more complete definition of the pure moving average process (MA) is found in Appendix I. It is only necessary at this point to know that a pure moving average process relates future observations to past disturbances only. Thus a MA (1) process relates a one-period-ahead production to the last two realized disturbances. A more elaborate model would relate a one-period-ahead production to the last two realized disturbances plus the last realized observation. This is a model of form ARMA (1,1). (See Appendix I for further description of the mixed autoregressive moving average model (ARMA).) This more elaborate model has an additional parameter. The significance of the value of this added parameter is the key to determining whether any benefit is gained by adopting the ARMA (1,1) model. For each of the

Table V. Time Series Models for Oil Production Decline of Foothills Pool, Beaver Lodge Madison Pool, North Black Slough Pool, and Haas Madison Pool.

Pool	Type of Time Series Model	Model
Foothills	IMA (1,1)	$W_t = a_t - 0.573_{a+-1} - 0.00926$
Beaver Lodge Madison	IMA (1,1)	$W_t = a_t - 0.407_{a+-1} - 0.01564$
North Black Slough	IMA (1,1)	$W_t = a_t - 0.631_{a+-1} - 0.02138$
Haas Madison	IMA (1,1)	$W_t = a_t - 0.584_{a+-1} - 0.00772$

Note: The "I" preceding the "MA" for the model type is standard time series notation included to indicate that it was necessary to take differences to form the stationary process. The first "1" listed after "IMA" indicates that first differences produced the stationary process. Since it is assumed the reader realizes that all production rate series used in this thesis required a transformation to stationarity, the "I" and first digit in the paranthesis will be omitted throughout the thesis.

four models of Table V, the added parameters had relatively insignificant values, which indicates the adequacy of the original model.

A series of diagnostic checks was also conducted for each of the four time series models. The cumulative periodogram, Portmanteau test, and an analysis of the autocorrelation of the residuals were used to check the models of Table V (Box and Jenkins, 1970, p 287-299). These diagnostic checks also did not suggest any inadequacy of the original models.

Table V indicates that the time series models associated with the four declines are all of similar form, MA (1). A further study would be necessary before it would be possible to generally conclude that all transformed oil production declines can be modeled as a first order moving average process. However, the four cases presented do seem to show that the first order moving average model is indicative of oil production decline.

Comparison of Extrapolative and Time Series Forecasts

The most recent production data for each of the oil reservoirs was not included in the construction of the production decline models. These production rates instead were used to make a judgement of the predictive accuracy of each of the models. Figures III-VI graphically compare the decline curve predictions, time series predictions (point predictions), and actual production rates for the four sample cases. Also included in these figures are the 95% probability limits for the time series predictions.

To make a judgement of the relative accuracy of the different

Figure III. Average Daily Production Rates for Foothills Pool, North Dakota - Decline Curve Predictions, Time Series Predictions (with Associated 95% Probability Limits), and Realized Production Rates

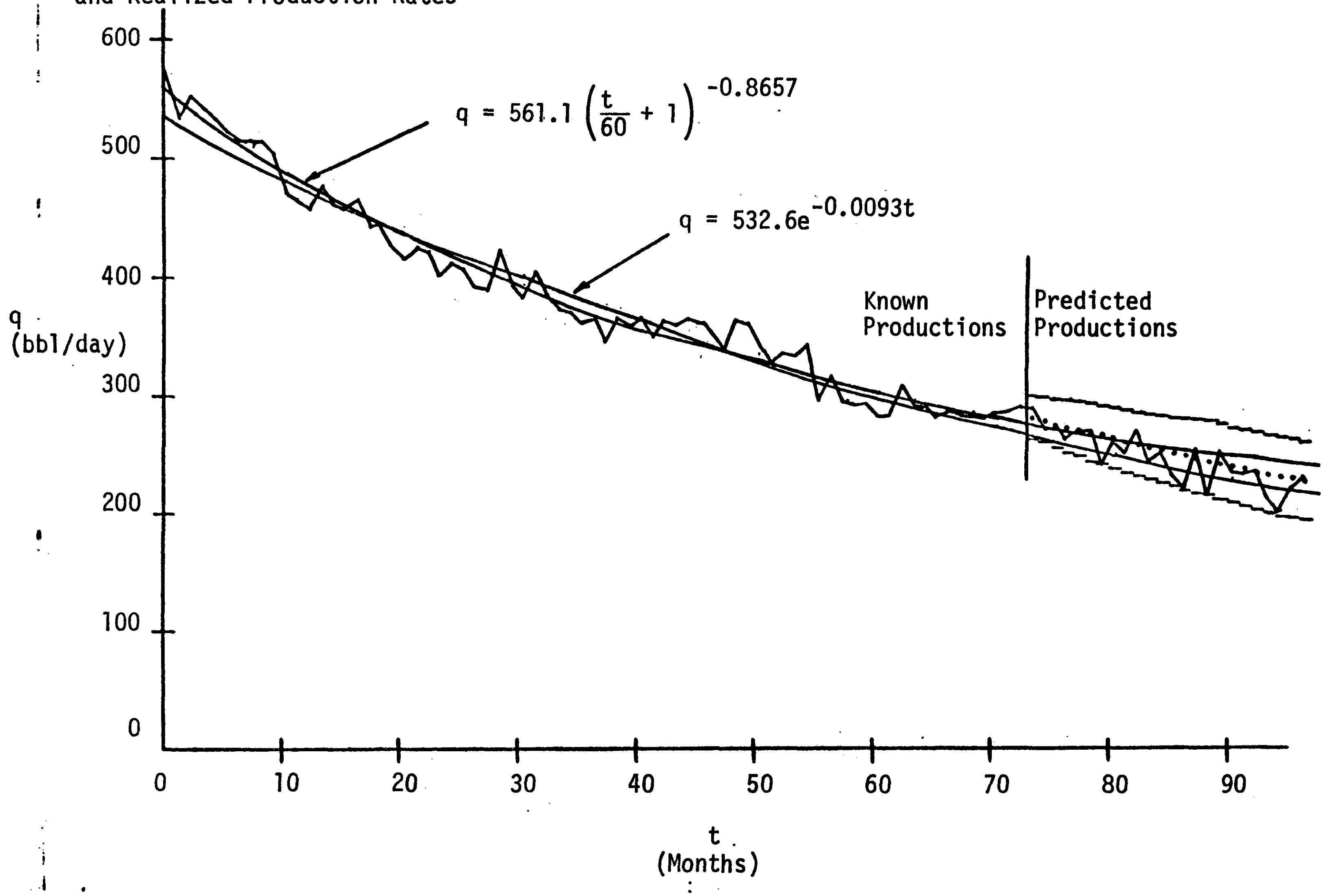


Figure IV. Average Daily Production Rates for Beaver Lodge Madison Pool, North Dakota - Decline Curve Predictions, Time Series Predictions (with Associated 95% Probability Limits), and Realized Production Rates

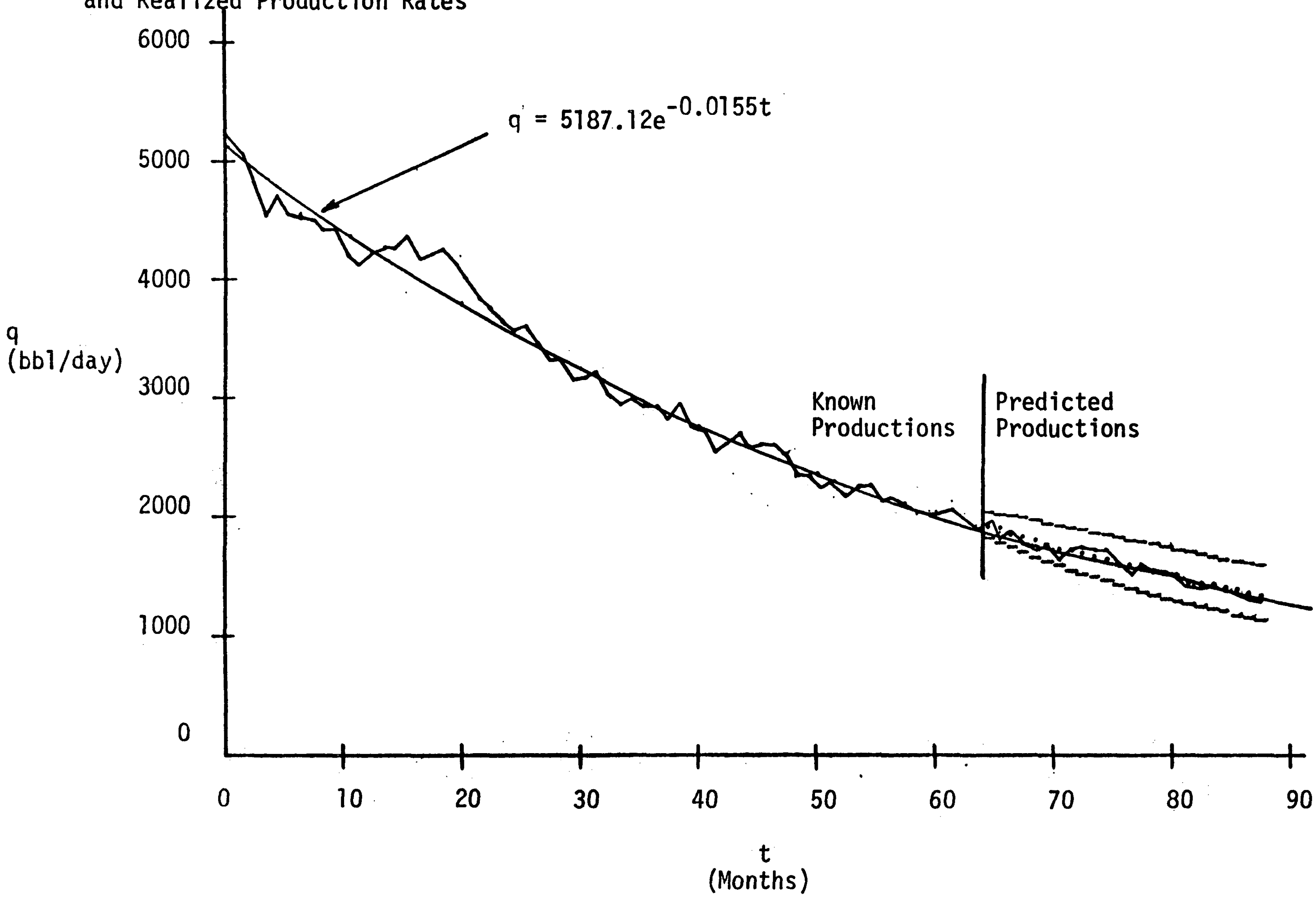


Figure V. Average Daily Production Rates for North Black Slough Pool, North Dakota - Decline Curve Predictions, Time Series Predictions (with Associated 95% Probability Limits), and Realized Production Rates

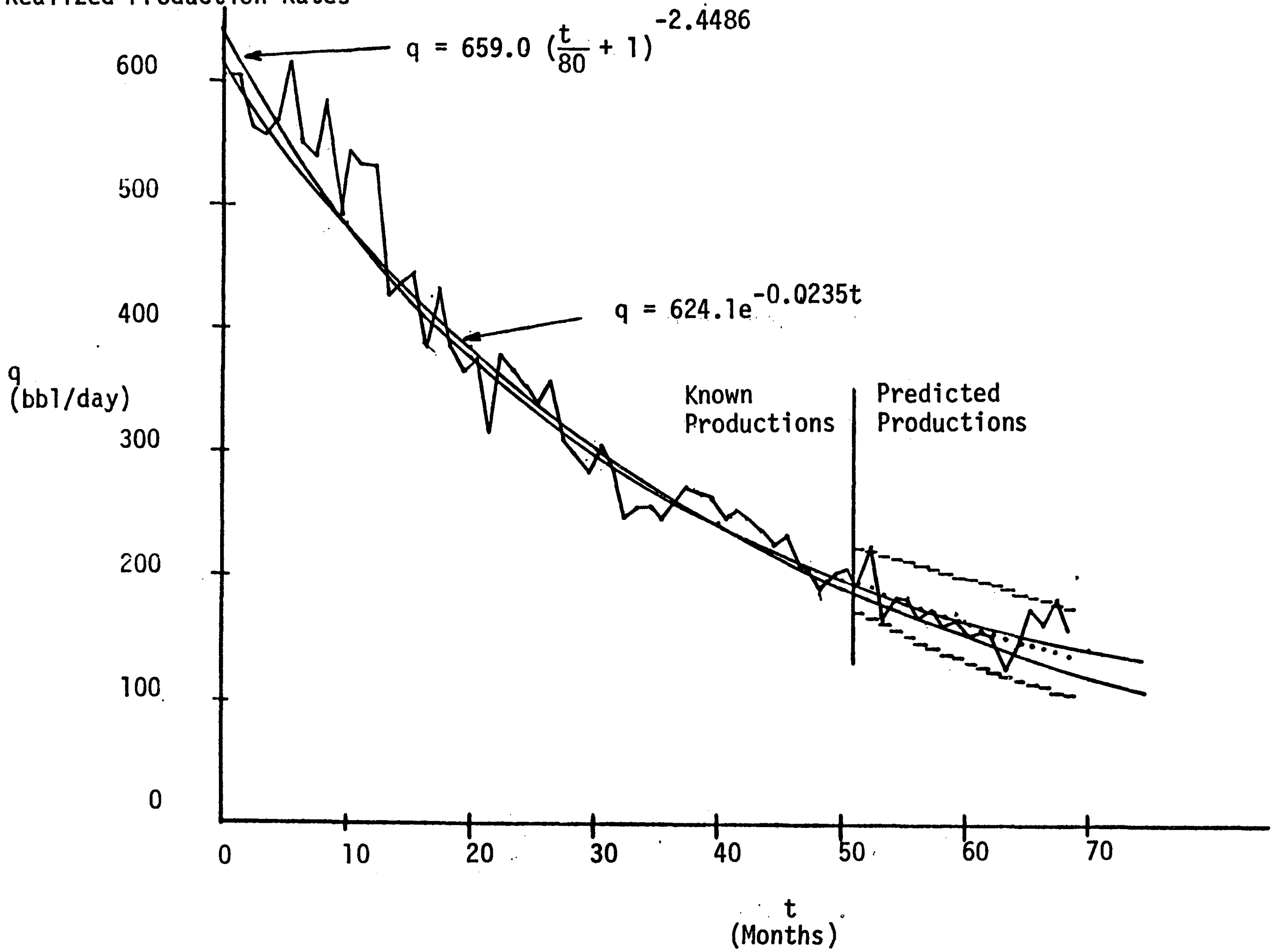
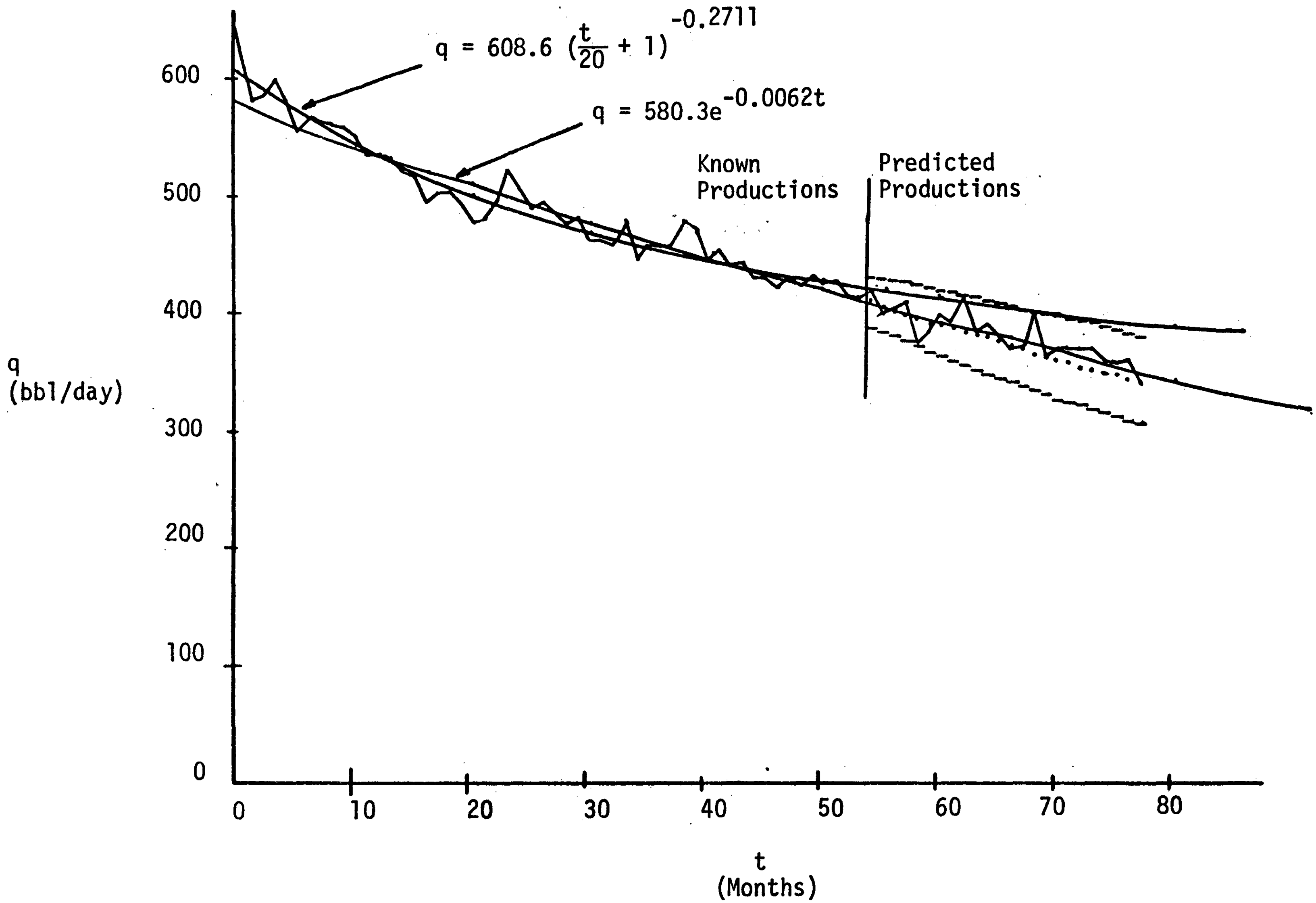


Figure VI. Average Daily Production Rates for Haas Madison Pool, North Dakota - Decline Curve Predictions, Time Series Predictions (with Associated 95% Probability Limits), and Realized Production Rates



methods of projections, total forecasted oil produced in a period of time was compared to the actual realized production in that same time period. The forecasted total amount of oil to be produced can be found by integrating the equations of the decline curves between the times of production. (See equations (2) and (6).) The time series predictions can simply be summed to yield their prediction of total production. Naturally, all of these figures require retransformation to a total barrel basis, since forecasts had been based on barrels per day values. Such a retransformation can very easily be made by multiplying the summed production rates by 30.4 - the average number of days per month. Projections of total oil to be produced, along with the actual productions, are listed in Table VI, and the corresponding percent errors in the predictions are given in Table VII.

The advantage of using time series models versus regression models is shown by Table VII. Future observations are assumed by extrapolative methods (regression) to decline as did their predecessors. Table VII shows that this assumption may not always be true (at least in the short run). Haas Madison Pool which was predicted by the regression (on the basis of Table IV) as having a hyperbolic decline is shown in Table VII to actually possess an exponential type of decline.

Table VII indicates that for the four sample cases, time series predictions were on the average more reliable than the best extrapolative predictions. The best extrapolative predictions were those made with the curve which "best" fit the production data using the criteria of Table IV. The application of time series analysis to many additional oil production declines would be necessary to substantiate any kind

Table VI. Predictions and Actual Productions of Total Oil from Foothills Pool, Beaver Lodge Madison Pool, North Black Slough Pool, and Haas Madison Pool (in barrels)

Pool	Years of Prediction	Predicted Productions			Actual Production
		Exponential	Hyperbolic	Time Series	
Foothills	2	177,363	191,064	186,450	188,626
Beaver Lodge Madison	2	1,175,848	1,176,694	1,191,636	1,181,739
North Black Slough	1.5	84,102	92,145	90,330	92,164
Haas Madison	2	280,154	298,406	276,029	280,802

Table VII. Percent Error of Predictions of Table VI.

Pool	Exponential	Hyperbolic	Predetermined Best Prediction	Time Series
Foothills	5.97	1.29	1.29	1.15
Beaver Lodge Madison	0.50	0.43	0.50	0.84
North Black Slough	8.75	0.02	0.02	1.99
Haas Madison	0.23	6.27	6.27	1.70
Total Error			8.08%	5.68%

of general conclusion concerning how much more accurate time series is at predicting future production; however, this study does show that time series methods can be applied to oil production decline analysis, resulting in reliable forecasts.

Besides providing a reliable forecast function whose form is determined directly by the data, time series methods provide the advantage of producing probability limits on individual forecasts. The probability limits give the production engineer an idea of the possible "spread" of future productions. Forecasts consisting of single projected values provide the production engineer with an idea of the magnitude of the production that may be realized, but yield no indication of how much the realized values may vary from the projection. A little knowledge of the size of the intervals in which production values most probably will lie, allows the production engineer to make both good optimistic and good pessimistic assessments of the oil reservoir. These in turn are very helpful in making future decisions concerning the oil property.

The 95% probability limits computed for the forecasts of the four sample reservoirs can be found in Appendix IV and are graphically shown in Figures III - VI. These 95% probability limits mean that given the production histories, there is a 95% probability that the actual future production, when it is realized, will lie within its associated probability interval. Unfortunately, these probability limits pertain only to individual forecasts and not jointly to all forecasts. It is therefore not possible to determine probability limits for total production forecasts using those probability limits of indivudual production forecasts (Box and Jenkins, 1970, p 138).

It should be pointed out that two of the derived hyperbolic decline curves - those of Foothills Pool and Haas Madison Pool - would not be considered as purè hyperbolic decline by many production engineers. Almost all the literature of today restricts the value of n in the hyperbolic decline equation (equation (4)) to a value somewhere between 0.0 and 1.0, inclusive. A least squares fit of the data of these two oil reservoirs yielded n = 1.155 for Foothills Pool, and n = 3.689 for Haas Madison Pool. Arps (1945, p 242) analyzed the data of many hyperbolic declines provided by Cutler in his 1924 paper, and came to the conclusion that n rarely exceeds 1.0 in value. However, Arps never limited the value of n to any range, and in fact stated that he himself had come across actual declines with n greater than 1.0. Unfortunately, many have misinterpreted Arps analysis, and thus much of today's literature concerning production decline curves gives the reader the impression that there is some physical law which limits the value of n to a value between the limits 0.0 and 1.0 inclusive. The regression derived equations yield the best hyperbolic fit possible. It seem inconsistent to exclude from an analysis certain types of curves when the types of curves accepted are themselves selected on the basis of no physical, governing law, but rather on what type of curve has traditionally given a good extrapolation forecast.

Extrapolation and time series methods each make it possible to obtain estimates of the remaining time that an oil field may economically produce, provided future production methods are not changed. The procedure for obtaining such estimates is simply to "follow" the predicted declining oil production rates until they reach an economic limit or minimum production rate which is economical. (See Figure I.) This is equivalent to equating q to the minimum economic production rate limit and solving for t in equation (1) if the decline is exponential, or in equation (5) if the decline is hyperbolic. For the time series predictions, it is necessary to actually trace the predictions to the economic limit. The time required to reach the economic limit is, of course, the estimate of the remaining productive life for the reservoir.

Once the remaining productive life has been determined, it is possible to estimate the total remaining oil that can be pumped, given constant pumping methods. To make such an estimate, it is necessary to perform the integration of the regression decline curve (either exponential or hyperbolic) over the remaining time of economic production for the field. Alternately, it is also possible to simply substitute values in equation (2) if the decline is exponential or equation (6) if the decline is hyperbolic and automatically obtain the estimate of remaining oil that will be produced. For both equations (2) and (6) t_o is equal to the present period and t_f is equal to the period in which the economic production rate is reached. Time series predictions again are summed over the remaining productive life to estimate total future production.

It is impossible to compare the accuracy of the predicted estimates of economic life and remaining production for the four sample cases, since the actual lives and remaining productions have not yet been realized. It is possible, however, to make comparisons of these estimates

Table VIII. Projected Economic Lives (in Months) of Foothills Pool, Beaver Lodge Madison Pool, North Black Slough Pool, and Haas Madison Pool Using Economic Production Rate Limit of 100 bbl/day.

Pool	Exponential	Hyperbolic	Time Series
Foothills	106.8	306.9	115.0
Beaver Lodge Madison	191.5	193.9	193.0
North Black Slough	27.0	41.8	34.0
Haas Madison	229.7	15,568.4	186.0

provided by the production decline curves and time series methods. Arbitrarily setting the economic production rate at 100 bbl/day for each of the four sample cases gives the expected economic lives and expected remaining future productions for the four fields listed in Tables VIII and IX respectively.

The values listed in Table VIII again show the problems that can arise using an extrapolation to forecast future behavior. A good fit for the known data does not assure that the actual future behavior will follow the extrapolated curve. Both exponential and hyperbolic curves fit the data well (as exhibited by Table III); however, the predicted lives of the reservoirs using the two types of curves vary significantly. This results in significant differences in estimated future production as shown in Table IX. Unfortunately, it is impossible to check these predictions, but it appears that hyperbolic decline - more often the best fitting of the decline curves - gives inflated values of economic productive life and expected remaining production.

Table IX. Projected Remaining Production (in barrels) of Foothills Pool, Beaver Lodge Madison Pool, North Black Slough Pool, and Haas Madison Pool Using an Economic Production Rate Limit of 100 bbl/day.

Pool	Exponential	Hyperbolic	Time Series
Foothills	559,100	1,477,900	615,400
Beaver Lodge Madison	3,597,400	3,628,500	3,642,800
North Black Slough	114,600	179,000	146,400
Haas Madison	1,541,300	63,912,000	1,246,600

Conclusion

This study has shown that time series analysis provides reliable forecasts of production decline for oil reservoirs. The four examples used in the study cannot by themselves be considered as sufficient evidence to suggest that time series analysis yields more accurate or less accurate forecasts than the traditional extrapolative approaches. Both procedures would have to be applied to many more production declines before any kind of generalization of this type could be substantiated. The time series forecasts were shown to be sufficiently accurate. In addition, though, time series analysis provides probability limits for individual forecasts which can be very important to the production engineer in making decisions concerning the economic future of the reservoir. Time series analysis does not require the forecaster to make an initial specification of the form of the model, but allows the data itself to specify the model form. Finally, time series forecasts are developed by a preferable method of assuming past interdependence among the observations to continue to exist in the future rather than assuming that simple extensions of past trends will yield valid forecasts.

It may seem to some that the procedures used in this study to forecast production decline are overly complicated and the extra computation not really worth the extra information obtained. Such an argument is misleading, as the time series computations may be done quickly and efficiently using modern computers. It should not be assumed that the purpose of this study has been to present a computerized procedure that will remove the subjectivity required to make an oil property evaluation. There is no substitute for the knowledge obtained by the oil property appraiser over his many years of oil property valuation. The additional information provided by a time series analysis of a production decline should be considered by the oil property appraiser, and should aid him in making good decisions about a property.

Appendix I

A Step-by-Step Algorithm for Development of a Preliminary Nonseasonal Time Series Model

It has become apparent to this author during the undertaking of this study that relatively few mathematicians, let alone economists and engineers, have even preliminary knowledge of the mechanics or uses of modern time series analysis. The purpose of this appendix is to present the preliminary analysis which must be undertaken to develop a time series model in a simple step-by-step format. The reader familiar with elementary computer programming will quickly recognize the great ease with which most of these steps may be programmed.

It must be remembered that this algorithm only results in a preliminary forecasting function. Further refinements of the estimated parameter values are possible and necessary to obtain reliable forecasts. These refinement techniques involve a back forecasting procedure and the use of numerical derivatives requiring small perturbations of the parameter estimates in the preliminary model. To properly present these methods here would require considerable space and would do away with the initial goal of simplicity for this appendix. Also, diagnostic checking of the fitted model should be undertaken to check model adequacy and suggest possible refinements of the model. There are several of these checks available with varying degrees of difficulty of implementation and informative output. The reader interested in the final parameter estimation techniques and diagnostic checking of the model is referred to chapters 7 and 8 of <u>Time Series Analysis Forecasting and Control</u> by Box and Jenkins.

Given: a set of observed values which have been transformed into stationary series of T observations (W_1, W_2, W_3, ... , W_T). (The concept of stationarity is introduced and explained in the main portion of this paper.)

Goal: develop a model relating future observations of the stationary series to past observations and past disturbances (white noise) of the form

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots + \theta_0 \qquad (1)$$

where

W_t is the value of the stationary observation occurring at time t.

ϕ_i is the value of the coefficient of W_{t-i}

a_t is the value of the independant, identically distributed white noise at time t.

θ_j is the value of the coefficient of a_{t-j}

θ_o is the value of a constant term

The above model is to be used to predict future observations, that is W_{t+1}, W_{t+2}, ... To do so, parameter estimates for ϕ_i, θ_j, and θ_o may be determined from the data at hand by following the procedures outlined below. In addition, the estimated parameter value for the variance of white noise (σ_a^2) is a desirable quantity. An estimate of this parameter is produced.

STEP 1 Compute a approximation of the mean for the process, $\overline{W}$.

$$\overline{W} = \frac{1}{T} \sum_{t=1}^{T} W_t \qquad (2)$$

STEP 2. Compute the sample autocovariance function consisting of C_k for $k = 0, 1, 2, \ldots, \frac{T}{4}$. This sample autocovariance function displays the average relationship between observations k periods apart. That is, if higher than average observations (having value greater than $\overline{W}$) are generally followed k periods later by higher than average observations or if lower than average observations (having value less than $\overline{W}$) are generally followed k periods later by lower than average observations then C_k will be positive. If lower than average observations tend to be followed k periods later by higher than average observations or if higher than average observations tend to be followed k periods later by lower than average observations, then C_k will be negative.

$$C_k = \frac{1}{T} \sum_{i=1}^{T-k} (W_i - \overline{W})(W_{i+k} - \overline{W}) \qquad (3)$$

Actually, it is not absolutely necessary that exactly $\frac{T}{4}$ autocorrelations be computed. This also holds for other functions that will be computed in the remaining steps which call for $\frac{T}{4}$ values. The value $\frac{T}{4}$ should only be used as an upper bound for the number of autocovariances which should be computed for the analysis.

STEP 3 Compute the sample autocorrelation function consisting of r_k for $k = 0, 1, 2, \ldots, \frac{T}{4}$. This sample autocorrelation function is simply a standardization of the autocovariance function of STEP 2, so that all values of the autocorrelation function will lie within the range, $-1 \leq 0 \leq 1$. The interpretation of the algebraic sign of the autocorrelation function is exactly as it was for the autocovariance function. However, the standardization conducted for the autocorrelation function allows one to compare the relative strengths of the relationships between the observations of k periods of lag. (A value close to -1.0 or 1.0 indicates a strong linear relationship or correlation between the two observations while a value close to 0.0 indicates a very weak relationship or correlation.)

$$r_k = \frac{C_k}{C_o} \qquad (4)$$

STEP 4 Compute an estimate of the partial autocorrelation function consisting of $\hat{\phi}_{kk}$ for $k = 1, 2, 3, \ldots, \frac{T}{4}$. This function will be useful when it becomes necessary to decide how many past observations a future observation is directly dependant upon. This is equivalent to determining the necessary number of ϕ's (in equation (1)) which should be included in the model. The partial autocorrelation function is found by sequentially solving sets of simultaneous equations. That is:

for $k = 1$ the equation solved for $\hat{\phi}_{11}$ is

$$r_1 - \hat{\phi}_{11} r_o = 0$$

where r_o and r_1 are obtained from STEP 3.For $k = 2$ the equations to be solved for $\hat{\phi}_{22}$ are

$$r_1 - \hat{\phi}_{11} r_o - \hat{\phi}_{22} r_1 = 0$$

$$r_2 - \hat{\phi}_{11} r_1 - \hat{\phi}_{22} r_o = 0$$

where $\hat{\phi}_{11}$ and $\hat{\phi}_{22}$ are the unknowns and r_o, r_1, and r_2 are obtained from STEP 3. Generally, to determine $\hat{\phi}_{kk}$ the equations to be solved are

$$r_1 - \hat{\phi}_{11} r_o - \hat{\phi}_{22} r_1 - \hat{\phi}_{33} r_2 - \cdots - \hat{\phi}_{kk} r_{k-1} = 0$$

$$r_2 - \hat{\phi}_{11} r_1 - \hat{\phi}_{22} r_o - \hat{\phi}_{33} r_1 - \hat{\phi}_{44} r_2 - \cdots - \hat{\phi}_{kk} r_{k-2} = 0$$

$$r_3 - \hat{\phi}_{11} r_2 - \hat{\phi}_{22} r_1 - \hat{\phi}_{33} r_o - \hat{\phi}_{44} r_1 - \cdots - \hat{\phi}_{kk} r_{k-3} = 0$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$r_k - \hat{\phi}_{11} r_{k-1} - \hat{\phi}_{22} r_{k-2} - \hat{\phi}_{33} r_{k-3} - \cdots - \hat{\phi}_{kk} r_o = 0$$

Thus, a set of k equations will have to be solved to determine each member of the function, $\hat{\phi}_{kk}$. The derivation of these equations would require considerable space. For the purposes of this paper it is probably only necessary to further point out that if the actual autocorrelations (ρ_k) and actual coefficients of the past observations (ϕ_i) were known, the above equations would be satisfied with ρ_k substituted for r_k and ϕ_k substituted for $\hat{\phi}_{kk}$.

Note - When solving the sets of equations do not substitute previously determined $\hat{\phi}_{kk}$'s from other sets of equations. This will result in incorrect partial autocorrelation estimates. For example, when solving for $\hat{\phi}_{22}$ it is incorrect to immediately substitute the previously determined $\hat{\phi}_{11}$ into the two equations. This is because values of $\hat{\phi}_{kk}$ are not conditioned on previously determined values, $\hat{\phi}_{k-1\ k-1}$, $\hat{\phi}_{k-2\ k-2}$, $\cdots$. It is necessary to treat $\hat{\phi}_{11}$ as an unknown just as $\hat{\phi}_{22}$ is treated when solving the equations. However, only $\hat{\phi}_{22}$ is included as a member of the partial autocorrelation function since $\hat{\phi}_{11}$ was determined in the first equation solved. This idea is extended each time a set of equations is solved. That is, for each i, $\hat{\phi}_{ii}$ is assumed as unknown, but for each k, the solving of k equations determines only $\hat{\phi}_{kk}$ as a new member of the autocorrelation function.

STEP 5 Determine the behavior of the sample autocorrelation function (r_0, r_1, r_2, ... $r_{\frac{T}{4}}$) . Specifically, it is necessary to determine whether 1) the sample autocorrelation function gradually diminishes or "dies off" to a value of zero, or 2) the sample autocorrelation function suddenly has all terms effectively zero after a certain point, that is it "cuts off". If the sample autocorrelation function exhibits the cutoff property, the exact point of cutoff must be determined. Some judgement must be used in making these decisions. However, the use of hypothesis testing takes away a portion of the subjectivity involved in determining the point of the cutoff. The hypothesis test does however require the user to assume a value for α, the level of significance or simply the percentage of time the user is willing to concede that he will wrongly reject his null hypothesis (Type I error). For all uses in this paper α is assumed to be equal to 0.05.

The testing procedure consists of checking the r_k with small values to determine whether ρ_k, ρ_{k+1}, ρ_{k+2}, ..., $\rho_{\frac{T}{4}}$, ... are significantly different from zero. To check any r_k in this manner use STEP 5A and STEP 5B below.

STEP 5A Compute the standard error (standard deviation of the sampling distribution) of r_k

$$SE(r_k) = \frac{1}{\sqrt{T}} (1 + 2r_1^2 + 2r_2^2 + \dots 2r_{k-1}^2)^{1/2}$$

STEP 5B Test the null hypothesis,

H_0: ρ_k, ρ_{k+1}, ρ_{k+2}, ... = 0 against the alternative hypothesis.

H_1: $\rho_k, \rho_{k+1}, \rho_{k+2}, \ldots \neq 0$

This procedure tests whether one is justified in claiming that actual autocorrelations after ρ_{k-1} have values of zero. The null hypothesis is rejected if a significant number (greater than 5% for $\alpha = 0.05$) of the estimates of $\rho_k, \rho_{k+1}, \rho_{k+2}, \ldots$, (that is, $r_k, r_{k+1}, r_{k+2}, \ldots$,) lie outside a two standard error interval about zero. Especially important to the rejection of the null hypothesis is the failure of the first few sample autocorrelations being tested to lie within this interval. If the null hypothesis is tested and cannot be rejected, the autocorrelation cutoff has been found, and further testing of sample autocorrelations is no longer necessary. This is because all remaining autocorrelations have passed the test for having value zero.

STEP 6 Determine the behavior of the sample partial autocorrelation function ($\hat{\phi}_{11}, \hat{\phi}_{22}, \ldots \hat{\phi}_{\frac{T}{4}\frac{T}{4}}$). The procedure for doing this is very similar to that just presented in STEP 5 for the autocorrelation function. Again, it must be determined whether the function shows a slow diminishing or "dying off" to a value of zero, or whether it suddenly becomes zero or "cuts off" at some point.

Hypothesis testing is again used in making judgements concerning the point of cutoff. The sample partial autocorrelation function is only an estimate of the actual partial autocorrelation function. The actual partial autocorrelation function is unknown but consists of elements that are denoted as $\phi_{11}, \phi_{22}, \phi_{33}, \ldots, \phi_{\frac{T}{4}\frac{T}{4}}$. It is necessary to check for $\hat{\phi}_{kk}$ by the following procedure so that it may be determined whether ϕ_{kk}, $\phi_{k+1\ k+1}$, ..., $\phi_{\frac{T}{4}\frac{T}{4}}$ may be considered as being equal to zero.

STEP 6A Compute the standard error of $\hat{\phi}_{kk}$ (that is the standard deviation of the sampling distribution of $\hat{\phi}_{kk}$) under the null hypothesis of STEP 6B.

$$SE(\hat{\phi}_{kk}) = \frac{1}{\sqrt{T}}$$

STEP 6B Test the null hypothesis

$$H_o: \phi_{k+1\ k+1}, \phi_{k+2\ k+2}, \ldots, \phi_{\frac{T}{4}\frac{T}{4}} = 0$$

against the alternate hypothesis

$$H_1: \phi_{k+1\ k+1}, \phi_{k+2\ k+2}, \ldots, \phi_{\frac{T}{4}\frac{T}{4}} \neq 0$$

That is, test whether it is valid to conclude that actual partial autocorrelations after ϕ_{kk} have a value of zero.

The null hypothesis is rejected if a significant number (greater than 5% if α is set at 0.05) of the estimates of $\phi_{k+1\ k+1}, \phi_{k+2\ k+2}, \ldots, \phi_{\frac{T}{4}\frac{T}{4}}$ (that is $\hat{\phi}_{k+1\ k+1}$, $\hat{\phi}_{k+2\ k+2}$, ..., $\hat{\phi}_{\frac{T}{4}\frac{T}{4}}$) lie outside a two standard error interval about zero. Again it is especially important that the first few of these partial autocorrelations lie within the interval. If the null hypothesis is tested and cannot be rejected from the evidence at hand, the cutoff point is determined. Further testing of the sample partial autocorrelations is unnecessary because the remaining partial autocorrelations have passed the test for having value zero.

STEP 7 Preliminary identification of the model to be used in the analysis is now possible by noting the combined behavior of the autocorrelation function and the partial autocorrelation function. Three combinations of behavior for these functions are possible.

1) For a particular q, a sharp cutoff of the autocorrelation function after r_q and a slow dying off of the partial autocorrelation function indicates a pure moving average model of order q [in symbols, MA(q)]. A pure moving average model relates future predicted observations to strictly past disturbances or white noise (a_t). The form of the model is

$$W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} + \theta_o$$

2) For a particular p, a sharp cutoff of the partial autocorrelation function after $\hat{\phi}_{pp}$ and a slow dying off of the autocorrelation function indicates a pure autoregressive model of order p [in symbols AR(p)]. A pure autoregressive model relates future observations to only past observations. The pure autoregressive model is of the form

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + a_t + \theta_o$$

3) For a particular p and a particular q an eventual dying out of both the autocorrelation function and the partial autocorrelation function indicates a mixed model containing both autoregressive and moving average components [symbolically ARMA (p,q)]. This mixed model, of course, relates future observations to both past disturbances and past observations. The form of the mixed model is

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \cdots - \theta_q a_{t-q} + \theta_o$$

Determination of the values for p and q is not nearly as easy as it was when these parameters were estimated in the pure moving average and pure autoregressive models. In general, though, either the autocorrelation function or partial autocorrelation function will die immediately, and the other function will have a period of irregular values before it dies. This period of irregular behavior is determined by the difference between the values of p and q. Specifically, if the autocorrelation function starts dying immediately and the partial autocorrelation function begins to die only after L periods, $p - q = L$. Similarly, when the partial autocorrelation function dies immediately and the autocorrelation function begins dying after L periods, then $q - p = L$. It is possible therefore, to determine the relative values of q and p. Several models having different test values for p and q may be needed to determine the model which "best fits" the data. It should be pointed out that most time series models are quite simple with neither p nor q exceeding 2 in value. The behavior of mixed models is discussed further in Box and Jenkins, p 76-78.

Note - If the autocorrelation function shows a very slow dying out characteristic covering many periods, the data has not been properly transformed to exhibit stationarity. Another transformation should be made. (See the discussion of stationarity in the main body of this paper).

STEP 8 Determine initial estimates of the model parameters. Estimates of parameters will be denoted with "^".

For Pure Autoregressive Model

If it was determined in STEP 7 that a pure autoregressive model of order p is most representative of the process then $\phi_1, \phi_2, \cdots, \phi_p$, σ_a^2, and θ_0 must be estimated. To obtain initial estimates of $\phi_1, \phi_2, \cdots, \phi_p$, and σ_a^2, solve the equations

$$C_o - \hat{\phi}_1 C_1 - \hat{\phi}_2 C_2 - \hat{\phi}_3 C_3 - \ldots - \hat{\phi}_p C_p = \hat{\sigma}_a^2$$

$$C_1 - \hat{\phi}_1 C_o - \hat{\phi}_2 C_1 - \hat{\phi}_3 C_2 - \ldots - \hat{\phi}_p C_{p-1} = 0$$

$$C_2 - \hat{\phi}_1 C_1 - \hat{\phi}_2 C_o - \hat{\phi}_3 C_1 - \ldots - \hat{\phi}_p C_{p-2} = 0$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$C_p - \hat{\phi}_1 C_{p-1} - \hat{\phi}_2 C_{p-2} - \hat{\phi}_3 C_{p-3} - \ldots - \hat{\phi}_p C_o = 0$$

It can be shown that if the actual covariances (γ_k) rather than the sample covariances (C_k) are known and actual autoregressive coefficients (ϕ_j) rather than the estimates ($\hat{\phi}_j$) are known, then the above equations must hold. Substitution of sample covariances for actual covariances permits one to therefore compute estimates for the autoregressive coefficients and the variance of white noise (σ_a^2).

To compute the estimate of θ_o, solve the equation

$$\hat{\theta}_o = \overline{W}(1 - \hat{\phi}_1 - \hat{\phi}_2 - \ldots - \hat{\phi}_p)$$

For Pure Moving Average Model

If it was determined in STEP 7 that a pure moving average model of order q should be used, then θ_1, θ_2, ...,θ_q, σ_a^2 , and θ_o must be estimated. An iterative scheme is used to determine the initial estimates for these parameters. The iterative procedure is as follows:

STEP 8A Initialize all $\hat{\theta}_j$, j = 1, 2, ..., q by giving each a value of zero.

STEP 8B Solve for the estimate of the variance of white noise

$$\hat{\sigma}_a^2 = \frac{C_o}{1 + \hat{\theta}_1^2 + \hat{\theta}_2^2 + \ldots \hat{\theta}_q^2}$$

STEP 8C Solve for estimates of $\theta_1, \theta_2, \ldots, \theta_q$, by solving the equations

$$\hat{\theta}_q = \frac{-C_q}{\hat{\sigma}_a^2}$$

$$\hat{\theta}_{q-1} = \frac{-C_{q-1}}{\hat{\sigma}_a^2} + \hat{\theta}_1\hat{\theta}_q$$

$$\hat{\theta}_{q-2} = \frac{-C_{q-2}}{\hat{\sigma}_a^2} + \hat{\theta}_1\hat{\theta}_{q-1} + \hat{\theta}_2\hat{\theta}_q$$

. . . .

$$\hat{\theta}_1 = \frac{-C_1}{\hat{\sigma}_a^2} + \hat{\theta}_1\hat{\theta}_2 + \hat{\theta}_2\hat{\theta}_3 + \ldots \hat{\theta}_{q-1}\hat{\theta}_q$$

STEP 8D Check to see if $\hat{\theta}_1, \hat{\theta}_2, \ldots \hat{\theta}_p$ have changed significantly in this iteration from their values in the previous iteration. If these changes are within the user's specified tolerance, then go on to STEP 8E. Otherwise return to STEP 8B and continue with the next iteration.

STEP 8E Determine the value for the constant, θ_o. Set $\hat{\theta}_o = \overline{W}$.

For Mixed Autoregressive Moving Average Model

If it was determined in STEP 7 that a mixed autoregressive moving average model of order (p,q) should be used to model the data, then $\phi_1, \phi_2, \ldots \phi_p, \theta_1, \theta_2, \ldots \theta_q, \sigma_a^2$, and θ must be estimated. To obtain estimates of $\phi_1, \phi_2, \ldots \phi_p$, solve the p equations similar to those solved for the pure autoregressive case.

$$C_k - \hat{\phi}_1 C_{k-1} - \hat{\phi}_2 C_{k-2} - \dots - \hat{\phi}_p C_{k-p} = 0$$

$$C_{k+1} - \hat{\phi}_1 C_k - \hat{\phi}_2 C_{k-1} - \dots - \hat{\phi}_p C_{k-p+1} = 0$$

. . . .

$$C_{k+p} - \hat{\phi}_1 C_{k+p-1} - \hat{\phi}_2 C_{k+p-2} - \dots - \hat{\phi}_p C_k = 0$$

for $k = q + 1$.

Estimates of θ_1, θ_2, ... θ_q, and σ_a^2 can be found by using the iterative scheme of STEP 8A through STEP 8D for the pure moving average model. Estimation of θ_o is found as in the autoregressive model, that is

$$\hat{\theta}_o = \overline{W}(1 - \hat{\phi}_1 - \hat{\phi}_2 - \dots \hat{\phi}_p)$$

STEP 9 Determine whether θ_o can be assumed as equal to zero. Some simplification is possible if θ_o may be considered as having a value of zero. To check whether this assumption may be made, it is necessary to find the variance of $\overline{W}$.

$$\text{Var}(\overline{W}) = \frac{\hat{\sigma}^2(W)}{T} \left(1 + 2 \sum_{i=1}^{\infty} r_i\right) \quad (4)$$

where

$$\hat{\sigma}^2(W) = \frac{1}{T-1} \sum_{i=1}^{T} (W_1 - \overline{W})^2$$

T = number of data points

r_i = sample autocorrelation for two observations i periods apart

$\overline{W}$ is defined as in equation (2)

Assume $\overline{W}$ is distributed normally. Then it is only necessary to check whether the computed value of $\overline{W}$ lies within a two standard error interval about zero (if α is set equal to 0.05). That is whether

$$-2(\text{Var}(\overline{W}))^{1/2} \leq \overline{W} \leq 2(\text{Var}(\overline{W}))^{1/2} .$$

If $\overline{W}$ is within the two standard error interval it is impossible to say that $\overline{W}$ is significantly different from zero and it is therefore set equal to zero. This means that $\hat{\theta}_0$ may also be set equal to zero. If $\overline{W}$ lies outside of this two standard error interval, then $\hat{\theta}_0$ must be considered as having a value significantly different from zero and obviously must be included in the model.

STEP 10 Forecasting from the model. A model has now been formulated with the desired form

$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \ldots + \phi_p W_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} + \theta_0$ with $\hat{\phi}_i$ substituted for ϕ_i and $\hat{\theta}_i$ substituted for θ_i. To forecast future W_t's, as one may guess, substitution of past W_t's that have been either observed or predicted themselves is necessary. Thus, the values for W_{t-1}, W_{t-2}, ... , W_{t-p}, are quite easily obtained.

It is a little more difficult, however, to arrive at values for the independant, identically distributed white noise, that is a_t, a_{t-1}, ... , a_{t-q}. The white noise is actually the random disturbance component in the observations which have occurred or will occur. For disturbances, a_t, which have occurred $a_t = W_t(\text{actual}) - W_t(\text{predicted})$. If W_t(actual) is an observation which has not yet been realized, then the disturbance, a_t, has not yet occurred. For such disturbances a_t is set equal to zero, its expected value. It is necessary to build up a listing of past disturbances in order to have good approximates for the white noise in the predictions for the moving average and mixed autoregressive moving average models. The procedure for doing this is illustrated in the sample problem of Appendix III.

Appendix II

Sample Problem Number 1

The Pure Autoregressive Model

To illustrate the algorithm of Appendix I consider the following observed data.

t time period	W_t observation
1	110
2	92
3	107
4	90
5	116
6	99
7	104
8	92
9	103
10	87
11	115
12	96
13	104
14	84
15	101

It is desired to develop a model which will predict future observations, that is, predict W_{16}, W_{17}, W_{18}, ... The assumption is made that the given observations form a stationary process and that no further stationary transformations are necessary.

Actually, at least 50 known observations are needed to perform a proper time series analysis, but this short series is ideal for illustrating the algorithm of Appendix I. Method:

STEP 1 Using equation (2) to compute an approximation of the mean for the series results in

$$\overline{W} = \frac{1}{T} \sum_{t=1}^{15} W_t = \frac{1}{15}(1500) = 100$$

STEP 2 The sample autocovariance function for k = 0, 1, 2, ... 8 is computed using equation (3), and results in

$C_0 = 89.47$

$C_1 = -55.47$

$C_2 = 45.67$

$C_3 = -59.07$

$C_4 = 44.27$

$C_5 = -30.67$

$C_6 = 43.07$

$C_7 = -32.33$

$C_8 = 24.20$

Note - The $\frac{T}{4}$ upper bound restriction was not used in this sample problem for purposes of illustration.

STEP 3 The corresponding sample autocorrelation function from equation (4) results in

$r_o = 1$

$r_1 = -0.620$

$r_2 = 0.510$

$r_3 = -0.660$

$r_4 = 0.495$

$r_5 = -0.343$

$r_6 = 0.481$

$r_7 = -0.362$

$r_8 = 0.271$

STEP 4 The estimate of the partial autocorrelation function can now be found. For $\hat{\phi}_{11}$, solving the corresponding equation yields

$\hat{\phi}_{11} = r_1 = -0.620.$

For $\hat{\phi}_{22}$, the 2 equations solved are

$-0.620 - \hat{\phi}_{11} + 0.62\,\hat{\phi}_{22} = 0$

$0.510 + 0.62\,\hat{\phi}_{11} - \hat{\phi}_{22} = 0$

yielding

$\hat{\phi}_{22} = 0.204$

In similar fashion the remainder of the sample partial autocorrelation function is found

$\hat{\phi}_{33} = -0.477$

$\hat{\phi}_{44} = -0.102$

$\hat{\phi}_{55} = 0.123$

$\hat{\phi}_{66} = 0.181$

$\hat{\phi}_{77} = 0.123$

$\hat{\phi}_{88} = 0.084$

STEP 5 The behavior of the sample autocorrelation function of STEP 3 must now be analyzed. It is easily observed that the autocorrelations do not show any point of sudden cutoff. Instead they seem to be slowly dying off as they alternate in sign. One should compute the standard errors for the sample autocorrelations and confirm they actually are significant by means of the hypothesis test of STEP 5A and STEP 5B.

STEP 6 The sample partial autocorrelation function of STEP 4 seems to show a cutoff characteristic. There is a question, however, whether the sample autocorrelation cuts off after $\hat{\phi}_{11}$ or after $\hat{\phi}_{33}$. Essentially we need to know whether the value of 0.204 for $\hat{\phi}_{22}$ was an unusually low sample observation of ϕ_{22}, or whether the -0.477 was an unusually high sample observation of ϕ_{33}.

To help make a determination of the true cutoff point, the hypothesis test of STEP 6A and STEP 6B is conducted.

STEP 6A The standard error of $\hat{\phi}_{kk}$ is computed as

$$SE(\hat{\phi}_{kk}) = \frac{1}{\sqrt{15}} = .258.$$

STEP 6B For the testing of whether the cutoff occurs after $\hat{\phi}_{11}$, the corresponding null and alternative hypothesis are

$$H_0: \phi_{22}, \phi_{33}, \phi_{44}, \cdots, \phi_{88} = 0$$

$$H_1: \phi_{22}, \phi_{33}, \phi_{44}, \cdots, \phi_{88} \neq 0 .$$

Because all the estimates of the partial autocorrelations being tested lie within a two standard error interval about zero, that is they lie within the interval

$$-0.516 \leq \hat{\phi}_{kk} \leq 0.516 \qquad k > 1$$

the null hypothesis is not rejected at $\alpha = 0.05$. Thus, $\phi_{22}, \phi_{33}, \cdots, \phi_{88}$ may all be considered as equal to zero (even though the statistician would like to reserve judgement about accepting the null hypothesis) and the cutoff can be considered as occurring after ϕ_{11}. If the null hypothesis had been rejected by the hypothesis test, then further testing would have been required to check for the actual cutoff point. Even though the null hypothesis was not rejected, the time series analyst would probably see what type of model would result if he would assume the cutoff to occur after $\hat{\phi}_{33}$. This procedure is known as "overfitting" the model since such action consists of testing a more complex model (that is AR(3) instead of the original AR(1)). This new model contains two more parameters which must be determined by the remaining steps of the algorithm. The significance of these additional parameters determine whether the AR(3) model is preferable to the AR(1) model.

STEP 7 Preliminary identification of the model to be used to forecast future observations is now possible. Checking the characteristics of the models listed in STEP 7 of Appendix I would indicate that a pure autoregressive model of order 1 (AR(1)) best fits the case at hand. That is, the sample autocorrelation function dies slowly, and the sample partial autocorrelation function cuts off after $\hat{\phi}_{11}$. The model to be used then is of the form

$$W_t = \phi_1 W_{t-1} + a_t + \theta_o$$

STEP 8 Following the procedure for estimation of parameters for the pure autoregressive model, it is possible to estimate ϕ_1,

σ_a^2, and θ_o. Solving the equations for the estimates of ϕ_1 and σ_a^2 results in

$$89.47 + 55.47\ \hat{\phi}_1 = \hat{\sigma}_a^2$$

$$-55.47 - 89.47\ \hat{\phi}_1 = 0$$

$$\hat{\phi}_1 = -0.62$$

$$\hat{\sigma}_a^2 = 55.08.$$

Solving for the estimate of θ_o gives

$$\hat{\theta}_o = 100(1 + .62)$$

$$= 162$$

STEP 9 To determine whether θ_o can be assumed as equal to zero, the variance of $\overline{W}$ is computed from equation 4 of Appendix I.

$$Var(\overline{W}) = \frac{95.86}{15}(1 + 2(-.288))$$

$$Var(\overline{W}) = 3.48 \quad .$$

For θ_o to be considered as having value zero, $\overline{W}$ must lie within the two standard error interval

$$-3.73 \leq \overline{W} \leq 3.73.$$

Since the value of $\overline{W}$ is 100, it can be seen that θ_o must be considered as nonzero in the model.

STEP 10 The model developed is then

$$W_t = -0.62W_{t-1} + a_t + 162.$$

Forecasting from this model is a very straightforward procedure. The white noise, a_t, in the model is set to its expected value of zero, because this is a future disturbance that cannot be determined until the observation being predicted is actually realized. Therefore, to predict future observations, the model

$$W_t = -0.62W_{t-1} + 162$$

is used.

Prediction of the next 5 future observations for the original series are computed below.

$$W_{16} = -0.62(101) + 162 = 99.38$$

$$W_{17} = -0.62(99.38) + 162 = 100.38$$

$$W_{18} = -0.62(100.38) + 162 = 99.76$$

$$W_{19} = -0.62(99.76) + 162 = 100.15$$

$$W_{20} = -0.62(100.15) + 162 = 99.91$$

Appendix III

Sample Problem Number 2

Forecasting of Future Oil Production Using a Pure Moving Average Model

To illustrate the procedure that was used in this study to forecast future oil production, the following sample problem is presented. The initial raw data consists of average daily productions found in Table I for Foothills Pool, North Dakota. Obviously this raw data does not form a stationary process, as the observations show a definite downward trend. As mentioned in the body of this thesis, a transformation including differences of the natural logarithms of the raw oil production data results in the formation of a stationary series. The data base resulting from this transformation is listed in Table AI.

STEP 1 Equation (1) gives the approximation of the mean of the transformed series as

$\overline{W} = -0.009262$

STEP 2 The sample autocovariance function for k = 0, 1, 2, ... 18 is found using equation (2).

$C_0 = 0.0015116$

$C_1 = -0.0006530$

$C_2 = -0.0000514$

$C_3 = 0.0000242$

$C_4 = 0.0001451$

.

.

.

$C_{17} = -0.0001602$

$C_{18} = 0.0002902$

Table AI. Differences of Natural Logarithms of Average Daily Production for Foothills Pool, North Dakota

t	W_t	t	W_t	t	W_t
1	-0.05754	25	-0.00925	49	-0.01123
2	0.03199	26	-0.03738	50	-0.06542
3	-0.02103	27	-0.00739	51	-0.03014
4	-0.01227	28	0.08786	52	0.02896
5	-0.02539	29	-0.07515	53	-0.00561
6	-0.00809	30	-0.02792	54	0.02456
7	0.00000	31	0.05582	55	-0.14854
8	-0.00349	32	-0.04986	56	0.06642
9	-0.02437	33	-0.03014	57	-0.06274
10	-0.06450	34	-0.01232	58	-0.01615
11	-0.00724	35	-0.01959	59	-0.00136
12	-0.02139	36	0.00794	60	-0.03421
13	0.04525	37	-0.05548	61	-0.00529
14	-0.03655	38	0.05711	62	0.09113
15	-0.00260	39	-0.01868	63	-0.04960
16	0.01529	40	0.01541	64	-0.00646
17	-0.05401	41	-0.04324	65	-0.02034
18	0.00988	42	0.04406	66	0.00729
19	-0.04828	43	-0.01207	67	0.00448
20	-0.02324	44	0.01561	68	-0.00552
21	0.02488	45	-0.00901	69	-0.01359

t	W_t	t	W_t	t	W_t
22	-0.01272	46	-0.02837	70	0.01186
23	-0.04558	47	-0.03386	71	-0.00035
24	0.02329	48	0.06962	72	0.00897

STEP 3 The sample autocorrelation function corresponding to the above sample autocovariance function is easily computed, and is listed in computer output of Foothills Pool found in Appendix IV.

STEP 4 The sample partial autocorrelation function found by solving sequences of simultaneous equations is also listed in Appendix IV.

STEP 5 The sample autocorrelation function seems to cut off after r_1, since r_2, r_3, r_4, ... all tend to have values close to zero when compared to r_1. The value of r_1, though is also relatively close to zero itself, and it may be questioned as to how significantly it differs from zero. Hypothesis testing can be used to help determine the validity of these conjectures. First, consider the case that r_1 is an observation of ρ_1 which is actually zero.

STEP 5A The standard error of r_1 is given by

$$SE(r_1) = \frac{1}{\sqrt{72}} (1) = .118$$

STEP 5B The null hypothesis,

$$H_0: \ \rho_1, \rho_2, \rho_3, \dots = 0,$$

is tested against the alternative hypothesis,

$$H_1: \ \rho_1, \rho_2, \rho_3, \dots \neq 0 \ .$$

The value obtained for r_1 was -0.432 which lies outside the two standard error interval about zero ($\alpha = 0.05$),

$$-0.236 \leq r_k \leq 0.236. \qquad k = 1,2,\dots$$

Thus, the value of the observation, r_1, was either a very rare occurrance of r_1, or ρ_1 should not be considered as having a value equal to zero. Because -0.432 is so far outside the above interval, it appears almost certain that ρ_1 is significantly different from zero. The null hypothesis is rejected.

Next the case is considered for ρ_1 having a value significantly different from zero and $\rho_2, \rho_3, \rho_4, \ldots = 0$. The algorithm of STEP 5 is again used.

STEP 5A The standard error of r_2 is computed.

$$SE(r_2) = \frac{1}{\sqrt{72}}(1 + 2(-0.432)^2)^{1/2}$$

$$= 0.138.$$

STEP 5B The null hypothesis is

$H_0: \rho_2, \rho_3, \rho_4, \ldots = 0$

and the corresponding alternative hypothesis is

$H_1: \rho_2, \rho_3, \rho_4, \ldots \neq 0.$

Looking at the values for $r_2, r_3, r_4, \ldots$, all values lie within the two standard error interval, $-.276 \leq r_k \leq .276$, $k = 2,3,4,\ldots$. This indicates the null hypothesis cannot be rejected and the autocorrelation function may be considered as having cutoff after ρ_1. This conclusion would have been made even if one or two of the sample autocorrelations had values outside the two standard error interval. Such behavior is expected because the two standard error interval contains approximately 95% of the observations of the random variable, and inevitably a portion (about 5%) of the observations will lie outside the interval if α is assumed to be equal to 0.05.

STEP 6 The partial autocorrelation function must next be examined. The partial autocorrelations show no sudden cutoff characteristic as did the previously analyzed autocorrelation function. Instead, they seem to be dieing out to a value of zero.

STEP 7 This series then exhibits the first of the model characteristics listed in STEP 7 of Appendix I. Its autocorrelation function cuts off after r_1 while its partial autocorrelation function dies out, indicating a pure moving average model of order 1 (MA(1)). The form of the proposed model is

$$W_t = a_t - \theta_1 a_{t-1} + \theta_o$$

STEP 8 To determine the estimates of the parameters, θ_1, and θ_o, for the pure moving average model, the moving average iterative scheme must be used.

Iteration 1 STEP 8A. Set $\hat{\theta} = 1$.

STEP 8B. $\hat{\sigma}_a^2 = 0.0015116$

STEP 8C. $\hat{\theta}_1 = \dfrac{6.53 \times 10^{-4}}{1.5116 \times 10^{-3}} = -0.432$

STEP 8D. Continue iterating until the change in $\hat{\theta}_1$ is less than an arbitrarily specified tolerance of 0.01.

Iteration 2 STEP 8B. $\hat{\sigma}_a^2 = \dfrac{1.5116 \times 10^{-3}}{1 + (-0.432)^2} = 1.274 \times 10^{-3}$

STEP 8C. $\hat{\theta}_1 = \dfrac{6.53 \times 10^{-4}}{1.274 \times 10^{-3}} = -0.513$

STEP 8D. The change in the value for $\hat{\theta}_1$ is larger than the 0.01 tolerance specified so the algorithm must be continued.

Iterations 3 through 5 yield the following results

Iteration 3 $\hat{\sigma}_a^2 = 1.197 \times 10^{-3}$; $\hat{\theta}_1 = 0.546$

Iteration 4 $\hat{\sigma}_a^2 = 1.165 \times 10^{-3}$; $\hat{\theta}_1 = 0.561$

Iteration 5 $\hat{\sigma}_a^2 = 1.150 \times 10^{-3}$; $\hat{\theta}_1 = 0.568$

The value of $\hat{\theta}_1$ in iteration 5 satisfies the tolerance specified, and thus is considered as the preliminary estimate of θ_1.

STEP 8E The value for the constant term, θ_0, has been computed in STEP 1,

$\hat{\theta}_0 = \overline{W} = -0.009262.$

STEP 9 To check if the value for θ_0 is significantly different from zero, the variance of $\overline{W}$ is computed.

$$\text{Var}(\overline{W}) = \frac{1.533 \times 10^{-3}}{72} \quad (1 + 2(-0.491))$$

$$= 3.833 \times 10^{-7}$$

The corresponding two standard error interval is then

$$-1.238 \times 10^{-3} \leq \overline{W} \leq 1.238 \times 10^{-3}$$

The value of $\overline{W}$ obtained above lies outside this interval, and as a result, θ_0 must be included in the model.

STEP 10 The model used to forecast future observations is then

$$W_t = a_t - 0.568\ a_{t-1} - 0.009262\ .$$

Again, it must be emphasized that the parameter values obtained using the outline procedure in Appendix I are only preliminary estimates of the parameter values. Refinement techniques were used to obtain better parameters values for the final forecast function used in forecasting future oil production. The above preliminary model compares to the actual model

$$W_t = a_t - .573\ a_{t-1} - .009262$$

used in forecasting future oil production.

In this case, the preliminary estimate of θ_i is very close to the actual value of the parameter used. It should be pointed out that the preliminary parameter estimate often is not as close to the final parameter estimate as it was in this case.

Forecasting with the Model Transformed observations 1 through 72 of Table AI have been used to obtain the parameter values for the forecasting function. It is now desired to forecast future observations from the function containing the final parameter estimates,

$$W_t = a_t - 0.573a_{t-1} - 0.009262 \quad .$$

At this point, it is best to partially retransform our series. The stationary series had been formed by taking differences of the natural logarithms of the raw data. That is $W_t = X_t - X_{t-1}$ where X_t and X_{t-1} are natural logs of the raw data. The forecasting function can then be rewritten

$$W_t = X_t - X_{t-1} = a_t - 0.573a_{t-1} - 0.009262$$

or

$$X_t = X_{t-1} + a_t - 0.573\ a_{t-1} - 0.009262 \quad .$$

Then if it is desired to predict the natural logarithm of observation 74, the forecasting function becomes

$$X_{74} = X_{73} + a_{74} - 0.573\ a_{73} - 0.009262 \quad .$$

Values for a_{74}, a_{73}, and X_{73} are necessary to determine a value for X_{74}. The value for a_{74} cannot be determined at this point because a_{74} is actually a future disturbance. That is a_{74} is defined as

$$a_{74} = X_{74(\text{actual})} - X_{74(\text{predicted})} .$$

Then obviously X_{74} must be realized before a_{74} can be determined, but X_{74} is the observation that is presently being predicted. To alleviate this predicament, the disturbance, a_{74}, is set at its expected value, zero. It is possible, though, to determine a value for a_{73}. The disturbance a_{73} is of course defined as

$$a_{73} = X_{73(actual)} - X_{73(predicted)}.$$

The value for $X_{73(actual)}$ has been observed to be 5.6733 (that is ln(291.0)). The problem is to determine a value for $W_{72(predicted)}$. At this point, it is necessary to build the listing of past disturbances which was mentioned in Appendix I.

This listing is computed by compiling a series of one period ahead forecasts beginning at some selected place in the observed data. For this example arbitrarily start at t = 63. At this point the raw data point was 310 making X_{63} = 5.73657. Then the forecasting function reads

$$X_{64(predicted)} = X_{63} - 0.573\ a_{63} - 0.009262\ .$$

Again a value for a_{63} is needed. This problem is solved by setting it equal to zero. This initial zero creates a transient effect in the predicted values which dies out as the listing process continues and hopefully will have died out completely by the time X_{74} is predicted.

Thus $X_{64(predicted)} = 5.73657 - 0.009262 = 5.7273$.

Next $X_{65(predicted)}$ is to be computed by

$$X_{65(predicted)} = X_{64} - 0.573\ a_{64} - 0.009362.$$

This time a value for a_{64} can be easily obtained because the past prediction error is available.

$$a_{64} = X_{64(actual)} - X_{64(predicted)}$$

$$= 5.68698 - 5.7273$$

$$= -0.04032$$

Continuing this process gives the following table of results:

t	X_t(predicted) $=X_{t-1}$(actual)$-0.573a_{t-1}-0.009262$	Z_t(actual)	X_t(actual) $=\ln(Z_t$(actual))	a_t $=X_t$(actual)$-X_t$(predicted)
63	- - -	310.0	5.73657	0
64	5.7273	295.0	5.68698	-0.04032
65	5.7008	293.1	5.68051	-0.02031
66	5.6829	287.2	5.66018	-0.02271
67	5.6639	289.3	5.66746	0.00353
68	5.6562	290.6	5.67195	0.01577
69	5.6537	289.0	5.66643	0.01278
70	5.6498	285.1	5.65284	0.00300
71	5.6419	288.5	5.66470	0.02284
72	5.6423	288.4	5.66435	0.02200
73	5.6425	291.0	5.67332	0.03084

Therefore, this procedure results in a forecast value of X_{74} of 5.6464.

Computation of X_{75}, X_{76}, X_{77} ... is straightforward once it is realized that the term, 0.573 a_{t-1}, becomes zero for predictions of two or more periods in the future. Again assume the last observed value occurred at t = 73, and it is desired to predict the observation when t = 75. Then our function gives

$$X_{75} = X_{74} - 0.0573\ a_{74} - 0.009361938.$$

But a_{74} as previously discussed is a future disturbance which is set at its expected value of zero. This result extends to all other future forecasts assuming we know only observations 1 through 73. The first five predictions for the natural logarithm of the average daily well production are listed below.

$X_{74} \doteq 5.6464$

$X_{75} = 5.6371$

$X_{76} = 5.6279$

$X_{77} = 5.6186$

$X_{78} = 5.6094$

Finally, it is desired to obtain the raw data predictions, Z_t, from the natural logarithm of the data predictions, X_t. It is very tempting at this point to suggest that the raw data prediction is the antilog of the natural logarithm of the data prediction. Such is not the case, as the exponential of an expected value does not equal the expected value of the exponential of a value. That is

$$e^{E(Z_t)} \neq E(e^{Z_t})$$

where E( ) indicates expected value (Nelson, 1973, p 162).

The raw data predictions can be found, though, using the relation

$$Z_{T+\ell} = e^{\left[X_{T+\ell} + \frac{1}{2}\sigma_a^2(1 + (\ell-1)(1-\theta_1)^2)\right]}$$

for all MA(1) models of the form used in this study. A value of σ_a^2 was found in STEP 8 as 1.150×10^{-3}. This value was improved on by a more advanced technique to yield a revised value of 1.119×10^{-3}. The raw data prediction can be found then by using the forecasting function

$$Z_{73+\ell} = e^{\left[X_{73+\ell} + \frac{1}{2}(1.119\times10^{-3})(1+(\ell-1)(.427)^2\right]}$$

The five raw data predictions are then

$Z_{74} = 283.43$

$Z_{75} = 280.83$

$Z_{76} = 278.29$

$Z_{77} = 275.74$

$Z_{78} = 273.24$

The corresponding variances for these predictions are given by

$$\text{Var}(Z_{T+\ell}) = \sigma_a^2(1 + (\ell -1)(1 - \theta_1)^2) .$$

These variances can easily be used to determine corresponding standard errors for the population which, of course, can be used to construct probability limits for the above predictions.

Appendix IV

Computer Output

For each of the four reservoirs, the computer output used to conduct the time series analysis is included. The information included as output consists of:

1) the data file of average daily productions by month

2) the mean of the stationary process ($\overline{W}$) , the autocorrelation function, and the partial autocorrelation function

3) the values for the sum of the squares function used to initially approximate the value of θ_1

4) the resulting values for the parameter estimates obtained from the final iterative algorithm including the standard deviation of $\overline{W}$, the variance of white noise (σ_a^2), the final value of θ_1, and the variances of the distribution of θ_1

5) the natural log forecasts with their associated probability limits, and the actual forecasts with their associated probability limits.

Computer Output For Foothills Pool, North Dakota

```
0073 566.9 535.2 552.6 541.1 534.5 521.1 516.9 516.9 515.1
502.7 471.3 467.9 458.0 479.2 462.0 460.8 467.9 443.3
447.7 426.6 416.8 427.3 421.9 403.1 412.6 408.8 393.8
390.9 426.8 395.9 385.0 407.1 387.3 375.8 371.2 364.0
366.9 347.1 367.5 360.7 366.3 350.8 366.6 362.2 367.9
364.6 354.4 342.6 367.3 363.2 340.2 330.1 339.8 337.9
346.3 298.5 319.0 299.6 294.8 294.4 284.5 283.0 310.0
295.0 293.1 287.2 289.3 290.6 289.0 285.1 288.5 288.4
291.0
```

```
THE MEAN IS:-0.9261938E-02
THE AUTOCORRELATION FUNCTION

-0.432-0.034 0.016 0.096-0.051 0.096-0.214 0.126-0.088 0.199
-0.130-0.045-0.100 0.243-0.154 0.023-0.106 0.192-0.222 0.171
-0.080 0.025-0.001 0.038-0.013 0.096-0.166 0.114-0.034-0.080
 0.083 0.029-0.095 0.056-0.020 0.080-0.109 0.088-0.129 0.093
 0.075-0.118 0.014-0.032 0.022 0.088-0.054-0.036 0.059 0.018
-0.017-0.029-0.077 0.113-0.040 0.008-0.014-0.014 0.007 0.040
THE STANDARD ERROR

 0.118 0.138 0.138 0.138 0.139 0.139 0.140 0.145 0.146 0.147
 0.151 0.152 0.152 0.153 0.159 0.161 0.161 0.162 0.165 0.169
 0.171 0.172 0.172 0.172 0.172 0.172 0.173 0.175 0.176 0.176
 0.177 0.177 0.177 0.178 0.178 0.178 0.179 0.180 0.180 0.182
 0.182 0.183 0.184 0.184 0.184 0.184 0.184 0.185 0.185 0.185
 0.185 0.185 0.185 0.186 0.186 0.187 0.187 0.187 0.187 0.187
THE PARTIAL AUTOCORRELATION FUNCTION
-0.432-0.272-0.160 0.033 0.028 0.157-0.131-0.047-0.151 0.142
 0.081-0.031-0.218 0.006-0.035-0.003-0.070 0.094-0.186-0.056
 0.010 0.059 0.113-0.048 0.049 0.095 0.007-0.050 0.021-0.175
-0.015 0.003 0.093-0.018-0.030 0.021 0.001 0.125-0.109-0.013
 0.120-0.032-0.009-0.085-0.056-0.030 0.076 0.100 0.042 0.024
-0.080-0.041-0.008 0.057-0.010-0.019-0.054-0.043-0.021 0.058
-0.099-0.046
```

Computer Output For Foothills Pool, North Dakota

```
THETA =-0.1000000E+01SUM OF SQUARES = 0.3449198E+02
THETA =-0.9000000E+00SUM OF SQUARES = 0.1635809E+01
THETA =-0.8000000E+00SUM OF SQUARES = 0.6109153E+00
THETA =-0.7000000E+00SUM OF SQUARES = 0.3611576E+00
THETA =-0.6000000E+00SUM OF SQUARES = 0.2600358E+00
THETA =-0.5000000E+00SUM OF SQUARES = 0.2062966E+00
THETA =-0.4000000E+00SUM OF SQUARES = 0.1727657E+00
THETA =-0.3000000E+00SUM OF SQUARES = 0.1496123E+00
THETA =-0.2000000E+00SUM OF SQUARES = 0.1325223E+00
THETA =-0.1000000E+00SUM OF SQUARES = 0.1193288E+00
THETA =-0.1862645E-08SUM OF SQUARES = 0.1088334E+00
THETA = 0.1000000E+00SUM OF SQUARES = 0.1003352E+00
THETA = 0.2000000E+00SUM OF SQUARES = 0.9342906E-01
THETA = 0.3000000E+00SUM OF SQUARES = 0.8791824E-01
THETA = 0.4000000E+00SUM OF SQUARES = 0.8378758E-01
THETA = 0.5000000E+00SUM OF SQUARES = 0.8122242E-01
THETA = 0.6000000E+00SUM OF SQUARES = 0.8069327E-01
THETA = 0.7000000E+00SUM OF SQUARES = 0.8314878E-01
THETA = 0.8000000E+00SUM OF SQUARES = 0.9030676E-01
THETA = 0.9000000E+00SUM OF SQUARES = 0.1046728E+00
THETA = 0.1000000E+01SUM OF SQUARES = 0.4403537E+00
```

```
THE NO. OF ITERATIONS   3
THE STANDARD DEVIATION OF W 0.1672425E-02
THE VARIANCE OF WHITE NOISE 0.1119346E-02
THE THETAS
 0.5729037E+00
THE COVARIANCE MATRIX
 0.9623569E-02
```

Computer Output For Foothills Pool, North Dakota

```
THETA SUBZERO
-0.9261938E-02
THE LOG FORCASTS
THE FORECASTS ARE
 0.5646192E+01 0.5636930E+01 0.5627669E+01 0.5618407E+01 0.5609145E+01
 0.5599883E+01 0.5590621E+01 0.5581359E+01 0.5572097E+01 0.5562835E+01
 0.5553573E+01 0.5544311E+01 0.5535049E+01 0.5525787E+01 0.5516525E+01
 0.5507263E+01 0.5498001E+01 0.5488739E+01 0.5479477E+01 0.5470215E+01
 0.5460953E+01 0.5451691E+01 0.5442429E+01 0.5433167E+01
THE LOWER PROBABILITY LIMITS
 0.5580617E+01 0.5565625E+01 0.5551060E+01 0.5536839E+01 0.5522903E+01
 0.5509207E+01 0.5495719E+01 0.5482410E+01 0.5469261E+01 0.5456254E+01
 0.5443373E+01 0.5430608E+01 0.5417948E+01 0.5405383E+01 0.5392907E+01
 0.5380512E+01 0.5368193E+01 0.5355944E+01 0.5343761E+01 0.5331639E+01
 0.5319575E+01 0.5307566E+01 0.5295608E+01 0.5283699E+01
THE UPPER PROBABILITY LIMITS
 0.5711767E+01 0.5708236E+01 0.5704277E+01 0.5699974E+01 0.5695386E+01
 0.5690558E+01 0.5685523E+01 0.5680307E+01 0.5674932E+01 0.5669416E+01
 0.5663772E+01 0.5658013E+01 0.5652150E+01 0.5646191E+01 0.5640143E+01
 0.5634014E+01 0.5627809E+01 0.5621534E+01 0.5615194E+01 0.5608791E+01
 0.5602331E+01 0.5595816E+01 0.5589251E+01 0.5582636E+01
THE FORECASTS ARE
 0.2833696E+03 0.2807858E+03 0.2782256E+03 0.2756887E+03 0.2731750E+03
 0.2706842E+03 0.2682161E+03 0.2657704E+03 0.2633471E+03 0.2609459E+03
 0.2585666E+03 0.2562090E+03 0.2538729E+03 0.2515580E+03 0.2492643E+03
 0.2469915E+03 0.2447394E+03 0.2425079E+03 0.2402967E+03 0.2381056E+03
 0.2359346E+03 0.2337833E+03 0.2316517E+03 0.2295395E+03
THE LOWER PROBABILITY LIMITS
 0.2652353E+03 0.2612885E+03 0.2575104E+03 0.2538743E+03 0.2503608E+03
 0.2469553E+03 0.2436466E+03 0.2404255E+03 0.2372848E+03 0.2342183E+03
 0.2312209E+03 0.2282880E+03 0.2254160E+03 0.2226015E+03 0.2198415E+03
 0.2171334E+03 0.2144749E+03 0.2118638E+03 0.2092983E+03 0.2067766E+03
 0.2042971E+03 0.2018583E+03 0.1994588E+03 0.1970975E+03
THE UPPER PROBABILITY LIMITS
 0.3024051E+03 0.3013390E+03 0.3001484E+03 0.2988596E+03 0.2974916E+03
 0.2960587E+03 0.2945717E+03 0.2930394E+03 0.2914686E+03 0.2898652E+03
 0.2882339E+03 0.2865788E+03 0.2849033E+03 0.2832105E+03 0.2815030E+03
 0.2797829E+03 0.2780523E+03 0.2763130E+03 0.2745665E+03 0.2728143E+03
 0.2710575E+03 0.2692974E+03 0.2675350E+03 0.2657712E+03
```

Computer Output For Beaver Lodge Madison Pool, North Dakota

```
0064 5206. 5092. 4834. 4567. 4709. 4579. 4547. 4520. 4421.
4435. 4201. 4142. 4254. 4297. 4295. 4388. 4189. 4218.
4278. 4102. 4010. 3869. 3792. 3651. 3592. 3613. 3494.
3337. 3356. 3174. 3197. 3220. 3033. 2979. 3030. 2975.
2954. 2839. 2967. 2760. 2718. 2580. 2639. 2706. 2598.
2625. 2619. 2519. 2379. 2370. 2275. 2328. 2189. 2272.
2271. 2140. 2169. 2128. 2022. 2058. 2043. 2137. 2041.
1943.
```

```
THE MEAN IS:-0.1564410E-01
THE AUTOCORRELATION FUNCTION

-0.310-0.035 0.051-0.111 0.040-0.110 0.006 0.240-0.167-0.035
 0.070-0.192 0.090-0.151-0.020 0.094-0.099 0.128 0.047-0.115
 0.077-0.123 0.173-0.131 0.064 0.070-0.054-0.032 0.052 0.002
 0.143-0.136-0.061 0.073-0.174 0.132-0.066 0.043 0.102-0.143
 0.019 0.073-0.098 0.034 0.019-0.009 0.126-0.043 0.038-0.014
-0.155 0.079 0.022-0.043 0.057-0.018 0.027-0.070-0.045 0.036
THE STANDARD ERROR

 0.126 0.138 0.138 0.138 0.139 0.140 0.141 0.141 0.147 0.150
 0.150 0.151 0.155 0.156 0.158 0.158 0.159 0.160 0.161 0.162
 0.163 0.164 0.165 0.168 0.169 0.170 0.170 0.171 0.171 0.171
 0.171 0.173 0.174 0.175 0.175 0.178 0.180 0.180 0.180 0.181
 0.183 0.183 0.183 0.184 0.184 0.184 0.184 0.186 0.186 0.186
 0.186 0.188 0.188 0.189 0.189 0.189 0.189 0.189 0.189 0.190
THE PARTIAL AUTOCORRELATION FUNCTION
-0.310-0.145-0.007-0.113-0.033-0.144-0.087 0.211-0.021-0.100
 0.001-0.178-0.040-0.166-0.187-0.123-0.121 0.013 0.048-0.057
-0.042-0.119 0.137-0.167-0.004-0.091-0.116-0.102-0.054 0.002
 0.153 0.006-0.080-0.082-0.151-0.059-0.121-0.102-0.044-0.065
-0.065 0.028 0.039-0.111 0.035-0.092 0.010-0.010-0.048-0.096
-0.133-0.077 0.073-0.037-0.031-0.051-0.013-0.027-0.044-0.074
-0.101 0.015
```

Computer Output For Beaver Lodge Madison Pool, North Dakota

```
THETA =-0.1000000E+01SUM OF SQUARES = 0.8166092E+00
THETA =-0.9000000E+00SUM OF SQUARES = 0.3075592E+00
THETA =-0.8000000E+00SUM OF SQUARES = 0.1988935E+00
THETA =-0.7000000E+00SUM OF SQUARES = 0.1450047E+00
THETA =-0.6000000E+00SUM OF SQUARES = 0.1151878E+00
THETA =-0.5000000E+00SUM OF SQUARES = 0.9637670E-01
THETA =-0.4000000E+00SUM OF SQUARES = 0.8338511E-01
THETA =-0.3000000E+00SUM OF SQUARES = 0.7391257E-01
THETA =-0.2000000E+00SUM OF SQUARES = 0.6679891E-01
THETA =-0.1000000E+00SUM OF SQUARES = 0.6138797E-01
THETA =-0.1862645E-08SUM OF SQUARES = 0.5727826E-01
THETA = 0.1000000E+00SUM OF SQUARES = 0.5421804E-01
THETA = 0.2000000E+00SUM OF SQUARES = 0.5205675E-01
THETA = 0.3000000E+00SUM OF SQUARES = 0.5072243E-01
THETA = 0.4000000E+00SUM OF SQUARES = 0.5021911E-01
THETA = 0.5000000E+00SUM OF SQUARES = 0.5065349E-01
THETA = 0.6000000E+00SUM OF SQUARES = 0.5231139E-01
THETA = 0.7000000E+00SUM OF SQUARES = 0.5580287E-01
THETA = 0.8000000E+00SUM OF SQUARES = 0.6222240E-01
THETA = 0.9000000E+00SUM OF SQUARES = 0.7237898E-01
THETA = 0.1000000E+01SUM OF SQUARES = 0.7857325E-01
```

```
THE NO. OF ITERATIONS  3
THE STANDARD DEVIATION OF W 0.2094301E-02
THE VARIANCE OF WHITE NOISE 0.7970942E-03
THE THETAS
 0.4065636E+00
THE COVARIANCE MATRIX
 0.1286541E-01
```

Computer Output For Beaver Lodge Madison Pool, North Dakota

```
THETA SUBZERO
-0.1564410E-01
THE LOG FORCASTS
THE FORECASTS ARE
 0.7570484E+01 0.7554840E+01 0.7539196E+01 0.7523551E+01 0.7507907E+01
 0.7492263E+01 0.7476619E+01 0.7460975E+01 0.7445331E+01 0.7429687E+01
 0.7414043E+01 0.7398399E+01 0.7382755E+01 0.7367111E+01 0.7351467E+01
 0.7335823E+01 0.7320179E+01 0.7304534E+01 0.7288890E+01 0.7273246E+01
 0.7257602E+01 0.7241958E+01 0.7226314E+01 0.7210670E+01
THE LOWER PROBABILITY LIMITS
 0.7515147E+01 0.7490493E+01 0.7466954E+01 0.7444196E+01 0.7422026E+01
 0.7400318E+01 0.7378985E+01 0.7357967E+01 0.7337215E+01 0.7316694E+01
 0.7296374E+01 0.7276234E+01 0.7256253E+01 0.7236416E+01 0.7216710E+01
 0.7197122E+01 0.7177644E+01 0.7158266E+01 0.7138981E+01 0.7119782E+01
 0.7100664E+01 0.7081621E+01 0.7062649E+01 0.7043743E+01
THE UPPER PROBABILITY LIMITS
 0.7625820E+01 0.7619186E+01 0.7611437E+01 0.7602907E+01 0.7593789E+01
 0.7584209E+01 0.7574253E+01 0.7563984E+01 0.7553447E+01 0.7542680E+01
 0.7531711E+01 0.7520564E+01 0.7509256E+01 0.7497805E+01 0.7486223E+01
 0.7474523E+01 0.7462713E+01 0.7450803E+01 0.7438800E+01 0.7426710E+01
 0.7414541E+01 0.7402295E+01 0.7389980E+01 0.7377597E+01
THE FORECASTS ARE
 0.1940852E+04 0.1910993E+04 0.1881594E+04 0.1852648E+04 0.1824146E+04
 0.1796083E+04 0.1768452E+04 0.1741246E+04 0.1714458E+04 0.1688083E+04
 0.1662113E+04 0.1636543E+04 0.1611366E+04 0.1586577E+04 0.1562168E+04
 0.1538136E+04 0.1514473E+04 0.1491174E+04 0.1468234E+04 0.1445646E+04
 0.1423406E+04 0.1401508E+04 0.1379947E+04 0.1358718E+04
THE LOWER PROBABILITY LIMITS
 0.1835638E+04 0.1790935E+04 0.1749270E+04 0.1709910E+04 0.1672418E+04
 0.1636504E+04 0.1601964E+04 0.1568644E+04 0.1536427E+04 0.1505219E+04
 0.1474943E+04 0.1445534E+04 0.1416938E+04 0.1389107E+04 0.1362001E+04
 0.1335582E+04 0.1309818E+04 0.1284681E+04 0.1260143E+04 0.1236181E+04
 0.1212772E+04 0.1189896E+04 0.1167533E+04 0.1145667E+04
THE UPPER PROBABILITY LIMITS
 0.2050461E+04 0.2036904E+04 0.2021181E+04 0.2004012E+04 0.1985823E+04
 0.1966890E+04 0.1947405E+04 0.1927509E+04 0.1907307E+04 0.1886881E+04
 0.1866297E+04 0.1845607E+04 0.1824856E+04 0.1804078E+04 0.1783305E+04
 0.1762560E+04 0.1741868E+04 0.1721245E+04 0.1700708E+04 0.1680271E+04
 0.1659946E+04 0.1639744E+04 0.1619673E+04 0.1599742E+04
```

Computer Output For North Black Slough Pool, North Dakota

```
0051 605.3 604.3 561.7 558.9 568.2 613.6 549.0 537.1 582.1
490.6 542.7 531.6 530.7 427.0 439.1 445.5 384.0 431.3
384.0 361.0 371.2 313.9 379.4 366.3 351.3 338.3 357.3
309.5 299.4 281.9 304.5 282.4 247.5 252.3 253.5 247.6
257.8 273.5 269.9 261.2 245.6 251.5 244.2 236.3 223.7
234.6 209.4 208.2 191.6 204.3 207.8
```

```
THE MEAN IS:-0.2138296E-01
THE AUTOCORRELATION FUNCTION

-0.469 0.026 0.059-0.048 0.022-0.122 0.146 0.009-0.144-0.108
 0.027 0.172-0.151 0.128 0.057 0.017-0.135-0.008 0.201-0.069
-0.071-0.023 0.054-0.200 0.129-0.056 0.116-0.146 0.033 0.023
THE STANDARD ERROR

 0.141 0.170 0.170 0.170 0.170 0.170 0.172 0.175 0.175 0.177
 0.178 0.178 0.182 0.184 0.186 0.186 0.186 0.188 0.188 0.193
 0.193 0.194 0.194 0.194 0.198 0.200 0.200 0.201 0.203 0.204
THE PARTIAL AUTOCORRELATION FUNCTION
-0.469-0.248-0.058-0.045-0.012-0.168 0.006 0.101-0.081-0.337
-0.354 0.018 0.031 0.090 0.105 0.223 0.075-0.131-0.053 0.035
 0.015-0.086 0.011-0.134 0.138-0.050-0.067-0.286-0.098-0.097
```

Computer Output For North Black Slough Pool, North Dakota

```
THETA =-0.1000000E+01SUM OF SQUARES = 0.5038313E+01
THETA =-0.9000000E+00SUM OF SQUARES = 0.2404290E+01
THETA =-0.8000000E+00SUM OF SQUARES = 0.1367558E+01
THETA =-0.7000000E+00SUM OF SQUARES = 0.9688351E+00
THETA =-0.6000000E+00SUM OF SQUARES = 0.7543426E+00
THETA =-0.5000000E+00SUM OF SQUARES = 0.6165582E+00
THETA =-0.4000000E+00SUM OF SQUARES = 0.5199093E+00
THETA =-0.3000000E+00SUM OF SQUARES = 0.4487054E+00
THETA =-0.2000000E+00SUM OF SQUARES = 0.3945000E+00
THETA =-0.1000000E+00SUM OF SQUARES = 0.3522366E+00
THETA =-0.1862645E-08SUM OF SQUARES = 0.3187350E+00
THETA = 0.1000000E+00SUM OF SQUARES = 0.2919578E+00
THETA = 0.2000000E+00SUM OF SQUARES = 0.2706063E+00
THETA = 0.3000000E+00SUM OF SQUARES = 0.2538941E+00
THETA = 0.4000000E+00SUM OF SQUARES = 0.2414476E+00
THETA = 0.5000000E+00SUM OF SQUARES = 0.2332984E+00
THETA = 0.6000000E+00SUM OF SQUARES = 0.2299045E+00
THETA = 0.7000000E+00SUM OF SQUARES = 0.2321414E+00
THETA = 0.8000000E+00SUM OF SQUARES = 0.2418425E+00
THETA = 0.9000000E+00SUM OF SQUARES = 0.2671781E+00
THETA = 0.1000000E+01SUM OF SQUARES = 0.3680440E+00
```

```
THE NO. OF ITERATIONS  3
THE STANDARD DEVIATION OF W 0.3674533E-02
THE VARIANCE OF WHITE NOISE 0.4597063E-02
THE THETAS
 0.6129678E+00
THE COVARIANCE MATRIX
 0.1249582E-01
```

Computer Output For North Black Slough Pool, North Dakota

```
THETA SUBZERO
-0.2138296E-01
THE LOG FORCASTS
THE FORECASTS ARE
 0.5276997E+01 0.5255614E+01 0.5234231E+01 0.5212848E+01 0.5191465E+01
 0.5170082E+01 0.5148699E+01 0.5127316E+01 0.5105933E+01 0.5084550E+01
 0.5063167E+01 0.5041784E+01 0.5020401E+01 0.4999018E+01 0.4977635E+01
 0.4956252E+01 0.4934869E+01 0.4913486E+01
THE LOWER PROBABILITY LIMITS
 0.5144106E+01 0.5113129E+01 0.5082758E+01 0.5052891E+01 0.5023452E+01
 0.4994383E+01 0.4965635E+01 0.4937173E+01 0.4908965E+01 0.4880986E+01
 0.4853214E+01 0.4825630E+01 0.4798220E+01 0.4770969E+01 0.4743865E+01
 0.4716897E+01 0.4690058E+01 0.4663337E+01
THE UPPER PROBABILITY LIMITS
 0.5409888E+01 0.5398100E+01 0.5385704E+01 0.5372805E+01 0.5359478E+01
 0.5345782E+01 0.5331763E+01 0.5317459E+01 0.5302901E+01 0.5288114E+01
 0.5273121E+01 0.5257938E+01 0.5242583E+01 0.5227068E+01 0.5211406E+01
 0.5195607E+01 0.5179681E+01 0.5163636E+01
THE FORECASTS ARE
 0.1962316E+03 0.1921462E+03 0.1881458E+03 0.1842287E+03 0.1803932E+03
 0.1766376E+03 0.1729601E+03 0.1693592E+03 0.1658332E+03 0.1623807E+03
 0.1590000E+03 0.1556898E+03 0.1524484E+03 0.1492745E+03 0.1461667E+03
 0.1431236E+03 0.1401439E+03 0.1372262E+03
THE LOWER PROBABILITY LIMITS
 0.1714181E+03 0.1661895E+03 0.1612180E+03 0.1564742E+03 0.1519349E+03
 0.1475818E+03 0.1433996E+03 0.1393757E+03 0.1354991E+03 0.1317605E+03
 0.1281516E+03 0.1246650E+03 0.1212943E+03 0.1180335E+03 0.1148773E+03
 0.1118208E+03 0.1088594E+03 0.1059892E+03
THE UPPER PROBABILITY LIMITS
 0.2236066E+03 0.2209860E+03 0.2182638E+03 0.2154664E+03 0.2126139E+03
 0.2097217E+03 0.2068022E+03 0.2038652E+03 0.2009189E+03 0.1979698E+03
 0.1950236E+03 0.1920850E+03 0.1891580E+03 0.1862459E+03 0.1833516E+03
 0.1804776E+03 0.1776261E+03 0.1747988E+03
```

Computer Output For Haas Madison Pool, North Dakota

```
0055 627.3 580.4 584.7 598.4 582.6 557.5 568.4 563.7 560.0
559.3 503.6 537.3 536.3 533.3 523.5 518.5 496.0 503.9
504.6 494.3 488.4 479.5 480.9 496.3 521.4 508.3 491.5
495.8 488.5 478.3 482.7 464.6 465.9 460.1 481.0 449.3
464.4 461.6 460.7 480.3 472.6 449.6 456.6 443.0 444.4
431.0 433.9 424.2 434.0 429.2 433.3 428.0 429.2 418.0
413.5
```

```
THE MEAN IS:-0.7717914E-02
THE AUTOCORRELATION FUNCTION

-0.360-0.035-0.043 0.021-0.043 0.232-0.251-0.120 0.163 0.043
-0.078 0.206-0.123-0.091 0.114 0.156-0.185-0.001-0.010-0.066
 0.145-0.063-0.124-0.086 0.283-0.242 0.138 0.102-0.225 0.009
THE STANDARD ERROR

 0.136 0.153 0.153 0.153 0.153 0.153 0.160 0.167 0.168 0.171
 0.172 0.172 0.177 0.178 0.179 0.181 0.183 0.186 0.186 0.186
 0.187 0.189 0.189 0.191 0.192 0.199 0.205 0.206 0.207 0.212
THE PARTIAL AUTOCORRELATION FUNCTION
-0.360-0.190-0.151-0.075-0.096 0.210-0.104-0.270-0.014 0.056
-0.051 0.195 0.145-0.060-0.058 0.240 0.057-0.147 0.046 0.031
-0.049-0.162-0.010-0.257-0.004-0.164-0.049 0.159-0.126-0.142
```

Computer Output For Haas Madison Pool, North Dakota

```
THETA =-0.1000000E+01SUM OF SQUARES = 0.4514618E+01
THETA =-0.9000000E+00SUM OF SQUARES = 0.4864428E+00
THETA =-0.8000000E+00SUM OF SQUARES = 0.2394247E+00
THETA =-0.7000000E+00SUM OF SQUARES = 0.1510611E+00
THETA =-0.6000000E+00SUM OF SQUARES = 0.1109430E+00
THETA =-0.5000000E+00SUM OF SQUARES = 0.8885375E-01
THETA =-0.4000000E+00SUM OF SQUARES = 0.7505016E-01
THETA =-0.3000000E+00SUM OF SQUARES = 0.6564397E-01
THETA =-0.2000000E+00SUM OF SQUARES = 0.5882417E-01
THETA =-0.1000000E+00SUM OF SQUARES = 0.5365785E-01
THETA =-0.1862645E-08SUM OF SQUARES = 0.4962830E-01
THETA = 0.1000000E+00SUM OF SQUARES = 0.4643247E-01
THETA = 0.2000000E+00SUM OF SQUARES = 0.4388946E-01
THETA = 0.3000000E+00SUM OF SQUARES = 0.4189552E-01
THETA = 0.4000000E+00SUM OF SQUARES = 0.4041985E-01
THETA = 0.5000000E+00SUM OF SQUARES = 0.3949396E-01
THETA = 0.6000000E+00SUM OF SQUARES = 0.3923431E-01
THETA = 0.7000000E+00SUM OF SQUARES = 0.3993803E-01
THETA = 0.8000000E+00SUM OF SQUARES = 0.4258415E-01
THETA = 0.9000000E+00SUM OF SQUARES = 0.5180506E-01
THETA = 0.1000000E+01SUM OF SQUARES = 0.1655242E+00
```

```
THE NO. OF ITERATIONS  4
THE STANDARD DEVIATION OF W 0.1510185E-02
THE VARIANCE OF WHITE NOISE 0.7263430E-03
THE THETAS
 0.5844333E+00
THE COVARIANCE MATRIX
 0.1145807E-01
```

Computer Output For Haas Madison Pool, North Dakota

```
THETA SUBZERO
-0.7717914E-02
THE LOG FORCASTS
THE FORECASTS ARE
 0.6022028E+01 0.6014310E+01 0.6006592E+01 0.5998874E+01 0.5991156E+01
 0.5983438E+01 0.5975720E+01 0.5968003E+01 0.5960285E+01 0.5952567E+01
 0.5944849E+01 0.5937131E+01 0.5929413E+01 0.5921695E+01 0.5913977E+01
 0.5906259E+01 0.5898541E+01 0.5890823E+01 0.5883106E+01 0.5875388E+01
 0.5867670E+01 0.5859952E+01 0.5852234E+01 0.5844516E+01
THE LOWER PROBABILITY LIMITS
 0.5969204E+01 0.5957098E+01 0.5945305E+01 0.5933767E+01 0.5922441E+01
 0.5911295E+01 0.5900304E+01 0.5889450E+01 0.5878716E+01 0.5868090E+01
 0.5857561E+01 0.5847120E+01 0.5836758E+01 0.5826470E+01 0.5816250E+01
 0.5806092E+01 0.5795992E+01 0.5785946E+01 0.5775951E+01 0.5766003E+01
 0.5756099E+01 0.5746238E+01 0.5736416E+01 0.5726632E+01
THE UPPER PROBABILITY LIMITS
 0.6074851E+01 0.6071522E+01 0.6067879E+01 0.6063982E+01 0.6059872E+01
 0.6055582E+01 0.6051137E+01 0.6046555E+01 0.6041853E+01 0.6037043E+01
 0.6032137E+01 0.6027142E+01 0.6022068E+01 0.6016920E+01 0.6011705E+01
 0.6006427E+01 0.6001091E+01 0.5995701E+01 0.5990261E+01 0.5984773E+01
 0.5979240E+01 0.5973666E+01 0.5968052E+01 0.5962400E+01
THE FORECASTS ARE
 0.4125639E+03 0.4094177E+03 0.4062956E+03 0.4031973E+03 0.4001225E+03
 0.3970713E+03 0.3940433E+03 0.3910384E+03 0.3880564E+03 0.3850971E+03
 0.3821604E+03 0.3792461E+03 0.3763541E+03 0.3734840E+03 0.3706359E+03
 0.3678095E+03 0.3650047E+03 0.3622212E+03 0.3594590E+03 0.3567178E+03
 0.3539975E+03 0.3512980E+03 0.3486191E+03 0.3459605E+03
THE LOWER PROBABILITY LIMITS
 0.3911943E+03 0.3864870E+03 0.3819559E+03 0.3775741E+03 0.3733217E+03
 0.3691838E+03 0.3651485E+03 0.3612066E+03 0.3573503E+03 0.3535731E+03
 0.3498698E+03 0.3462357E+03 0.3426667E+03 0.3391594E+03 0.3357107E+03
 0.3323178E+03 0.3289783E+03 0.3256900E+03 0.3224508E+03 0.3192590E+03
 0.3161129E+03 0.3130109E+03 0.3099517E+03 0.3069338E+03
THE UPPER PROBABILITY LIMITS
 0.4347849E+03 0.4333397E+03 0.4317639E+03 0.4300844E+03 0.4283206E+03
 0.4264871E+03 0.4245954E+03 0.4226545E+03 0.4206717E+03 0.4186533E+03
 0.4166041E+03 0.4145287E+03 0.4124305E+03 0.4103129E+03 0.4081785E+03
 0.4060299E+03 0.4038692E+03 0.4016982E+03 0.3995187E+03 0.3973322E+03
 0.3951400E+03 0.3929434E+03 0.3907436E+03 0.3885414E+03
```

List of Symbols

$\hat{a}$ estimate of y-intercept

a_t independant, identically distributed white noise at time t

b defined as the reciprocal of n

$\hat{b}$ estimate of the slope

C constant of proportionality

C_k sample autocovariance of lag k

D constant fraction of decline per period of time

D_i initial fraction of decline per period of time

h defined as the reciprocal of nD_i

n power of production rate to which the decline in production is proportional

Np cummulative production

p order of autoregressive component of time series model

q (associated with time series) order of the moving average component of time series model

q (associated with decline curves) production rate

q_o initial production rate

r_k sample autocorrelation of lag k

t time period

t_o initial time period

t_f final time period

T total number of stationary observations used in the time series analysis

W_t stationary observation at time t

X_i realized values of independant variables

X_t logarithm of raw observation at time t

Y_i realized values of dependant variables

$\hat{Y}_i$ values of dependant variable predicted by the least squares line

Z_t raw observation at time t

α level of significance

θ_j coefficient of a_{t-j} in time series model

θ_o constant term in time series model

ρ_k autocorrelation of lag k

$\hat{\sigma}_a^2$ estimate of the variance of white noise

σ_a^2 variance of white noise

$\hat{\phi}_{kk}$ estimate of the kth partial autocorrelation

ϕ_{kk} kth partial autocorrelation

ϕ_i coefficient of W_{t-i} in time series model

Literature Cited

Arps, J. J., 1945, Analysis of Decline Curves: Trans. A.I.M.E., v. 160, p 228-247.

Box, G. P., and Jenkins, G. M., 1970, Time Series Analysis Forecasting and Control: San Francisco, Holden-Day, 553 p.

Brons, Folkert, 1963, On the Use and Misuse of Production Decline Curves: Producers Monthly, v. 27, no. 9, p 22 - 25.

Chisholm, R. K., and Whitaker G. R., Jr., 1971, Forecasting Methods: Homewood, Illinois, Richard D. Irwin, Inc., 177 p.

Cutler, W. W., Jr., 1924, Estimation of Underground Oil Reserves by Well Production Curves: U.S. Bur. Mines Bull. 228

Gopal, V. N., 1973, Computer Technique Simplifiers Decline - Curve Analysis: Oil and Gas Journal, v. 71, no. 6, p 50-52.

Hughes, R. V., 1967, Oil Property Valuation: New York, John Wiley and Sons, Inc., 313 p.

Johnston, J., 1972, Econometric Methods: New York, McGraw-Hill, 437 p.

Larkey, L. S., 1925, Mathematical Determination of Production Decline Curves: Trans. A.I.M.E., v. 71, p 1322 - 1328.

Mueller, R. K., Private communication, 1976.

Nelson, C. R., 1973, Applied Time Series Analysis for Managerial Forecasting: San Francisco, Holden-Day, 231 p.

Nind, T. E., 1964, Principles of Oil Well Production: New York, McGraw-Hill, 364 p.

Oduolowu, Akin, 1976, A Case Study of the Recoverable Hydrocarbon Volumes for the Minnelusa and Muddy Sands, Powder River Basin, Wyoming: Log Analyst, March-April, p 24-32.

Stevens, W. F., and Thodos, Geroge, 1961, Estimating Oil Reserves From Production Data: The Petroleum Engineer, v. 33, no. 2, p B-46 - B-51.